FILE HISTORY

US 11,840,730

PATENT:	11,840,730
INVENTORS:	PORRECA GREGORY LASERSON URI
TITLE:	Methods and compositions for evaluating genetic markers
APPLICATION NO:	US16952764A
FILED:	19 NOV 2020
ISSUED:	12 DEC 2023
COMPILED:	12 JUL 2024

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/952,764	Porreca et al.
	Examiner	**Art Unit**
	STEVEN C POHNERT PhD	1634

CPC						
Symbol					**Type**	**Version**
C12Q	/	1	/	6874	F	2013-01-01
C12Q	/	1	/	6827	I	2013-01-01
C12Q	/	1	/	6883	I	2013-01-01

CPC Combination Sets								
Symbol					**Type**	**Set**	**Ranking**	**Version**
	/		/					

NONE		Total Claims Allowed:	
(Assistant Examiner)	(Date)	9	
/Steven Pohnert/ Primary Examiner, Art Unit 1634	18 September 2023	O.G. Print Claim(s)	O.G. Print Figure
(Primary Examiner)	(Date)	1	none

U.S. Patent and Trademark Office — Part of Paper No.: 20230911

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/952,764	Porreca et al.
	Examiner	**Art Unit**
	STEVEN C POHNERT PhD	1634

INTERNATIONAL CLASSIFICATION					
CLAIMED					
C12Q1/6874	/	1	/	6874	
C12Q1/6827	/	1	/	6827	
C12Q1/6883	/	1	/	6883	

NON-CLAIMED				
	/		/	

US ORIGINAL CLASSIFICATION	
CLASS	**SUBCLASS**

CROSS REFERENCES(S)						
CLASS	**SUBCLASS (ONE SUBCLASS PER BLOCK)**					

NONE		**Total Claims Allowed:**	
(Assistant Examiner)	(Date)	9	
/Steven Pohnert/ Primary Examiner, Art Unit 1634	18 September 2023	O.G. Print Claim(s)	O.G. Print Figure
(Primary Examiner)	(Date)	1	none

U.S. Patent and Trademark Office — Part of Paper No.: 20230911

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/952,764	Porreca et al.
	Examiner	**Art Unit**
	STEVEN C POHNERT PhD	1634

☑ Claims renumbered in the same order as presented by applicant ☐ CPA ☐ T.D. ☐ R.1.47

CLAIMS

Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original
-	1	-	11	-	21	-	31	-	41	-	51	-	61	4	71
-	2	-	12	-	22	-	32	-	42	-	52	-	62	5	72
-	3	-	13	-	23	-	33	-	43	-	53	-	63	6	73
-	4	-	14	-	24	-	34	-	44	-	54	-	64	7	74
-	5	-	15	-	25	-	35	-	45	-	55	-	65	8	75
-	6	-	16	-	26	-	36	-	46	-	56	-	66	9	76
-	7	-	17	-	27	-	37	-	47	-	57	-	67		
-	8	-	18	-	28	-	38	-	48	-	58	1	68		
-	9	-	19	-	29	-	39	-	49	-	59	2	69		
-	10	-	20	-	30	-	40	-	50	-	60	3	70		

NONE		**Total Claims Allowed:**	
(Assistant Examiner)	(Date)	9	
/Steven Pohnert/ Primary Examiner, Art Unit 1634	18 September 2023	O.G. Print Claim(s)	O.G. Print Figure
(Primary Examiner)	(Date)	1	none

U.S. Patent and Trademark Office — Part of Paper No.: 20230911

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/952,764	Porreca et al.
	Examiner	Art Unit
	STEVEN C POHNERT PhD	1634

CPC - Searched*		
Symbol	Date	Examiner

CPC Combination Sets - Searched*		
Symbol	Date	Examiner

US Classification - Searched*			
Class	Subclass	Date	Examiner

* See search history printout included with this form or the SEARCH NOTES box below to determine the scope of the search.

Search Notes		
Search Notes	Date	Examiner
consulted Joe Woitach on 112, 101 and art	02/04/2021	SCP
searched EAST	02/06/2021	SCP
inventor search EAST and PALM	02/06/2021	SCP
Assignee search EAST	02/06/2021	SCP
google search	02/06/2021	SCP

Interference Search			
US Class/CPC Symbol	US Subclass/CPC Group	Date	Examiner

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/952,764	Porreca et al.
	Examiner	Art Unit
	STEVEN C POHNERT PhD	1634

CPC - Searched*		
Symbol	Date	Examiner

CPC Combination Sets - Searched*		
Symbol	Date	Examiner

US Classification - Searched*			
Class	Subclass	Date	Examiner

* See search history printout included with this form or the SEARCH NOTES box below to determine the scope of the search.

Search Notes		
Search Notes	Date	Examiner
consulted Joe Woitach on 112, 101 and art	02/04/2021	SCP
searched EAST	02/06/2021	SCP
inventor search EAST and PALM	02/06/2021	SCP
Assignee search EAST	02/06/2021	SCP
google search	02/06/2021	SCP
consulted Kevin Orwig on declaration, 112, 101 and art	04/06/2021	SCP
searched EAST	04/06/2021	SCP
inventor search EAST and PALM	04/06/2021	SCP
Assignee search EAST	04/06/2021	SCP
google search	04/06/2021	SCP

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/952,764	Porreca et al.
	Examiner	Art Unit
	STEVEN C POHNERT PhD	1634

Interference Search			
US Class/CPC Symbol	US Subclass/CPC Group	Date	Examiner

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/952,764	Porreca et al.
	Examiner	Art Unit
	STEVEN C POHNERT PhD	1634

CPC - Searched*		
Symbol	Date	Examiner

CPC Combination Sets - Searched*		
Symbol	Date	Examiner
C12Q1/6874 C12Q1/6827 C12Q1/6883	09/16/2023	SCP

US Classification - Searched*			
Class	Subclass	Date	Examiner

* See search history printout included with this form or the SEARCH NOTES box below to determine the scope of the search.

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/952,764	Porreca et al.
	Examiner	Art Unit
	STEVEN C POHNERT PhD	1634

Search Notes		
Search Notes	**Date**	**Examiner**
consulted Joe Woitach on 112, 101 and art	02/04/2021	SCP
searched EAST	02/06/2021	SCP
inventor search EAST and PALM	02/06/2021	SCP
Assignee search EAST	02/06/2021	SCP
google search	02/06/2021	SCP
consulted Kevin Orwig on declaration, 112, 101 and art	04/06/2021	SCP
searched EAST	04/06/2021	SCP
inventor search EAST and PALM	04/06/2021	SCP
Assignee search EAST	04/06/2021	SCP
google search	04/06/2021	SCP
searched P2E2	09/16/2023	SCP
inventor search P2E2 and PDP	09/16/2023	SCP
assignee search P2E2	09/16/2023	SCP
reviewed global dossier	09/16/2023	SCP
consulted Joe Woitach on PTAB decision	09/15/2023	SCP
google search	09/18/2023	SCP

Interference Search			
US Class/CPC Symbol	**US Subclass/CPC Group**	**Date**	**Examiner**
	C12Q1/6874 C12Q1/6827 C12Q1/6883	09/16/2023	SCP
	text searched claims	09/16/2023	SCP

Web Search History

date, time	web site	search string
9/18/2023 1:42:47 PM	Google	duplication bias tag barcode
9/18/2023 1:42:53 PM	Google	duplication bias barcode
9/18/2023 1:43:19 PM	Google	molecular barcode amplification artifact bias
9/18/2023 1:43:23 PM	Google Scholar	molecular barcode amplification artifact bias
9/18/2023 1:43:42 PM	Google Scholar	molecular barcode amplification artifact bias [before:2009]
9/18/2023 1:44:16 PM	Google Scholar	[before:2009] molecular barcode amplification artifact bias sequencing
9/18/2023 1:44:32 PM	Google Scholar	[before:2009] umi amplification artifact bias sequencing

PE2E SEARCH - Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	British Equivalents	Time Stamp
L1	57	((("PORRECA") near3 ("Gregory")) OR (("LASERSON") near3 ("Uri"))).INV.	(US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT)	OR	ON	ON	2021/04/06 10:16 AM
L2	85	(good near2 start near2 genetics).as. or (molecular near2 loop near2 biosolutions).as.	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2021/04/06 10:16 AM
L3	35	duplication near2 bias	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2021/04/06 10:18 AM
L4	1529	((duplication or amplification) near2 bias) with sequenc$4	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2021/04/06 10:18 AM
L5	1344	L4 and (tags or barcodes or apatmer or label)	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2021/04/06 10:19 AM
L6	673	L4 same (tags or barcodes or apatmer or label)	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2021/04/06 10:19 AM
L7	633	L6 and (error or correct$4)	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2021/04/06 10:20 AM
L9	633	L7 and (error or correct$4)	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2021/04/06 10:22 AM
L10	8	L7 and @ad<"20090430"	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2021/04/06 10:23 AM
L11	12	(L1 or L2) and L9	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2021/04/06 10:25 AM
L12	71	((("PORRECA") near3 ("Gregory")) OR (("LASERSON") near3 ("Uri"))).INV.	(US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT)	OR	ON	ON	2023/09/16 12:32 PM
L13	41	(good near2 start near2 genetics).as. OR (molecular near2 loop near2 biosolutions).as.	(US-PGPUB; USPAT; USOCR)	OR	ON	ON	2023/09/16 12:32 PM
L14	2187	((duplication OR	(US-PGPUB; USPAT;	OR	ON	ON	2023/09/16

Petitioner Illumina, Inc., EX1004, p. 11

		amplification) near2 bias) WITH sequenc$4	USOCR; FPRS; EPO; JPO; DERWENT)				12:33 PM
L15	131	L14 AND @ad<"20090430"	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2023/09/16 12:34 PM
L16	1932	(L13 OR L14) AND L5	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2023/09/16 12:35 PM
L17	11	(L13 orl12) AND L5	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2023/09/16 12:35 PM
L18	15	(L13 OR L12) AND L5	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2023/09/16 12:36 PM
L19	0	(L13 OR L12) AND L5.clm.	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2023/09/16 12:37 PM
L20	127902	C12Q1/6874.cpc. OR C12Q1/6827.cpc. OR C12Q1/6883.cpc.	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2023/09/16 12:43 PM
L21	71	((("PORRECA") near3 ("Gregory")) OR (("LASERSON") near3 ("Uri"))).INV.	(US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT)	OR	ON	ON	2023/09/16 12:44 PM
L22	94	(good near2 start near2 genetics).as. or (molecular near2 loop near2 biosolutions).as.	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2023/09/16 12:44 PM
L23	36	duplication near2 bias	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2023/09/16 12:44 PM
L24	2187	((duplication or amplification) near2 bias) with sequenc$4	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2023/09/16 12:44 PM
L25	1932	L4 and (tags or barcodes or apatmer or label)	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2023/09/16 12:44 PM
L26	1009	L4 same (tags or barcodes or apatmer or label)	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2023/09/16 12:44 PM
L27	954	L6 and (error or correct$4)	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2023/09/16 12:44 PM

L28	954	L7 and (error or correct$4)	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2023/09/16 12:44 PM
L29	8	L7 and @ad<"20090430"	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2023/09/16 12:44 PM
L30	15	(L1 or L2) and L9	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2023/09/16 12:44 PM
L31	71	((("PORRECA") near3 ("Gregory")) OR (("LASERSON") near3 ("Uri"))).INV.	(US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT)	OR	ON	ON	2023/09/16 12:44 PM
L32	41	(good near2 start near2 genetics).as. OR (molecular near2 loop near2 biosolutions).as.	(US-PGPUB; USPAT; USOCR)	OR	ON	ON	2023/09/16 12:44 PM
L33	2187	((duplication OR amplification) near2 bias) WITH sequenc$4	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT)	OR	ON	ON	2023/09/16 12:44 PM
L34	131	L14 AND @ad<"20090430"	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2023/09/16 12:44 PM
L35	1932	(L13 OR L14) AND L5	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2023/09/16 12:44 PM
L36	11	(L13 orl12) AND L5	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2023/09/16 12:44 PM
L37	15	(L13 OR L12) AND L5	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2023/09/16 12:44 PM
L38	0	(L13 OR L12) AND L5.clm.	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2023/09/16 12:44 PM
L39	127902	C12Q1/6874.cpc. OR C12Q1/6827.cpc. OR C12Q1/6883.cpc.	(US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2023/09/16 12:44 PM

PE2E SEARCH - Search History (Interference)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	British Equivalents	Time Stamp
N1	29739	C12Q1/6874.cpc. OR C12Q1/6827.cpc. OR	(US-PGPUB; USPAT)	OR	ON	ON	2023/09/16 12:43 PM

		C12Q1/6883.cpc.					
N2	1992	((duplication OR amplification) near2 bias) WITH sequenc$4	(US-PGPUB; USPAT)	OR	ON	ON	2023/09/16 12:43 PM
N3	742	N2 AND N1	(US-PGPUB; USPAT)	OR	ON	ON	2023/09/16 12:44 PM
N4	370	N3 AND workflow	(US-PGPUB; USPAT)	OR	ON	ON	2023/09/16 12:45 PM
N5	0	N3 AND workflow.clm.	(US-PGPUB; USPAT)	OR	ON	ON	2023/09/16 12:45 PM
N6	370	N3 AND workflow	(US-PGPUB; USPAT)	OR	ON	ON	2023/09/16 12:45 PM
N7	15	N2 SAME workflow	(US-PGPUB; USPAT)	OR	ON	ON	2023/09/16 12:45 PM

EAST Search History

EAST Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	57	((("PORRECA") near3 ("Gregory")) OR (("LASERSON") near3 ("Uri"))).INV.	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2021/04/06 10:16
L2	85	(good near2 start near2 genetics).as. or (molecular near2 loop near2 biosolutions).as.	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/04/06 10:16
L3	35	duplication near2 bias	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/04/06 10:18
L4	1,529	((duplication or amplification) near2 bias) with sequenc$4	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/04/06 10:18
L5	1,344	l4 and (tags or barcodes or apatmer or label)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/04/06 10:19
L6	673	l4 same (tags or barcodes or apatmer or label)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/04/06 10:19
L7	633	l6 and (error or correct$4)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/04/06 10:20
L9	633	l7 and (error or correct$4)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/04/06 10:22
L10	8	l7 and @ad<"20090430"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/04/06 10:23
L11	12	(l1 or l2) and l9	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/04/06 10:25

EAST Search History (Interference)

<This search history is empty>

4/6/2021 10:34:37 AM
C:\Users\spohnert\Documents\EAST\Workspaces\16952764.wsp

Web Search History

date, time	web site	search string
4/6/2021 10:31:46 AM	Google Scholar	duplication bias tag barcode
4/6/2021 10:31:54 AM	Google Scholar	amplifcation bias tag barcode
4/6/2021 10:31:57 AM	Google Scholar	amplification bias tag barcode
4/6/2021 10:32:06 AM	Google Scholar	molecular barcodes amplification artifacts
4/6/2021 10:32:16 AM	Google Scholar	molecular barcodes amplification artifacts [before:2009]
4/6/2021 10:32:26 AM	Google Scholar	[before:2009] molecular barcodes amplification artifacts sequencing
4/6/2021 10:32:47 AM	Google Scholar	[before:2009] umi sequencing error
4/6/2021 10:33:01 AM	Google Scholar	[before:2009] cleaning up sequencing reads
4/6/2021 10:33:28 AM	Google Scholar	[before:2009] duplication bias tag massively parallel sequencing

Web Search History

date, time	web site	search string
2/4/2021 8:48:51 AM	Google Scholar	reaD DUPLICATION BIAS
2/4/2021 8:49:09 AM	Google Scholar	reaD DUPLICATION BIAS [before:2008]
2/4/2021 8:49:30 AM	Google Scholar	[before:2008] reaD DUPLICATION BIAS SEQUENCING
2/4/2021 8:52:55 AM	Google Scholar	[before:2008] reaD DUPLICATION BIAS SEQUENCING TAGS
2/4/2021 8:53:01 AM	Google Scholar	[before:2008] reaD DUPLICATION BIAS SEQUENCING TAGS COLLPASE
2/4/2021 8:54:35 AM	Google Scholar	[before:2008] reaD DUPLICATION BIAS SEQUENCING TAGS COLLAPSE
2/4/2021 8:54:46 AM	Google Scholar	[before:2008] reaD DUPLICATION BIAS COLLAPSE
2/4/2021 8:55:04 AM	Google Scholar	reaD DUPLICATION BIAS COLLAPSE [before:2008]
2/4/2021 8:57:39 AM	Google Scholar	[before:2008] SEQUENCING reaD DUPLICATION BIAS COLLAPSE
2/4/2021 8:58:06 AM	Google Scholar	SEQUENCING reaD DUPLICATION BIAS COLLAPSE [before:2008]
2/4/2021 8:59:47 AM	Google Scholar	SEQUENCING reaD DUPLICATION BIAS COLLAPSE
2/4/2021 9:04:10 AM	Google Scholar	umi sequencing error
2/4/2021 9:07:06 AM	Google Scholar	umi sequencing error [before:2009]
2/4/2021 10:04:36 AM	Google Scholar	duplication bias in pcr for massively parallel sequencing
2/4/2021 10:05:13 AM	Google Scholar	duplication bias in pcr for massively parallel sequencing [before:2009]
2/4/2021 10:05:27 AM	Google Scholar	[before:2009] cleaning up sequencing reads

date, time	web site	search string
2/4/2021 10:06:21 AM	Google Scholar	cleaning up sequencing reads [before:2009]
2/4/2021 10:08:35 AM	Google Scholar	[before:2009] sequencing reads
2/4/2021 10:25:21 AM	Google Scholar	[before:2009] sequencing reads threshold
2/4/2021 10:25:25 AM	Google Scholar	sequencing reads threshold [before:2007]
2/6/2021 12:59:13 PM	Google	mips circulating tumor cells
2/6/2021 12:59:20 PM	Google Scholar	mips circulating tumor cells
2/6/2021 12:59:31 PM	Google Scholar	mips circulating tumor cells [before:2008]
2/6/2021 12:59:46 PM	Google Scholar	[before:2008] mips circulating tumor cells molecular inversion
2/6/2021 1:00:44 PM	Google Scholar	[before:2008] mips circulating nucleic acids molecular inversion
2/6/2021 1:01:10 PM	Google Scholar	[before:2008] mips circulating nucleic acids "molecular inversion probe"
2/6/2021 1:01:59 PM	Google Scholar	[before:2008] "cell free" nucleic acids "molecular inversion probe"
2/6/2021 1:02:14 PM	Google Scholar	[before:2008] cell free nucleic acids "molecular inversion probe"
2/6/2021 1:22:00 PM	Google Scholar	[before:2008] "molecular inversion probe"
2/6/2021 1:22:08 PM	Google Scholar	[before:2008] "molecular inversion probe" blood
2/6/2021 1:22:45 PM	Google Scholar	"molecular inversion probe" blood [before:2008]
2/6/2021 1:23:37 PM	Google Scholar	[before:2008] "molecular inversion probe" blood cell free
2/6/2021 1:28:15 PM	Google Scholar	[before:2008] "molecular inversion probe" tumor dna

EAST Search History

EAST Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	55	((("PORRECA") near3 ("Gregory")) OR (("LASERSON") near3 ("Uri"))).INV.	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2021/02/06 15:30
L2	78	(good near2 start near2 genetics).as. or (miolecular near2 loop near2 biosolutions).as.	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/02/06 15:31
L3	89	READ with duplication with (tag or differentiator or barcode)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/02/06 15:31
L4	11	L3 and bias	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/02/06 15:31
L5	0	(l1 or l2) and L4	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/02/06 15:31
S1	89	READ with duplication with (tag or differentiator or barcode)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/02/04 09:00
S2	11	S1 and bias	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/02/04 09:01
S3	1,155	umi and bias	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/02/04 09:10
S4	445	S3 and @pd<"20090430"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/02/04 09:12
S5	8,729	(tag or differentiator or barcode) same bias	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/02/04 09:14
S6	3,517	S5 and @pd<"20090430"	US-PGPUB; USPAT; USOCR; FPRS; EPO;	OR	ON	2021/02/04 09:14

			JPO; DERWENT; IBM_TDB			
S7	88,116	(molecular near2 inversion near2 probes) or mips	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/02/06 13:19
S8	7,302	S7 same ((circulating or (cell near2 free)) with rna or dna or nucleic)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/02/06 13:21
S9	1,844	S8 and @pd<"20080430"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/02/06 13:21
S10	124	S7 same ((circulating or (cell near2 free)) with (rna or dna or nucleic))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/02/06 13:25
S11	5	S10 and @pd<"20080430"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/02/06 13:26
S12	40,443	((circulating or (cell near2 free)) with (rna or dna or nucleic))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/02/06 13:27
S13	11,374	S12 and @pd<"20080430"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/02/06 13:27

EAST Search History (Interference)

<This search history is empty>

2/6/2021 3:32:11 PM
C:\Users\spohnert\Documents\EAST\Workspaces\16952764.wsp

US011840730B1

(12) **United States Patent**
Porreca et al.

(10) **Patent No.: US 11,840,730 B1**
(45) **Date of Patent: Dec. 12, 2023**

(54) **METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS**

(71) Applicant: **Molecular Loop Biosciences, Inc.**, Cambridge, MA (US)

(72) Inventors: **Gregory Porreca**, Cambridge, MA (US); **Uri Laserson**, Boston, MA (US)

(73) Assignee: **MOLECULAR LOOP BIOSCIENCES, INC.**, Cambridge, MA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 516 days.

(21) Appl. No.: **16/952,764**

(22) Filed: **Nov. 19, 2020**

Related U.S. Application Data

(63) Continuation of application No. 15/231,687, filed on Aug. 8, 2016, now abandoned, which is a continuation of application No. 13/266,862, filed as application No. PCT/US2010/001293 on Apr. 30, 2010.

(60) Provisional application No. 61/182,089, filed on May 28, 2009, provisional application No. 61/179,358, filed on May 18, 2009, provisional application No. 61/178,923, filed on May 15, 2009, provisional application No. 61/174,470, filed on Apr. 30, 2009.

(51) **Int. Cl.**
C12Q 1/6874 (2018.01)
C12Q 1/6827 (2018.01)
C12Q 1/6883 (2018.01)

(52) **U.S. Cl.**
CPC ***C12Q 1/6874*** (2013.01); ***C12Q 1/6827*** (2013.01); ***C12Q 1/6883*** (2013.01)

(58) **Field of Classification Search**
CPC ... C12Q 1/6874; C12Q 1/6827; C12Q 1/6883
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,683,195	A	7/1987	Mullis et al.
4,683,202	A	7/1987	Mullis
4,988,617	A	1/1991	Landegren et al.
5,060,980	A	10/1991	Johnson et al.
5,210,015	A	5/1993	Gelfand et al.
5,234,809	A	8/1993	Boom et al.
5,242,794	A	9/1993	Whiteley et al.
5,348,853	A	9/1994	Wang et al.
5,434,049	A	7/1995	Okano et al.
5,459,307	A	10/1995	Klotz, Jr.
5,486,686	A	1/1996	Zdybel, Jr. et al.
5,491,224	A	2/1996	Bittner et al.
5,494,810	A	2/1996	Barany et al.
5,567,583	A	10/1996	Wang et al.
5,583,024	A	12/1996	McElroy et al.
5,604,097	A	2/1997	Brenner
5,636,400	A	6/1997	Young
5,674,713	A	10/1997	McElroy et al.
5,695,934	A	12/1997	Brenner
5,700,673	A	12/1997	McElroy et al.
5,701,256	A	12/1997	Marr et al.
5,830,064	A	11/1998	Bradish et al.
5,846,719	A	12/1998	Brenner et al.
5,863,722	A	1/1999	Brenner
5,866,337	A	2/1999	Schon
5,869,252	A	2/1999	Bouma et al.
5,869,717	A	2/1999	Frame et al.
5,871,921	A	2/1999	Landegren et al.
5,888,788	A	3/1999	De Miniac
5,942,391	A	8/1999	Zhang et al.
5,971,921	A	10/1999	Timbel
5,993,611	A	11/1999	Moroney et al.
5,994,056	A	11/1999	Higuchi
6,020,127	A	2/2000	MacKenzie et al.
6,033,854	A	3/2000	Kurnit et al.
6,033,872	A	3/2000	Bergsma et al.
6,100,099	A	8/2000	Gordon et al.
6,138,077	A	10/2000	Brenner
6,150,516	A	11/2000	Brenner et al.
6,171,785	B1	1/2001	Higuchi
6,172,214	B1	1/2001	Brenner
6,172,218	B1	1/2001	Brenner
6,197,508	B1	3/2001	Stanley
6,197,574	B1	3/2001	Miyamoto et al.
6,210,891	B1	4/2001	Nyren et al.
6,223,128	B1	4/2001	Allex et al.
6,235,472	B1	5/2001	Landegren et al.
6,235,475	B1	5/2001	Brenner et al.
6,235,501	B1	5/2001	Gautsch et al.
6,235,502	B1	5/2001	Weissman et al.
6,258,568	B1	7/2001	Nyren
6,274,320	B1	8/2001	Rothberg et al.
6,306,597	B1	10/2001	Macevicz
6,352,828	B1	3/2002	Brenner
6,360,235	B1	3/2002	Tilt et al.
6,361,940	B1	3/2002	Van Ness et al.
6,403,320	B1	6/2002	Read et al.
6,462,254	B1	10/2002	Vernachio et al.
6,489,105	B1	12/2002	Matlashewski et al.
6,558,928	B1	5/2003	Landegren
6,569,920	B1	5/2003	Wen et al.
6,582,938	B1	6/2003	Su et al.
6,585,938	B1	7/2003	Machida et al.
6,613,516	B1	9/2003	Christians et al.

(Continued)

FOREIGN PATENT DOCUMENTS

EP	1 321 477	A1	6/2003
EP	1 564 306	A2	8/2005

(Continued)

OTHER PUBLICATIONS

Porecca(Nature methods (2007) vol. 4, pp. 931-936).*

(Continued)

Primary Examiner — Steven Pohnert
(74) *Attorney, Agent, or Firm* — Withers Bergman LLP; Thomas C. Meyers

(57) **ABSTRACT**

Aspects of the invention relates to methods and compositions that are useful to reduce bias and increase the reproducibility of multiplex analysis of genetic loci. In some configurations, predetermined preparative steps and/or nucleic acid sequence analysis techniques are used in multiplex analyses for a plurality of genetic loci in a plurality of samples.

9 Claims, 16 Drawing Sheets

Specification includes a Sequence Listing.

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,714,874 B1 3/2004 Myers et al.
6,716,580 B2 4/2004 Gold et al.
6,719,449 B1 4/2004 Laugharn, Jr. et al.
6,818,395 B1 11/2004 Quake et al.
6,828,100 B1 12/2004 Ronaghi
6,833,246 B2 12/2004 Balasubramanian
6,858,412 B2 2/2005 Willis et al.
6,911,345 B2 6/2005 Quake et al.
6,913,879 B1 7/2005 Schena
6,927,024 B2 8/2005 Dodge et al.
6,941,317 B1 9/2005 Chamberlin et al.
6,948,843 B2 9/2005 Laugharn, Jr. et al.
7,034,143 B1 4/2006 Preparata et al.
7,041,481 B2 5/2006 Anderson et al.
7,049,077 B2 5/2006 Yang
7,057,026 B2 6/2006 Barnes et al.
7,071,324 B2 7/2006 Preparata et al.
7,074,564 B2 7/2006 Landegren
7,074,586 B1 7/2006 Cheronis et al.
7,115,400 B1 10/2006 Adessi et al.
7,169,560 B2 1/2007 Lapidus et al.
7,211,390 B2 5/2007 Rothberg et al.
7,232,656 B2 6/2007 Balasubramanian et al.
7,244,559 B2 7/2007 Rothberg et al.
RE39,793 E 8/2007 Brenner
7,264,929 B2 9/2007 Rothberg et al.
7,282,337 B1 10/2007 Harris
7,297,518 B2 11/2007 Quake et al.
7,320,860 B2 1/2008 Landegren et al.
7,323,305 B2 1/2008 Leamon et al.
7,335,762 B2 2/2008 Rothberg et al.
7,351,528 B2 4/2008 Landegren
7,393,665 B2 7/2008 Brenner
7,510,829 B2 3/2009 Faham et al.
7,523,117 B2 4/2009 Zhang et al.
7,537,889 B2 5/2009 Sinha et al.
7,537,897 B2 5/2009 Brenner et al.
7,544,473 B2 6/2009 Brenner
7,582,431 B2 9/2009 Drmanac et al.
7,598,035 B2 10/2009 Macevicz
7,629,151 B2 12/2009 Gold et al.
7,642,056 B2 1/2010 Ahn et al.
7,666,593 B2 2/2010 Lapidus
7,700,323 B2 4/2010 Willis et al.
7,774,962 B1 8/2010 Ladd
7,776,616 B2 8/2010 Heath et al.
RE41,780 E 9/2010 Anderson et al.
7,790,388 B2 9/2010 Landegren et al.
7,809,509 B2 10/2010 Milosavljevic
7,835,871 B2 11/2010 Kain et al.
7,862,999 B2 1/2011 Zheng et al.
7,865,534 B2 1/2011 Genstruct
7,883,849 B1 2/2011 Dahl
7,957,913 B2 6/2011 Chinitz et al.
7,960,120 B2 6/2011 Rigatti et al.
7,985,716 B2 7/2011 Yershov et al.
7,993,880 B2 8/2011 Willis et al.
8,024,128 B2 9/2011 Rabinowitz et al.
8,114,027 B2 2/2012 Triva
8,165,821 B2 4/2012 Zhang
8,209,130 B1 6/2012 Kennedy et al.
8,283,116 B1 10/2012 Bhattacharyya et al.
8,462,161 B1 6/2013 Barber
8,463,895 B2 6/2013 Arora et al.
8,474,228 B2 7/2013 Adair et al.
8,496,166 B2 7/2013 Burns et al.
8,529,744 B2 9/2013 Marziali et al.
8,738,300 B2 5/2014 Porreca et al.
8,778,609 B1 7/2014 Umbarger
8,812,422 B2 8/2014 Nizzari et al.
8,847,799 B1 9/2014 Kennedy et al.
8,976,049 B2 3/2015 Kennedy et al.
9,074,244 B2 7/2015 Sparks et al.
9,115,387 B2 8/2015 Umbarger
9,228,233 B2 1/2016 Kennedy et al.
9,292,527 B2 3/2016 Kennedy et al.
9,535,920 B2 1/2017 Kennedy et al.
9,567,639 B2 2/2017 Oliphant et al.
9,677,124 B2 6/2017 Umbarger
10,066,259 B2 9/2018 Gore et al.
10,202,637 B2 2/2019 Umbarger
10,227,635 B2 3/2019 Umbarger et al.
10,604,799 B2 3/2020 Porreca et al.
10,683,533 B2 6/2020 Umbarger et al.
2001/0007742 A1 7/2001 Landergren
2001/0046673 A1 11/2001 French et al.
2002/0001800 A1 1/2002 Lapidus
2002/0040216 A1 4/2002 Dumont et al.
2002/0042052 A1 4/2002 Nilsen et al.
2002/0091666 A1 7/2002 Rice et al.
2002/0164629 A1 11/2002 Quake et al.
2002/0172954 A1 11/2002 Mao et al.
2002/0182609 A1 12/2002 Arcot
2002/0187496 A1 12/2002 Andersson et al.
2002/0190663 A1 12/2002 Rasmussen
2003/0166057 A1 9/2003 Hildebrand et al.
2003/0175709 A1 9/2003 Murphy et al.
2003/0177105 A1 9/2003 Xiao et al.
2003/0203370 A1 10/2003 Yakhini et al.
2003/0208454 A1 11/2003 Rienhoff et al.
2003/0224384 A1 12/2003 Sayood et al.
2004/0029264 A1 2/2004 Robbins
2004/0053275 A1 3/2004 Shafer
2004/0106112 A1 6/2004 Nilsson et al.
2004/0121373 A1 6/2004 Friedlander et al.
2004/0142325 A1 7/2004 Mintz et al.
2004/0152108 A1 8/2004 Keith et al.
2004/0161773 A1 8/2004 Rogan et al.
2004/0170965 A1 9/2004 Scholl et al.
2004/0171051 A1 9/2004 Holloway
2004/0197813 A1 10/2004 Hoffman et al.
2004/0209299 A1 10/2004 Pinter et al.
2005/0003369 A1 1/2005 Christians et al.
2005/0026204 A1 2/2005 Landegren
2005/0032095 A1 2/2005 Wigler et al.
2005/0048505 A1 3/2005 Fredrick et al.
2005/0059048 A1 3/2005 Gunderson et al.
2005/0100900 A1 5/2005 Kawashima et al.
2005/0112590 A1 5/2005 Boom et al.
2005/0186589 A1 8/2005 Kowalik et al.
2005/0214811 A1 9/2005 Margulies et al.
2005/0244879 A1 11/2005 Schumm et al.
2005/0250147 A1 11/2005 Macevicz
2005/0272065 A1 12/2005 Lakey et al.
2006/0008824 A1 1/2006 Ronaghi et al.
2006/0019304 A1 1/2006 Hardenbol et al.
2006/0024681 A1 2/2006 Smith et al.
2006/0030536 A1 2/2006 Yu et al.
2006/0078894 A1 4/2006 Winkler et al.
2006/0149047 A1 7/2006 Nanduri et al.
2006/0177837 A1 8/2006 Borozan et al.
2006/0183132 A1 8/2006 Fu et al.
2006/0192047 A1 8/2006 Goossen
2006/0195269 A1 8/2006 Yeatman et al.
2006/0246500 A1 11/2006 Browne
2006/0263789 A1 11/2006 Kincaid
2006/0281098 A1 12/2006 Miao et al.
2006/0286577 A1 12/2006 Jia
2006/0292585 A1 12/2006 Nautiyal et al.
2006/0292611 A1 12/2006 Berka et al.
2007/0009925 A1 1/2007 Fang et al.
2007/0020640 A1 1/2007 McCloskey et al.
2007/0042369 A1 2/2007 Reese et al.
2007/0092883 A1 4/2007 Schouten et al.
2007/0114362 A1 5/2007 Feng et al.
2007/0128624 A1 6/2007 Gormley et al.
2007/0161013 A1 7/2007 Hantash
2007/0162983 A1 7/2007 Hesterkamp et al.
2007/0166705 A1 7/2007 Milton et al.
2007/0212704 A1 9/2007 Dong et al.
2007/0225487 A1 9/2007 Nilsson et al.
2007/0238122 A1 10/2007 Allbritton et al.
2007/0244675 A1 10/2007 Shai et al.
2007/0264653 A1 11/2007 Berlin et al.

(56) **References Cited**

U.S. PATENT DOCUMENTS

2008/0003142 A1 1/2008 Link et al.
2008/0014589 A1 1/2008 Link et al.
2008/0076118 A1 3/2008 Tooke et al.
2008/0081330 A1 4/2008 Kahvejian
2008/0085836 A1 4/2008 Kearns et al.
2008/0090239 A1 4/2008 Shoemaker et al.
2008/0125324 A1 5/2008 Petersdorf et al.
2008/0176209 A1 7/2008 Muller et al.
2008/0269068 A1 10/2008 Church et al.
2008/0280955 A1 11/2008 McCamish
2008/0293589 A1 11/2008 Shapero
2009/0009904 A1 1/2009 Yasuna et al.
2009/0019156 A1 1/2009 Mo et al.
2009/0026082 A1 1/2009 Rothberg et al.
2009/0029385 A1 1/2009 Christians et al.
2009/0035777 A1 2/2009 Kokoris et al.
2009/0042206 A1 2/2009 Schneider et al.
2009/0098551 A1 4/2009 Landers et al.
2009/0099041 A1 4/2009 Church et al.
2009/0105081 A1 4/2009 Rodesch et al.
2009/0119313 A1 5/2009 Pearce
2009/0127589 A1 5/2009 Rothberg et al.
2009/0129647 A1 5/2009 Dimitrova et al.
2009/0156412 A1 6/2009 Boyce, IV et al.
2009/0163366 A1 6/2009 Nickerson et al.
2009/0181389 A1 7/2009 Li et al.
2009/0191565 A1 7/2009 Lapidus et al.
2009/0192047 A1 7/2009 Parr et al.
2009/0202984 A1 8/2009 Cantor
2009/0203014 A1 8/2009 Wu et al.
2009/0220955 A1 9/2009 Verrant
2009/0226975 A1 9/2009 Sabot et al.
2009/0233814 A1 9/2009 Bashkirov et al.
2009/0298064 A1 12/2009 Batzoglou et al.
2009/0301382 A1 12/2009 Patel
2009/0318310 A1 12/2009 Liu et al.
2010/0035243 A1 2/2010 Muller et al.
2010/0035252 A1 2/2010 Rothberg et al.
2010/0063742 A1 3/2010 Hart et al.
2010/0069263 A1 3/2010 Shendure et al.
2010/0076185 A1 3/2010 Adey et al.
2010/0086914 A1 4/2010 Bentley et al.
2010/0086926 A1 4/2010 Craig et al.
2010/0105107 A1 4/2010 Hildebrand et al.
2010/0137143 A1 6/2010 Rothberg et al.
2010/0137163 A1 6/2010 Link et al.
2010/0143908 A1 6/2010 Gillevet
2010/0159440 A1 6/2010 Messier et al.
2010/0188073 A1 7/2010 Rothberg et al.
2010/0196911 A1 8/2010 Hoffman et al.
2010/0197507 A1 8/2010 Rothberg et al.
2010/0216151 A1 8/2010 Lapidus et al.
2010/0216153 A1 8/2010 Lapidus et al.
2010/0248984 A1 9/2010 Shaffer et al.
2010/0282617 A1 11/2010 Rothberg et al.
2010/0285578 A1 11/2010 Selden et al.
2010/0297626 A1 11/2010 McKernan et al.
2010/0300559 A1 12/2010 Schultz et al.
2010/0300895 A1 12/2010 Nobile et al.
2010/0301042 A1 12/2010 Kahlert
2010/0301398 A1 12/2010 Rothberg et al.
2010/0304982 A1 12/2010 Hinz et al.
2010/0311061 A1 12/2010 Korlach et al.
2010/0330619 A1 12/2010 Willis et al.
2011/0004413 A1 1/2011 Carnevali et al.
2011/0009278 A1 1/2011 Kain et al.
2011/0015863 A1 1/2011 Pevzner et al.
2011/0021366 A1 1/2011 Chinitz et al.
2011/0034342 A1 2/2011 Fox
2011/0092375 A1 4/2011 Zamore et al.
2011/0098193 A1 4/2011 Kingsmore et al.
2011/0117544 A1 5/2011 Lexow
2011/0159499 A1 6/2011 Hindson et al.
2011/0166029 A1 7/2011 Margulies et al.
2011/0224105 A1 9/2011 Kum et al.
2011/0230365 A1 9/2011 Rohlfs et al.
2011/0257889 A1 10/2011 Klammer et al.
2011/0288780 A1 11/2011 Rabinowitz et al.
2011/0301042 A1 12/2011 Steinmann et al.
2012/0015050 A1 1/2012 Abkevich et al.
2012/0021930 A1 1/2012 Schoen et al.
2012/0046877 A1 2/2012 Hyland et al.
2012/0059594 A1 3/2012 Hatchwell et al.
2012/0074925 A1 3/2012 Oliver
2012/0079980 A1 4/2012 Taylor et al.
2012/0115736 A1 5/2012 Bjornson et al.
2012/0164630 A1 6/2012 Porreca et al.
2012/0165202 A1 6/2012 Porreca et al.
2012/0179384 A1 7/2012 Kuramitsu et al.
2012/0214678 A1 8/2012 Rava et al.
2012/0216151 A1 8/2012 Sarkar et al.
2012/0236861 A1 9/2012 Ganeshalingam et al.
2012/0245041 A1 9/2012 Brenner et al.
2012/0252020 A1 10/2012 Shuber
2012/0252684 A1 10/2012 Selifonov et al.
2012/0258461 A1 10/2012 Weisbart
2012/0270212 A1 10/2012 Rabinowitz et al.
2012/0270739 A1 10/2012 Rava et al.
2013/0129755 A1 5/2013 Song
2013/0130921 A1 5/2013 Gao et al.
2013/0178378 A1 7/2013 Hatch et al.
2013/0183672 A1 7/2013 de Laat et al.
2013/0222388 A1 8/2013 McDonald
2013/0268206 A1 10/2013 Porreca et al.
2013/0268474 A1 10/2013 Nizzari et al.
2013/0274146 A1 10/2013 Umbarger et al.
2013/0275103 A1 10/2013 Struble et al.
2013/0288242 A1 10/2013 Stoughton et al.
2013/0323730 A1 12/2013 Curry et al.
2013/0332081 A1 12/2013 Reese et al.
2013/0337447 A1 12/2013 Porreca et al.
2013/0344096 A1 12/2013 Chiang et al.
2014/0129201 A1 5/2014 Kennedy et al.
2014/0136120 A1 5/2014 Colwell et al.
2014/0206552 A1 7/2014 Rabinowitz et al.
2014/0222349 A1 8/2014 Higgins et al.
2014/0228226 A1 8/2014 Yin et al.
2014/0255931 A1 9/2014 Porreca et al.
2014/0308667 A1 10/2014 Umbarger
2014/0318274 A1 10/2014 Zimmerman et al.
2014/0342354 A1 11/2014 Evans et al.
2014/0361022 A1 12/2014 Finneran
2015/0051085 A1 2/2015 Vogelstein et al.
2015/0056613 A1 2/2015 Kural
2015/0111208 A1 4/2015 Umbarger et al.
2015/0178445 A1 6/2015 Cibulskis et al.
2015/0258170 A1 9/2015 McCabe et al.
2015/0299767 A1 10/2015 Armour et al.
2015/0310163 A1 10/2015 Kingsmore et al.
2015/0354003 A1 12/2015 Umbarger
2016/0003812 A1 1/2016 Porreca et al.
2016/0034638 A1 2/2016 Spence et al.
2016/0068889 A1 3/2016 Gole et al.
2016/0210486 A1 7/2016 Porreca et al.
2016/0251719 A1 9/2016 Umbarger
2017/0044610 A1 2/2017 Johnson
2017/0129964 A1 5/2017 Cheung
2017/0183731 A1 6/2017 Mann et al.
2017/0275676 A1 9/2017 Umbarger
2018/0371533 A1 12/2018 Gore et al.
2019/0233881 A1 8/2019 Umbarger et al.
2020/0181696 A1 6/2020 Porreca et al.

FOREIGN PATENT DOCUMENTS

EP 10770071.8 11/2010
EP 2425240 A2 3/2012
EP 2 437 191 A2 4/2012
EP 2716766 A1 4/2014
WO 95/011995 A1 5/1995
WO 1996/019586 A1 6/1996
WO 98/014275 A1 4/1998
WO 98/044151 A1 10/1998
WO 00/018957 A1 4/2000

(56) **References Cited**

FOREIGN PATENT DOCUMENTS

WO	02/093453	A2	11/2002
WO	2004/018497	A2	3/2004
WO	2004/083819	A2	9/2004
WO	2005/003304	A2	1/2005
WO	2006/084132	A2	8/2006
WO	2007/010251	A2	1/2007
WO	2007/061284	A1	5/2007
WO	2007/107717	A1	9/2007
WO	2007/123744	A2	11/2007
WO	2007/135368	A2	11/2007
WO	2008067551	A2	6/2008
WO	2009/036525	A2	3/2009
WO	2009/076238	A2	6/2009
WO	2010/024894	A1	3/2010
WO	2010/115154	A1	10/2010
WO	2010/126614	A2	11/2010
WO	2011/006020	A1	1/2011
WO	2011066476	A1	6/2011
WO	2011067378	A1	6/2011
WO	2011/102998	A2	8/2011
WO	2011/155833	A2	12/2011
WO	2012/006291	A2	1/2012
WO	2012040387	A1	3/2012
WO	2012/051208	A2	4/2012
WO	2012/087736	A1	6/2012
WO	2012/109500	A2	8/2012
WO	2012/134884	A1	10/2012
WO	2012/149171	A1	11/2012
WO	2012/170725	A2	12/2012
WO	2013/058907	A1	4/2013
WO	2013/148496	A1	10/2013
WO	2013/177086	A1	11/2013
WO	2013/191775	A2	12/2013
WO	2014/052909	A2	4/2014
WO	2014/074246	A1	5/2014
WO	2015/119941	A2	8/2015

OTHER PUBLICATIONS

Wang (Genome Biology (2007) vol. 8, R246.1-246.14).*

Schiffman (Journal of Clinical Oncology (2007) vol. 25, p. 530).*

Tan, 2014, Clinical outcome of preimplantation genetic diagnosis and screening using next generation sequencing, GigaScience 3(30):1-9.

Thauvin-Robinet, 2009, The very low penetrance of cystic fibrosis for the R117H mutation: a reappraisal for genetic counseling and newborn screening, J Med Genet 46:752-758.

Thiyagarajan, 2006, PathogenMIPer: a tool for the design of molecular inversion probes to detect multiple pathogens, BMC Bioinformatics 7:500.

Thompson, 1994, Clustal W: improving the sensitivity of progressive mulitple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nuc Acids Res 22:4673-80.

Thompson, 2011, The properties and applications of single-molecule DNA sequencing, Genome Biol 12(2):217.

Thorstenson, 1998, An Automated Hydrodynamic Process for Controlled, Unbiased DNA Shearing, Genome Res 8(8): 848-855.

Thorvaldsdottir, 2012, Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration, Brief Bioinform 24(2):178-92.

Tkachuk, 1990, Detection of bcr-abl Fusion in Chronic Myelogeneous Leukemia by in Situ Hybridization, Science 250:559.

Tobler, 2005, The SNPlex Genotyping System: A Flexible and Scalable Platform for SNP Genotyping, J Biomol Tech 16(4):398.

Tokino, 1996, Characterization of the human p57 KIP2 gene: alternative splicing, insertion/deletion polymorphisms in VNTR sequences in the coding region, and mutational analysis, Human Genetics 96:625-31.

Treangen, 2011, Repetitive DNA and next-generation sequencing: computational challenges and solutions, Nat Rev Gen 13(1):36-46.

Turner, 2009, Massively parallel exon capture and library-free resequencing across 16 genomes, Nat Meth 6:315-316.

Turner, 2009, Methods for genomic partitioning, Ann Rev Hum Gen 10:263-284.

Umbarger, 2013, Detecting contamination in Next Generation DNA sequencing libraries, American Society of Human Genetics 63rd Annual Meeting, Abstract, Oct. 22, 2013.

Umbarger, 2014, Next-generation carrier screening, Gen Med 16(2):132-140.

Veeneman, 2012, Oculus: faster sequence alignment by streaming read compression, BMC Bioinformatics 13:297.

Wahl, 1979, Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridization by using dextran sulfate, PNAS 76:3683-3687.

Wallace, 1979, Hybridization of synthetic oligodeoxyribonucteotides to dp x 174DNA:the effect of single base pair mismatch, Nucl Acids Res 6:3543-3557.

Wallace, 1987, Oligonucleotide probes for the screening of recombinant DNA libraries, Meth Enz 152:432-442.

Wang, 2005, Allele quantification using molecular inversion probes (MIP), Nucleic Acids Res 33(21):e183.

Warner, 1996, A general method for the detection of large CAG repeat expansions by fluorescent PCR, J Med Genet 33(12):1022-6.

Warren, 2007, Assembling millions of short DNA sequences using SSAKE, Bioinformatics, 23:500-501.

Waszak, 2010, Systematic inference of copy-number genotypes from personal genome sequencing data reveals extensive olfactory gene content diversity, PLoS Comp Biol 6(11):e1000988.

Watson, 2004, Cystic fibrosis population carrier screening: 2004 revision of American College of Medical Genetics mutation panel, Genetics in Medicine 6(5):387-391.

Williams, 2003, Restriction endonucleases classification, properties, and applications, Mol Biotechnol 23(3):225-43.

Wirth, 1999, Quantitative analysis of survival motor neuron copies, Am J Hum Genet 64:1340-1356.

Wittung, 1997, Extended DNA-Recognition Repertoire of Peptide Nucleic Acid (PNA): PNA-dsDNA Triplex Formed with Cytosine-Rich Homopyrimidine PNA, Biochemistry 36:7973.

Wu, 1998, Sequencing regular and labeled oligonucleotides using enzymatic digestion and ionspray mass spectrometry, Anal Biochem 263:129-138.

Wu, 2001, Improved oligonucleotide sequencing by alkaline phosphatase and exonuclease digestions with mass spectrometry, Anal Biochem 290:347-352.

Xu, 2012, FastUniq: A fast de novo duplicates removal tool for paired short reads, PLoS One 7(12):e52249.

Yau, 1996, Accurate diagnosis of carriers of deletions and duplications in Duchenne/Becker muscular dystrophy by fluorescent dosage analysis, J Med Gen 33(7):550-8.

Ye, 2009, Pindel: a pattern growth approach to detect break points of large deletions and medium size insertions from paired-end short reads, Bioinformatics 25(21):2865-2871.

Yershov, 1996, DNA analysis and diagnostics on oligonucleotide microchips, PNAS 93:4913-4918.

Yoo, 2009, Applications of DNA microarray in disease diagnostics, J Microbiol Biotech19(7):635-46.

Yoon, 2014, MicroDuMIP: target-enrichment technique for microarray-based duplex molecular inversion probes, Nucl Ac Res 43(5):e28.

Yoshida, 2004, Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage, Cancer Science 95(11)866-71.

Yu, 2007, A novel set of DNA methylation markers in urine sediments for sensitive/specific detection of bladder cancer, Clin Cancer Res 13(24):7296-7304.

Yuan, 1981, Structure and mechanism of multifunctional restriction endonucleases, Ann Rev Biochem 50:285-319.

Zerbino, 2008, Velvet: Algorithms for de novo short read assembly using de Bruijn graphs, Genome Research 18 (5):821-829.

Zhang, 2011, Is Mitochondrial tRNAphe Variant m.593T.Ca Synergistically Pathogenic Mutation in Chinese LHON Families with m.11778G.A? PLoS ONE 6(10) e26511.

Zhao, 2009, PGA4genomics for comparative genome assembly based on genetic algorithm optimization, Genomics 94 (4):284-6.

(56) **References Cited**

OTHER PUBLICATIONS

Zheng, 2011, iAssembler: a package for de novo assembly of Roche-454/Sanger transcriptome sequences, BMC Bioinformatics 12:453.
Zhou, 2014, Bias from removing read duplication in ultra-deep sequencing experiments, Bioinformatics 30 (8):1073-1080.
Zhulidov, 2004, Simple cDNA normalization using kamchatka crab duplex-specific nuclease, Nucl Acids Res 32(3):e37.
Zimmerman, 2010, A novel custom resequencing array for dilated cardiomyopathy, Gen Med 12(5):268-78.
Zimran, 1990, A glucocerebrosidase fusion gene in Gaucher disease, J Clin Invest 85:219-222.
Zuckerman, 1987, Efficient methods for attachment of thiol specific probes to the 3'-ends of synthetic bligodeoxyribonucleotides, Nucl Acid Res 15(13):5305-5321.
Craig, 1997, Removal of repetitive sequences from FISH probes, Hum Genet 100:472.
Cremers, 1998, Autosomal Recessive Retinitis Pigmentosa and Cone-Rod Dystrophy Caused by Splice Site Mutations In the Stargardt's Disease Gene ABCR, Hum Mol Gen 7(3):355.
Cronin, 1996, Cystic Fibrosis Mutation Detection by Hybridization to Light-Generated DNA Probe Arrays Human Mutation 7:244.
Dahl, 2005, Multiplex amplification enabled by selective circularization of large sets of genomic DNA fragments, Nucleic Acids Res 33(8):e71.
Danecek, 2011, The variant call format and VCFtools, Bioinformatics 27(15):2156-2158.
De la Bastide, 2007, Assembling genome DNA sequences with PHRAP, Current Protocols in Bioinformatics 17:11.4.1-11.4.15.
Delcher, 1999, Alignment of whole genomes, Nuc Acids Res 27(11):2369-2376.
Den Dunnen, 2003, Mutation Nomenclature, Curr Prot Hum Genet 7.13.1-7.13.8.
Deng et al., 2012, Supplementary Material, Nature Biotechnology, S1-1-S1-1 1, Retrieved from the Internet on Oct. 24, 2012.
Deng, 2009, targeted bisulfite sequencing reveals changes in DNA methylation, Nat Biotech 27(4):353-360.
Deorowicz, 2013, Data compression for sequencing data, Alg for Mole Bio 8:25.
Diep, 2012, Library-free methylation sequencing with bisulfite padlock probes, Nature Methods 9:270-272 (and supplemental information).
DiGuistini, 2009, De novo sequence assembly of a filamentous fungus using Sanger, 454 and Illumina sequence data, Genome Biology, 10:R94.
Dolinsek, 2013, Depletion of unwanted nucleic acid templates by selection cleavage: LNAzymes, catalytically active bligonucleotides containing locked nucleic acids, open a new window for detecting rare microbial community members, App Env Microbiol 79(5):1534-1544.
Dong, 2011, Mutation surveyor: An in silico tool for sequencing analysis, Methods Mol Biol 760:223-37.
Drmanac, 1992, Sequencing by hybridization: towards an automated sequencing of one million M13 clones arrayed on membranes, Elctrophoresis 13:566-573.
Dudley, 2009, A quick guide for developing effective bioinformatics programming skills, PLoS Comp Biol 5(12): e1000589.
Ericsson, 2008, A dual-tag microarray platform for high-performance nucleic acid and protein analyses, Nucl Acids Res 36:e45.
Fares, 2008, Carrier frequency of autosomal-recessive disorders in the Ashkenazi Jewish population: should the rationale for mutation choice for screening be reevaluated?, Prenatal Diagnosis 28:236-41.
Faulstich, 1997, A sequencing method for RNA oligonucleotides based on mass spectrometry, Anal Chem 69:4349-4353.
Faust, 2014, SAMBLASTER: fast duplicate marking and structural variant read extraction, Bioinformatics published online May 7, 2014.
Fitch, 1970, Distinguishing homologs from analogous proteins, Syst Biol 19(2):99-113.
Flaschker, 2007, Description of the mutations in 15 subjects with variant forms of maple syrup urine disease, J Inherit Metab Dis 30:903-909.
Frey, 2006, Statistics Hacks 108-115.
Friedenson, 2005, BRCA1 and BRCA2 Pathways and the Risk of Cancers Other Than Breast or Ovarian, Medscape General Medicine 7(2):60.
Furtado, 2011, Characterization of large genomic deletions in the FBN1 gene using multiplex ligation-dependent probe amplification, BMC Med Gen 12:119-125.
Garber, 2008, Fixing the front end, Nat Biotech 26(10):1101-1104.
Gemayel, 2010, Variable tandem repeats accelerate evolution of coding and regulatory sequences, Ann Rev Genet 44:445-77.
Giusti, 1993, Synthesis and Characterization of f'-Fluorescent-dye-labeled Oligonucleotides, PCR Meth Appl 2:223-227.
Glover, 1995, Sequencing of oligonucleotides using high performance liquid chromatography and electrospray mass spectrometry, Rapid Com Mass Spec 9:897-901.
Gnirke, 2009, Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing, nature biotechnology 27:182-9.
Goto, 1994, A Study on Development of a Deductive Object-Oriented Database and Its Application to Genome Analysis, PhD Thesis, Kyushu University, Kyushu, Japan (106 pages).
Goto, 2010, BioRuby: bioinformatics software for the Ruby programming language, Bioinformatics 26(20):2617-2619.
Green, 2005, Suicide polymerase endonuclease restriction, a novel technique for enhancing PCR amplification of minor DNA template, Appl Env Microbiol 71(8):4721-4727.
Guerrero-Fernandez, 2013, FQbin: a compatible and optimize dformat for storing and managing sequence data, IWBBIO Proceedings, Granada 337-344.
Gupta, 1991, A general method for the synthesis of 3'-sulfhydryl and phosphate group containing oligonucleotides, Nucl Acids Res 19(11):3019-3025.
Gupta, 2014, Expanding the genetic toolkit: ZFNs, TALENs, and CRISPR-Cas9, J Clin Invest 124(10):4154.
Gustincich, 1991, A fast method for high-quality genomic DNA extraction from whole human blood, BioTechniques 11 (3):298-302.
Gut, 1995, A procedure for selective DNA alkylation and detection by mass spectrometry, Nucl Acids Res 23 (8):1367-1373.
Hallam, 2014, Validation for Clinical Use of, and Initial Clinical Experience with, a Novel Approach to Population-Based Carrier Screening using High-Throughput Next-Generation DNA Sequencing, J Mol Diagn 16:180-9.
Hammond, 1996, Extraction of DNA from preserved animal specimens for use in randomly amplified polymorphic DNA analysis, Anal Biochem 240:298-300.
Hardenbol, 2003, Multiplexed genotyping with sequence-tagged molecular inversion probes, Nat Biotech 21:673-8.
Hardenbol, 2005, Highly multiplexed molecular inversion probe genotyping: over 10,000 targeted SNPs genotyped in a single tube assay, Genome Res 15:269-75.
Harris, 2006, Defects can increase the melting temperature of DNA-nanoparticle assemblies, J Phys Chem B 110 (33):16393-6.
Harris, 2008, Helicos True Single Molecule Sequencing (tSMS) Science 320:106-109.
Harris, 2008, Single-molecule DNA sequencing of a viral genome, Science 320(5872):106-9.
Heger, 2006, Protonation of Cresol Red in Acidic Aqueous Solutions Caused by Freezing, J Phys Chem B 110 (3):1277-1287.
Heid, 1996, Real time quantitative PCR, Genome Res 6:986-994.
Hiatt, 2013, Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation, Genome Res 23:843-54.
Hodges, 2007, Genome-wide in situ exon capture for selective resequencing, Nat Genet 39(12):1522-7.
Abravaya, 1995, Detection of point mutations with a modified ligase chain reaction (GAP-LCR), Nucleic Acids Research, 23(4): 675-682.

(56) **References Cited**

OTHER PUBLICATIONS

Adey, 2010, Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition, Genome Biol 11:R119.
Ageno, 1969, The alkaline denaturation of DNA, Biophys J 9(11):1281-1311.
Agrawal, 1990, Site-specific functionalization of oligodeoxynucleotides for non-radioactive labelling, Tetrahedron Let 31:1543-1546.
Akhras, 2007, Connector inversion probe technology: A powerful one-primer multiplex DNA amplification system for numerous scientific applications, PLoSOne 9:e915.
Akhras, 2007, PathogenMip Assay: a multiplex pathogen detection assay, PLOS One 2:e2230.
Alazard, 2002, Sequencing of production-scale synthetic oligonucleotides by enriching for coupling failures using matrix-assisted laser desorption/ ionization time of-flight mass spectrometry, Anal Biochem 301:57-64.
Alazard, 2006, Sequencing oligonucleotides by enrichment of coupling failures using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, Curr Protoc Nucleic Acid Chem, Chapter 10, Unit 10:1-7.
Albert, 2007, Direct selection of human genomic loci by microarray hybridization, Nature Methods 4(11):903-5.
Aljanabi, 1997, Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques, Nucl. Acids Res 25:4692-4693.
Antonarakis and the Nomenclature Working Group, 1998, Recommendations for a nomenclature system for human gene mutations, Human Mutation 11:1-3.
Archer, 2014, Selective and flexible depletion of problematic sequences from RNA-seq libraries at the cDNA stage, BMC Genomics 15(1):401.
Ball, 2009, Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells, Nat Biotech 27:361-8.
Balzer, 2013, Filtering duplicate reads from 454 pyrosequencing data, Bioinformatics 29(7):830-836.
Barany, 1991, Genetic disease detection and DNA amplification using cloned thermostable ligase, PNAS 88:189-193.
Barany, 1991, The Ligase Chain Reaction in a PCR World, Genome Research 1:5-16.
Bau, 2008, Targeted next-generation sequencing by specific capture of multiple genomic loci using low-volume microfluidic DNA arrays, Analytical and Bioanal Chem 393(1):171-5.
Beer, 1962, Determination of base sequence in nucleic acids with the electron microscope: visibility of a marker, PNAS 48(3):409-416.
Bell, 2011, Carrier testing for severe childhood recessive diseases by next-generation sequencing, Sci Trans Med 3 (65ra4).
Benner, 2001, Evolution, language and analogy in functional genomics, Trends Genet 17:414-8.
Bentzley, 1996, Oligonucleotide sequence and composition determined by matrix-assisted laser desorption/ionization, Anal Chem 68:2141-2146.
Bentzley, 1998, Base specificity of oligonucleotide digestion by calf spleen phosphodiesterase with matrix-assisted aser desorption ionization analysis, Anal Biochem 258:31-37.
Bhangale, 2006, Automating resequencing-based detection of insertion-deletion polymorphisms, Nature Genetics 38:1457-1462.
Bickle, 1993, Biology of DNA Restriction, Microbiol Rev 57(2):434-50.
Bonfield, 2013, Compression of FASTQ and SAM format sequencing data, PLoS One 8(3):e59190.
Bose, 2012, BIND—An algorithm for loss-less compression of nucleotide sequence data, J Biosci 37(4):785-789.
Boyden, 2013, High-throughput screening for SMN1 copy number loss by next-generation sequencing, American Society of Human Genetics 63rd Annual Meeting, Abstract, Oct. 22, 2013.
Boyer, 1971, DNA restriction and modification mechanisms in bacteria, Ann Rev Microbiol 25:153-76.
Braasch, 2001, Locked nucleic acid (LNA): fine-tuning the recognition of DNA and RNA, Chemistry & Biology 8(1):1-7.
Braslavsky, 2003, Sequence information can be obtained from single DNA molecules, PNAS 100:3960-4.
Brinkman, 2004, Splice Variants as Cancer Biomarkers, Clin Biochem 37:584.
Brison, 1982, General method for cloning amplified DNA by differential screening, Mol Cell Biol 2(5):578-587.
Brown, 1979, Chemical synthesis and cloning of a tyrosine tRNA gene, Methods Enzymol 68:109-51.
Browne, 2002, Metal ion-catalyzed nucleic Acid alkylation and fragmentation, J Am Chem Soc 124(27):7950-7962.
Brownstein, 2014, An international effort towards developing standards for best practices in analysis, interpretation and reporting of clinical genome sequencing results in the CLARITY Challenge, Genome Biol 15:R53.
Bunyan, 2004, Dosage analysis of cancer predisposition genes by multiplex ligation-dependent probe amplification, British Journal of Cancer, 91(6):1155-59.
Burrow, 1994, A block-sorting lossless data compression algorithm, Technical Report 124, Digital Equipment Corporation, CA. (24 pages).
Carpenter, 2013, Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries, Am J Hum Genet 93(5):852-864.
Caruthers, 1985, Gene synthesis machines: DNA chemistry and its uses, Science 230:281-285.
Castellani, 2008, Consenses on the use of and interpretation of cystic fibrosis mutation analysis in clinical practice, J Cyst Fib 7:179-196.
Challis, 2012, An integrative variant analysis suite for whole exome next-generation sequencing data, BMC Informatics 13(8):1-12.
Chan, 2011, Natural and engineered nicking endonucleases-from cleavage mechanism to engineering of strand- specificity, Nucl Acids Res 39(1):1-18.
Chen, 2010, Identification of racehorse and sample contamination by novel 24-plex STR system, Forensic Sci Int: Genetics 4:158-167.
Chennagiri, 2013, A generalized scalable database model for storing and exploring genetic variations detected using sequencing data, American Society of Human Genetics 63rd Annual Meeting, Abstract, Oct. 22, 2013.
Chevreux, 1999, Genome sequence assembly using trace signals and additional sequence information, Proc GCB 99:45-56.
Chirgwin, 1979, Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease, Biochemistry, 18:5294-99.
Choe, 2010, Novel CFTR Mutations in a Korean Infant with Cystic Fibrosis and Pancreatic Insufficiency, J Korean Med Sci 25:163-5.
Ciotti, 2004, Triplet repeat prmied PCR (TP PCR) in molecular diagnostic testing for Friedrich ataxia, J Mol Diag 6 (4):285-9.
Cock, 2010, The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants, Nucleic Acids Res 38(6):1767-1771.
Collins, 2004, Finishing the euchromatic sequence of the human genome, Nature 431(7011):931-45.
Maddalena, 2005, Technical standards and guidelines: molecular genetic testing for ultra-rare disorders, Genet Med 7:571-83.
Malewicz, 2010, Pregel: a system for large-scale graph processing, Proc. ACM SIGMOD Int Conf Mgmt Data 135-46.
Mamanova, 2010, Target-enrichment strategies for next-generation sequencing, Nat Meth 7(2):111-118.
Margulies, 2005, Genome sequencing in micro-fabricated high-density picotiter reactors, Nature, 437:376-380.
Marras, 1999, Multiplex detection of single-nucleotide variations using molecular beacons, Genetic Analysis: Biomolecular Engineering 14:151.
Maxam, 1977, A new method for sequencing DNA, PNAS, 74:560-564.
May 1988, How Many Species Are There on Earth?, Science 241(4872):1441-9.
McDonnell, 2007, Antisepsis, disinfection, and sterilization: types, action, and resistance, p. 239.
McKenna, 2010, The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data, Genome Research 20:1297-1303.

(56) **References Cited**

OTHER PUBLICATIONS

Messiaen, 1999, Exon 10b of the NF1 gene represents a mutational hotspot and harbors a recurrent missense mutation Y489C associated with aberrant splicing, Genetics in Medicine, 1(6):248-253.
Meyer, 2007, Targeted high-throughput sequencing of tagged nucleic acid samples, Nucleic Acids Research 35(15): e97 (5 pages).
Meyer, 2008, Parallel tagged sequencing on the 454 platform, Nat Protocol 3(2):267-278.
Miesenbock, 1998, Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins, Nature 394(6689):192-95.
Miller, 2010, Assembly algorithms for next-generation sequencing data, Genomics 95:315-327.
Mills, 2010, Mapping copy number variation by population-scale genome sequencing, Nature 470(7332):59-65.
Miner, 2004, Molecular barcodes detect redundancy and contamination in hairpin-bisulfite PCR, Nucl Acids Res 32 (17):e135.
Minton, 2011, Mutation Surveyor: software for DNA sequence analysis, Meth Mol Biol 688:143-53.
Miyake, 2009, PIK3CA gene mutations and umplification in uterine cancers, Canc Lett 261:120-126.
Miyazaki, 2009, Characterization of deletion breakpoints in patients with dystrophinopathy carrying a deletion of exons 15-55 of the Duchenne muscular dystrophy (DMD) gene, J Hum Gen 54:127-30.
Mockler, 2005, Applications of DNA tiling arrays for whole-genome analysis, Genomics 85(1):1-15.
Mohammed, 2012, Deliminate—a fast and efficient methods for loss-less compression of genomice sequences, Bioinformatics 28(19):2527-2529.
Moudrianakis, 1965, Base Sequence Determination in Nucleic Acids with the Electron Microscope, III. Chemistry and Microscopy of Guanine-Labeled DNA, PNAS, 53:564-71.
Mullan, 2002, Multiple sequence alignment—the gateway to further analysis, Brief Bioinform 3(3):303-5.
Munne, 2012, Preimplantation genetic diagnosis for aneuploidy and translocations using array comparative genomic hybridization, Curr Genomics 13(6):463-470.
Nan, 2006, A novel CFTR mutation found in a Chinese patient with cystic fibrosis, Chinese Med J 119(2):103-9.
Narang, 1979, Improved phosphotriester method for the synthesis of gene fragments, Meth Enz 68:90-98.
Nelson, 1989, Bifunctional oligonucleotide probes synthesized using a novel CPG support are able to detect single base pair mutations, Nucl Acids Res 17(18):7187-7194.
Ng, 2009, Targeted capture and massively parallel sequencing of 12 human exomes, Nature 461(7261):272-6.
Nicholas, 2002, Strategies for multiple sequence alignment, Biotechniques 32:572-91.
Nickerson, 1990, Automated DNA diagnostics using an ELISA-based oligonucleotide ligation assay, PNAS 87:8923-7.
Nielsen, 1999, Peptide Nucleic Acids, Protocols and Applications (Norfolk: Horizon Scientific Press, 1-19).
Nilsson, 2006, Analyzing genes using closing and replicating circles, Trends in Biotechnology 24:83-8.
Ning, 2001, SSAHA: a fast search method for large DNA databases, Genome Res 11(10):1725-9.
Nordhoff, 1993, Ion stability of nucleic acids in infrared matrix-assisted laser desorption/ ionization mass spectrometry, Nucl Acid Res 21(15):3347-57.
Nuttle, 2013, Rapid and accurate large-scale genotyping of duplicated genes and discovery of interlocus gene conversions, Nat Meth 10(9):903-909.
Nuttle, 2014, Resolving genomic disorder-associated breakpoints within segmental DNA duplications using massively parallel sequencing, Nat Prot 9(6):1496-1513.
O'Roak, 2012, Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders, Science 338(6114):1619-1622.
Oefner, 1996, Efficient random sub-cloning of DNA sheared in a recirculating point-sink flow system, Nucleic Acids Res 24(20):3879-3886.
Oka, 2006, Detection of loss of heterozygosity in the p53 gene in renal cell carcinoma and bladder cancer using the polymerase chain reaction, Mol Carcinogenesis 4(1):10-13.
Okoniewski, 2013, Precise breakpoint localization of large genomic deletions using PacBio and Illumina next-generation sequencers, Biotechniques 54(2):98-100.
Okou, 2007, Microarray-based genomic selection for high-throughput reseugencing, Nat Meth 4(11):907-909.
Oliphant, 2002, BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping, Biotechniques Suppl:56-8, 60-1.
Ordahl, 1976, Sheared DNA fragment sizing: comparison of techniques, Nucleic Acids Res 3:2985-2999.
Ostrer, 2001, A genetic profile of contemporary Jewish populations, Nat Rev Genet 2(11):891-8.
Owens, 1998, Aspects of oligonucleotide and peptide sequencing with MALDI and electrospray mass spectrometry, Bioorg Med Chem 6:1547-1554.
Parameswaran, 2007, A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing, Nucl Acids Rese 35:e130.
Parkinson, 2012, Preparation of high-quality next-generation sequencing libraries from picogram quantities of target DNA, Genome Res 22:125-133.
Pastor, 2010, Conceptual modeling of human genome mutations: a dichotomy between what we have and what we shoudl have, 2010 Proc BIOSTEC Bioinformatics, pp. 160-166.
Paton, 2000, Conceptual modelling of genomic information, Bioinformatics 16(6):548-57.
Pearson, 1988, Improved tools for biological sequence comparison, PNAS 85(8):2444-8.
Holland, 2008, BioJava: an open-source framework for bioinformatics, Bioinformatics 24(18):2096-2097.
Homer, 2008, Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays. PLoS One 4(8):e1000167.
Homer, 2009, BFAST: An alignment tool for large scale genome resequencing, PLoS ONE 4(11):e7767.
Housley, 2009, SNP discovery and haplotype analysis in the segmentally duplicated DRD5 coding region, Ann Hum Genet 73(3):274-282.
Huang, 2008, Comparative analysis of common CFTR polymorphisms poly-T, TGrepeats and M470V in a healthy Chinese population, World J Gastroenterol 14(12):1925-30.
Husemann, 2009, Phylogenetic Comparative Assembly, Algorithms in Bioinformatics: 9th International Workshop, pp. 145-156, Salzberg & Warnow, Eds. Springer-Verlag, Berlin, Heidelberg.
Illumina, 2010, De Novo assembly using Illumina reads, Technical Note (8 pages).
International Human Genome Sequencing Consortium, 2004, Finishing the euchromatic sequence of the human genome, Nature 431:931-945.
Iqbal, 2012, De novo assembly and genotyping of variants using colored de Bruijn graphs, Nature Genetics 44:226-232.
Isosomppi, 2009, Disease-causing mutations in the CLRN1 gene alter normal CLRN1 protien trafficking to the plasma membrane, Mol Vis 15:1806-1818.
Jaijo, 2010, Microarray-based mutation analysis of 183 Spanish families with Usher syndrome, Invest Ophthalmol Vis Sci 51(3):1311-7.
Jensen, 2001, Orthologs and paralogs—we need to get it right, Genome Biol 2(8):1002-1002.3.
Jones, 2008, Core signaling pathways in human pancreatic cancers revealed by global genomic analyses, Science 321(5897):1801-1806.
Kambara, 1988, Optimization of Parameters in a DNA Sequenator Using Fluorescence Detection, Nature Biotechnology 6:816-821.
Kennedy, 2013, Accessing more human genetic variation with short sequencing reads, American Society of Human Genetics 63rd Annual Meeting, Abstract, Oct. 22, 2013.

(56) **References Cited**

OTHER PUBLICATIONS

Kent, 2002, BLAT—The BLAST-like alignment tool, Genome Res 12(4): 656-664.
Kerem, 1989, Identification of the cystic fibrosis gene: genetic analysis, Science 245:1073-1080.
Kinde, 2012, FAST-SeqS: a simple an effective method for detection of aneuploidy by massively parallel sequencing, PLoS One 7(7):e41162.
Kircher, 2010, High-througput DNA sequencing—concepts and limitations, Bioassays 32:524-36.
Kirpekar, 1994, Matrix assisted laser desorption/ionization mass spectrometry of enzymatically synthesized RNA up to 150 kDa, Nucl Acids Res 22:3866-3870.
Klein, 2011, LOCAS—A low coverage sequence assembly tool for re-sequencing projects, PLoS One 6(8):e23455.
Kneen, 1998, Green fluorescent protein as a noninvasive intracellular pH indicator, Biophys J 74(3):1591-99.
Koboldt, 2009, VarScan: variant detection in massively parallel sequencing of individual and pooled samples, Bioinformatics 25:2283-85.
Krawitz, 2010, Microindel detection in short-read sequence data, Bioinformatics 26(6):722-729.
Kreindler, 2010, Cystic fibrosis: exploiting its genetic basis in the hunt for new therapies, Pharmacol Ther 125 (2):219-229.
Krishnakumar, 2008, A comprehensive assay for targeted multiplex amplification of human DNA sequences, PNAS 105:9296-301.
Kumar, 2010, Comparing de novo assemblers for 454 transcriptome data, Genomics 11:571.
Kurtz, 2004, Versatile and open software for comparing large genomes, Genome Biol 5:R12.
Lam, 2008, Compressed indexing and local alignment of DNA, Bioinformatics 24(6):791-97.
Langmead, 2009, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, Genome Biol 10:R25.
Larkin, 2007, Clustal W and Clustal X version 2.0, Bioinformatics, 23(21):2947-2948.
Lecompte, 2001, Multiple alignment of complete sequences (MACS) in the post-genomic era, Gene 270(1-2):17-30.
Li, 2003, DNA binding and cleavage by the periplasmic nuclease Vvn: a novel structure with a known active site, EMBO J 22(15):4014-4025.
Li, 2008, SOAP: short oligonucleotide alignment program, Bioinformatics 24(5):713-14.
Li, 2009, Fast and accurate short read alignment with Burrows-Wheeler transform, Bioinformatics, 25 (14):1754-60.
Li, 2009, SOAP2: an improved ultrafast tool for short read alignment, Bioinformatics 25(15):1966-67.
Li, 2009, The Sequence Alignment/Map format and SAMtools, Bioinformatics 25(16):2078-9.
Li, 2010, Fast and accurate long-read alignment with Burrows-Wheeler transform, Bioinformatics 26(5):589-95.
Li, 2011, Improving SNP discovery by base alignment quality, Bioinformatics 27:1157.
Li, 2011, Single nucleotide polymorphism genotyping and point mutation detection by ligation on microarrays, J Nanosci Nanotechnol 11(2):994-1003.
Li, 2012, A new approach to detecting low-level mutations in next-generation sequence data, Genome Biol 13:1-15.
Li, 2014, HUGO: Hierarchical mUlti-reference Genome compression for aligned reads, JAMIA 21:363-373.
Lin, 2008, ZOOM! Zillions of Oligos Mapped, Bioinformatics, 24:2431.
Lin, 2010, A molecular inversion prove assay for detecting alternative splicing, BMC Genomics 11(712):1-14.
Lin, 2012, Development and evaluation of a reverse dot blot assay for the simultaneous detection of common alpha and beta thalassemia in Chinese, Blood Cells Molecules, and Diseases 48(2):86-90.
Lipman, 1985, Rapid and sensitive protein similarity searches, Science 227(4693):1435-41.
Liu, 2012, Comparison of next-generation sequencing systems, J Biomed Biotech 2012:251364.
Llopis, 1998, Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins, PNAS 95(12):6803-08.
Ma, 2006, Application of real-time polymerase chain reaction (RT-PCR), J Am Soc 1-15.
MacArthur, 2014, Guidelines for investigating causality of sequence variants in human disease, Nature 508:469-76.
Pertea, 2003, TIGR gene indices clustering tools (TGICL), Bioinformatics 19(5):651-52.
Pieles, 1993, Matrix-assisted laser desorption ionization time of-flight mass spectrometry: A powerful tool for the mass and sequence analysis of natural and modified oligonucleotides, Nucleic Acids Res 21:3191-3196.
Pinho, 2013, MFCompress: a compression tool for FASTA and multi-FASTA data, Bioinformatics 30(1):117-8.
Porreca, 2007, Multiplex amplificaiton of large sets of human exons, Nat Meth 4(11):931-936.
Porreca, 2013, Analytical performance of a Next-Generation DNA sequencing-based clinical workflow for genetic carrier screening, American Society of Human Genetics 63rd Annual Meeting, Abstract, Oct. 22, 2013.
Pourmand, 2006, PathgoenMIPer: a tool for the design of molecular inversion probes, BMC informatics 7:500.
Procter, 2006, Molecular diagnosis of Prader-Willi and Angelman syndromes by methylation-specific melting analysis and methylation-specific multiplex ligation-dependent probe amplification, Clin Chem 52(7):1276-83.
Qiagen, 2011, Gentra Puregene handbook, 3d Ed. (72 pages).
Quail, 2010, DNA: Mechanical Breakage, In Encyclopedia of Life Sciences, John Wiley & Sons Ltd, Chicester (5 pages).
Rambaut, 1997, Seq-Gen:an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees, Bioinformatics 13:235-38.
Richards, 2008 ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions, Genet Med 10(4):294-300.
Richter, 2008, MetaSim—A Sequencing Simulator for Genomics and Metagenomics, PLoS ONE 3:e3373.
Roberts, 1980, Restriction and modification enzymes and their recognition sequences, Nucleic Acids Res 8(1):r63-r80.
Rodriguez, 2010, Constructions from Dots and Lines, Bull Am Soc Inf Sci Tech 36(6):35-41.
Rosendahl, 2013, CFTR, SPINK1, CTRC and PRSS1 variants in chronic pancreatitis: is the role of mutated CFTR overestimated?, Gut 62:582-592.
Rothberg, 2011, An integrated semiconductor device enabling non-optical genome sequencing, Nature 475:348-352.
Rowntree, 2003, The phenotypic consequences of CFTR mutations, Ann Hum Gen 67:471-485.
Baihan, 2009, Update on Usher syndrome, Curr Op Neurology 22(1):19-24.
Sanger, 1977, DNA Sequencing with chain-terminating inhibitors, PNAS 74(12):5463-5467.
Santa Lucia, 1998, A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics, PNAS 95(4):1460-5.
Sargent, 1987, Isolation of differentially expressed genes, Meth Enzym 152:423-432.
Sauro, 2004, How Do You Calculate a Z-Score/ Sigma Level?, https://www.measuringusability.com/zcalc.htm (online publication).
Sauro, 2004, What's a Z-score and Why Use It in Usability Testing?, https://www.measuringusability.com/z.htm (online publication).
Schadt, 2010, A window into third-generation sequencing, Human Mol Genet 19(R2):R227-40.
Schatz, 2010, Assembly of large genomes using second-generation sequencing, Genome Res., 20:1165-1173.
Schiffman, 2009, Molecular inversion probes reveal patterns of 9p21 deletion and copy number aberrations in childhood leukemia, Cancer Genetics and Cytogenetics 193:9-18.

(56) **References Cited**

OTHER PUBLICATIONS

Schneeberger, 2011, Reference-guided assembly of four diverse *Arabidopsis thaliana* genomes, PNAS 108 (25):10249-10254.
Schouten, 2002, Relative Quantification of 40 Nucleic Acid Sequences by Multiplex Ligation-Dependent Probe Amplification, Nucle Acids Res 30 (12):257.
Schrijver, 2005, Diagnostic testing by CFTR gene mutation analysis in a large group of Hispanics, J Mol Diag 7 (2):289-299.
Schuette, 1995, Sequence analysis of phosphorothioate oligonucleotides via matrix-assisted laser desorption ionization time-of-flight mass spectrometry, J Pharm Biomed Anal 13:1195-1203.
Schwartz, 2009, Identification of cystic fibrosis variants by polymerase chain reaction/oligonucleotide ligation assay, J Mol Diag 11(3):211-15.
Schwartz, 2011, Clinical utility of single nucleotide polymorphism arrays, Clin Lab Med 31(4):581-94.
Sequeira, 1997, Implementing generic, object-oriented models in biology, Ecological Modeling 94.1:17-31.
Shagin, 2002, A novel method for SNP detection, Genome Res 12:1935.
Shen, 2011, High quality DNA sequence capture of 524 disease candidate genes, PNAS 108(16):6549-6554.
Shen, 2013, Multiplex capture with double-stranded DNA probes, Genome Medicine 5(50):1-8.
Sievers, 2011, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, Mol Syst Biol 7:539.
Simpson, 2009, ABySS: A parallel assembler for short read sequence data, Genome Res., 19(6):1117-23.
Slater, 2005, Automated generation of heuristics for biological sequence comparison, BMC Bioinformatics 6:31.
Smirnov, 1996, Sequencing oligonucleotides by exonuclease digestion and delayed extraction matrix-assisted laser desorption ionization time-of-flight mass spectrometry, Anal Biochem 238:19-25.
Smith, 1985, The synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis, Nucl Acid Res 13:2399-2412.
Smith, 2010, Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples, Nucleic Acids Research 38(13):e142 (8 pages).
Soni, 2007, Progress toward ultrafast DNA sequencing using solid-state nanopores, Clin Chem 53(11):1996-2001.
Spanu, 2010, Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism, Science 330(6010):1543-46.
Sproat, 1987, The synthesis of protected 5'-mercapto-2',5'-dideoxyribonucleoside-3'-O-phosphoramidites; uses of 5'-mercapto-oligodeoxyribonucleotides, Nucl Acid Res 15:4837-4848.
Streit, 2003, CFTR gene: molecular analysis in patients from South Brazil, Molecular Genetics and Metabolism 78:259-264.
Strom, 2005, Mutation detection, interpretation, and applications in the clinical laboratory setting, Mutat Res 673:160-67.
Summerer, 2009, Enabling technologies of genomic-scale sequence enrichment for targeted high-throughput sequencing, Genomics 94(6):363-8.
Summerer, 2010, Targeted High Throughput Sequencing of a Cancer-Related Exome Subset by Specific Sequence Capture With a Fully Automated Microarray Platform, Genomics 95(4):241-246.
Sunnucks, 1996, Microsatellite and chromosome evolution of parthenogenetic sitobion aphids in Australia, Genetics 144:747-756.

* cited by examiner

FIG. 1

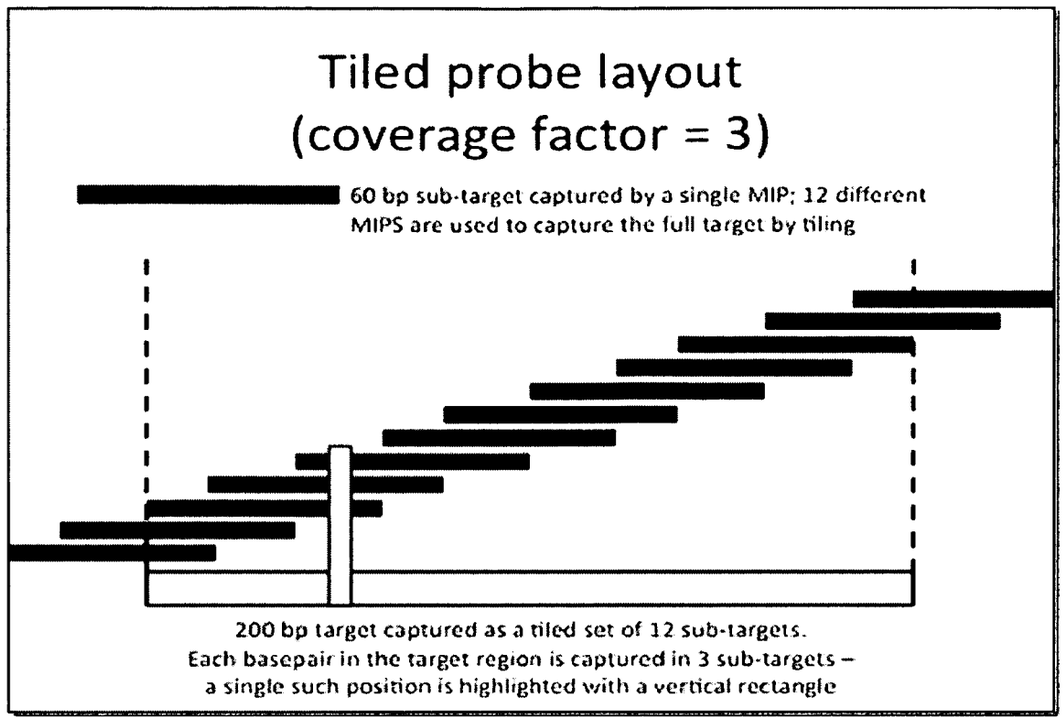

Tiled probe layout
(coverage factor = 3)
60 bp sub-target captured by a single MIP; 12 different MIPS are used to capture the full target by tiling
200 bp target captured as a tiled set of 12 sub-targets.
Each basepair in the target region is captured in 3 sub-targets –
a single such position is highlighted with a vertical rectangle

FIG. 2

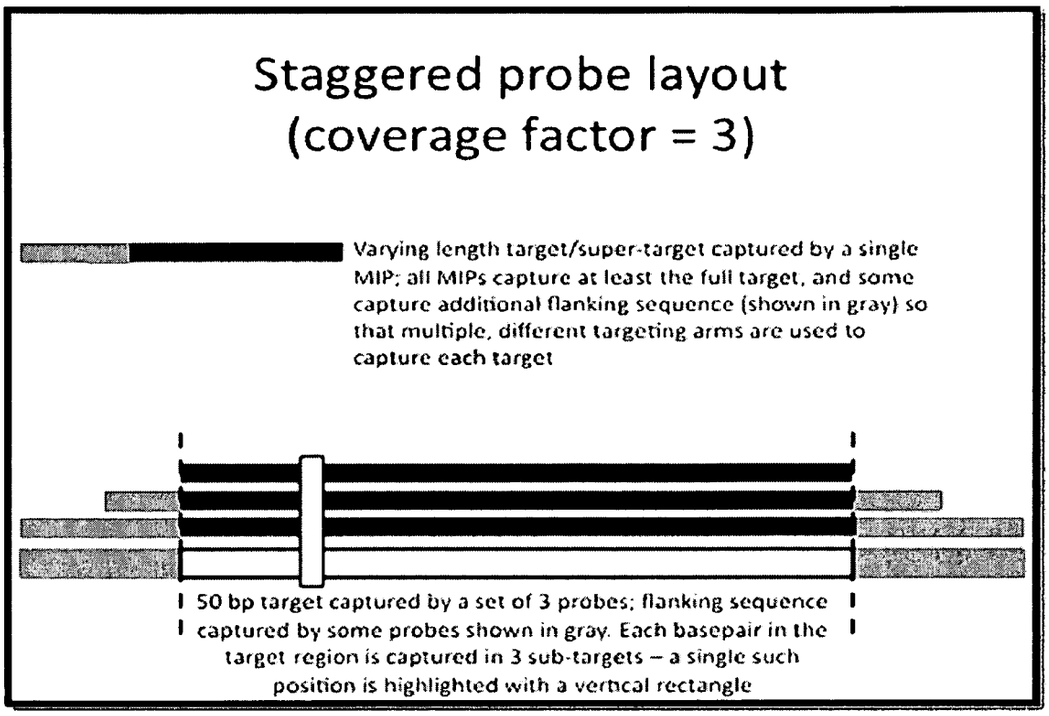

Staggered probe layout
(coverage factor = 3)
Varying length target/super-target captured by a single MIP; all MIPs capture at least the full target, and some capture additional flanking sequence (shown in gray) so that multiple, different targeting arms are used to capture each target
50 bp target captured by a set of 3 probes; flanking sequence captured by some probes shown in gray. Each basepair in the target region is captured in 3 sub-targets – a single such position is highlighted with a vertical rectangle

FIG. 3

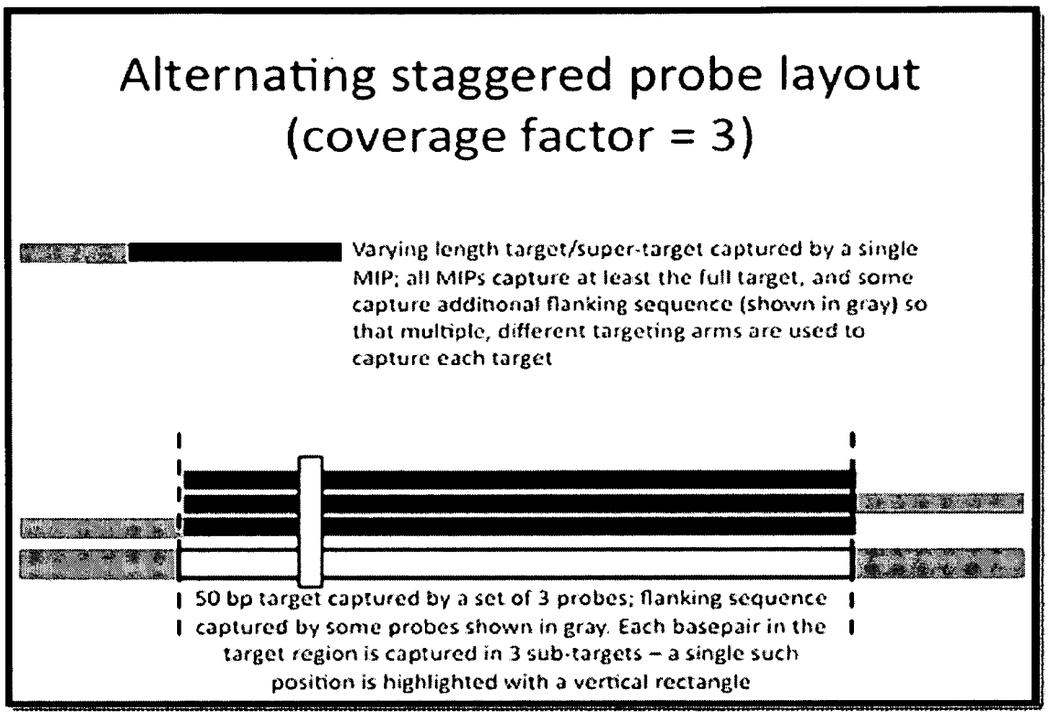

Alternating staggered probe layout
(coverage factor = 3)
Varying length target/super-target captured by a single MIP; all MIPs capture at least the full target, and some capture additional flanking sequence (shown in gray) so that multiple, different targeting arms are used to capture each target
50 bp target captured by a set of 3 probes; flanking sequence captured by some probes shown in gray. Each basepair in the target region is captured in 3 sub-targets – a single such position is highlighted with a vertical rectangle

FIG. 4A

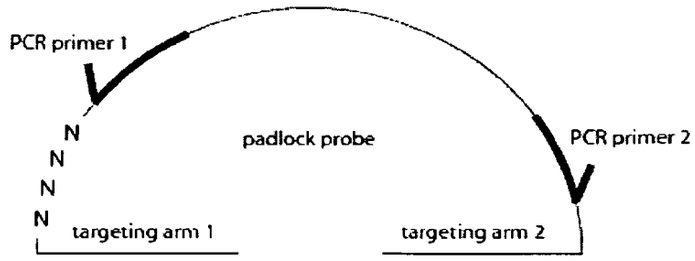


FIG. 4B

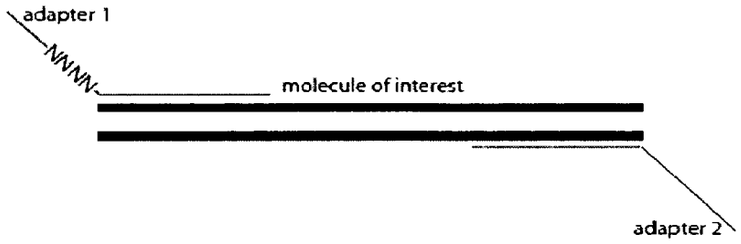


FIG. 4C

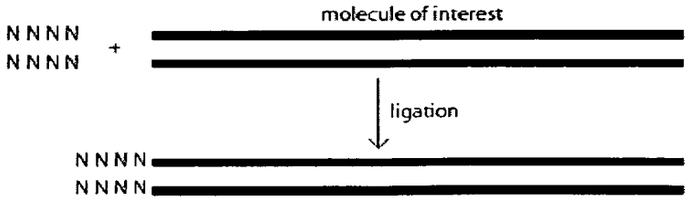


FIG. 5

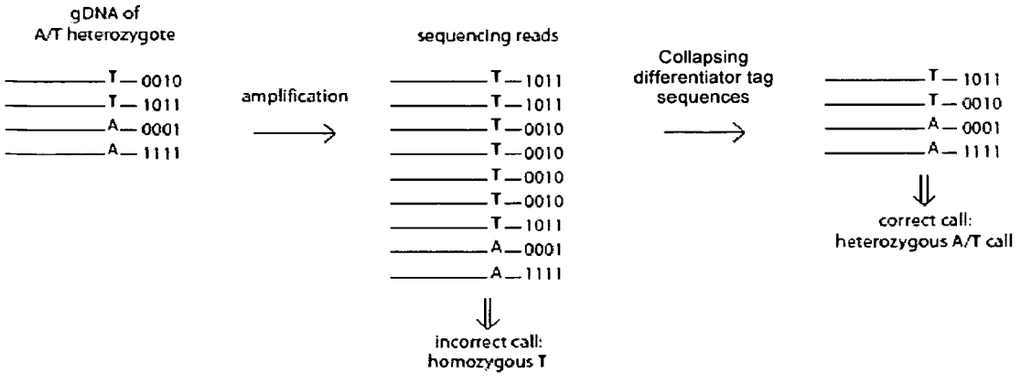

gDNA of
A/T heterozygote
T — 0010
T — 1011
A — 0001
A — 1111
amplification
sequencing reads
T — 1011
T — 1011
T — 0010
T — 0010
T — 0010
T — 0010
T — 1011
A — 0001
A — 1111
incorrect call:
homozygous T
Collapsing
differentiator tag
sequences
T — 1011
T — 0010
A — 0001
A — 1111
correct call:
heterozygous A/T call

FIG. 6

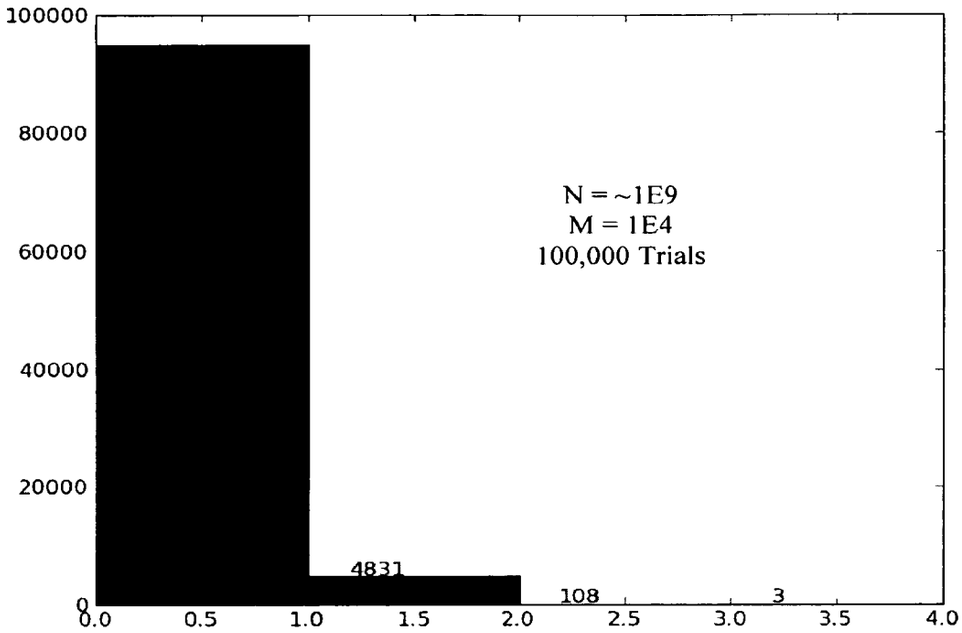

100000
80000
60000
40000
20000
0
N = ~1E9
M = 1E4
100,000 Trials
4831
108
3
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Independent Occurrences of
Target:Differentiator Tag Sequence
100,000 Trials

FIG. 7

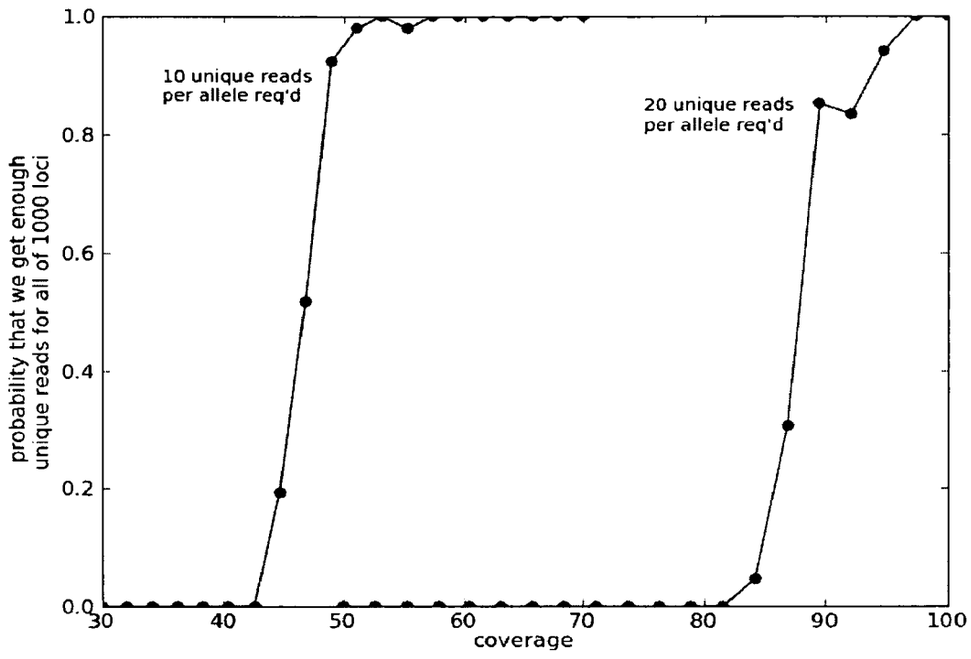

1.0
0.8
0.6
0.4
0.2
0.0
probability that we get enough unique reads for all of 1000 loci
10 unique reads per allele req'd
20 unique reads per allele req'd
30
40
50
60
70
80
90
100
coverage

FIG. 8

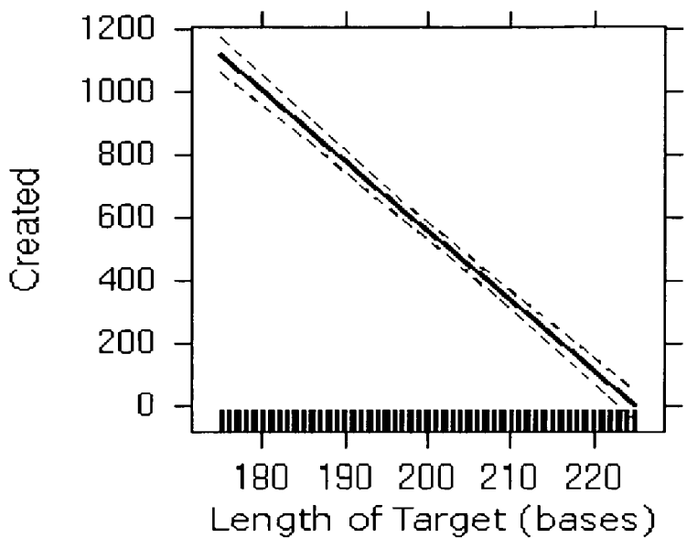

1200
1000
800
600
400
200
0
Created
180 190 200 210 220
Length of Target (bases)

FIG. 9

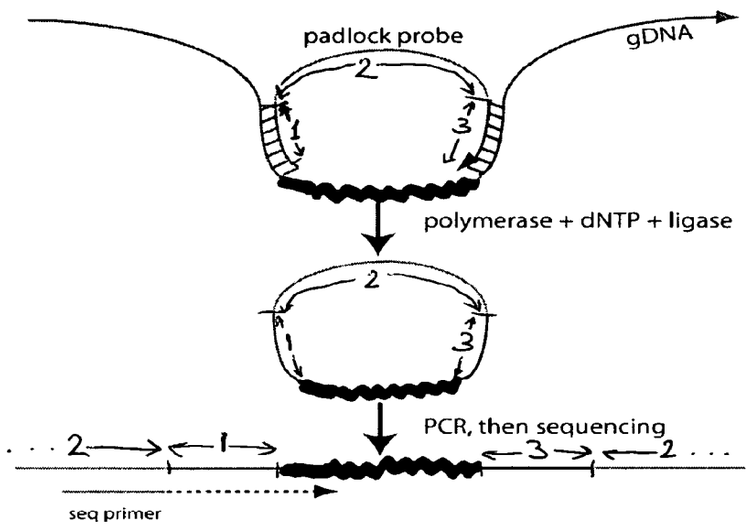

padlock probe
gDNA
2
1
3
polymerase + dNTP + ligase
2
1
3
PCR, then sequencing
. . . 2
1
3
2 . . .
seq primer

FIG. 10

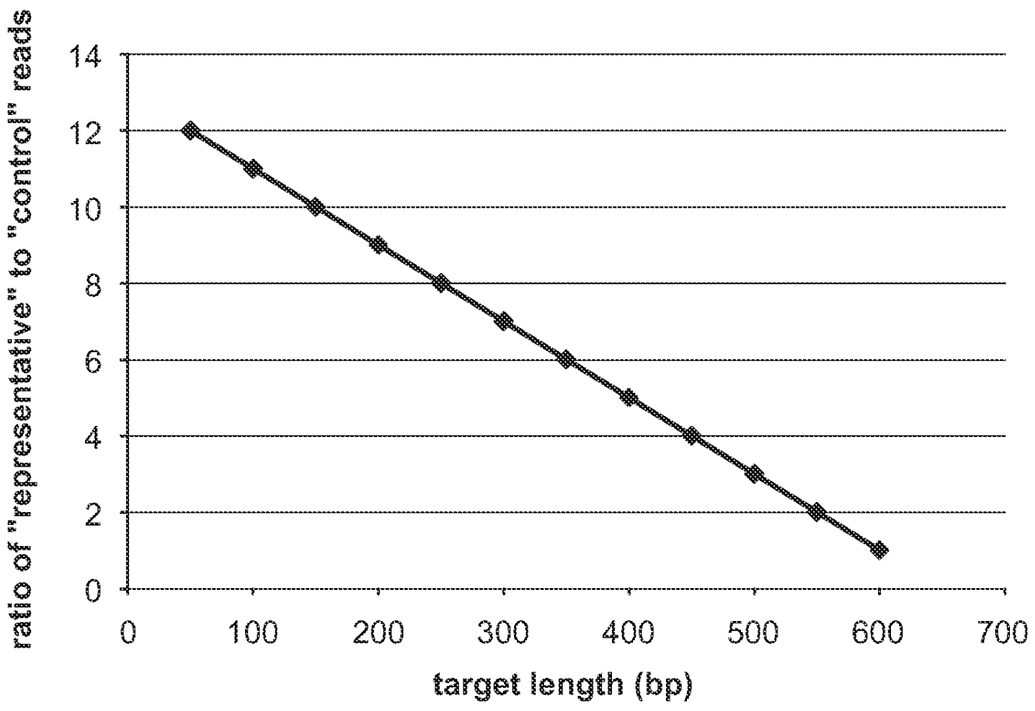

ratio of "representative" to "control" reads
14
12
10
8
6
4
2
0
0
100
200
300
400
500
600
700
target length (bp)

FIG. 11A

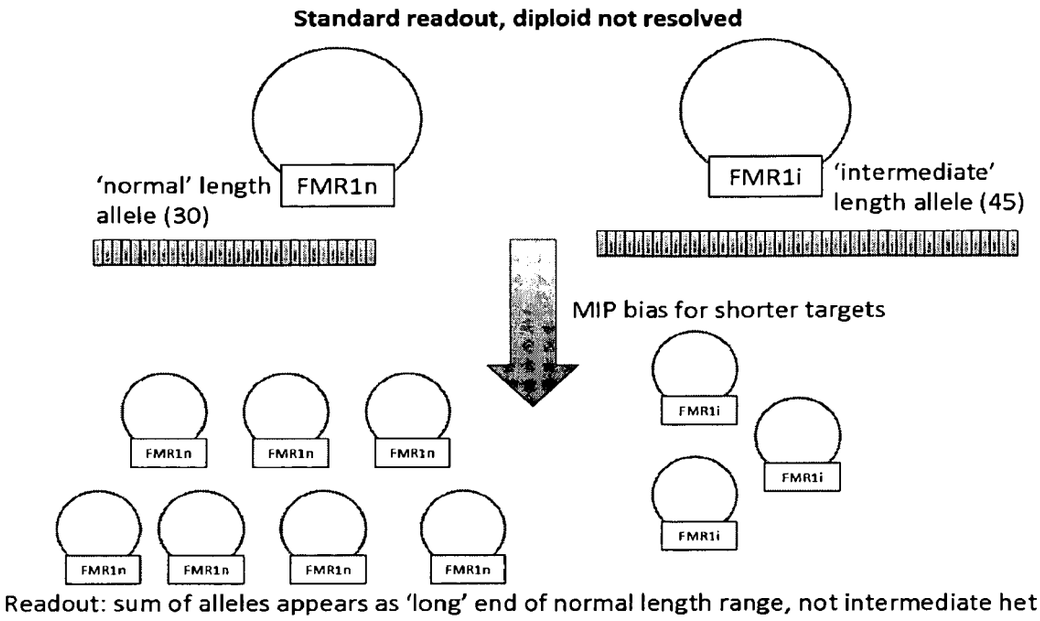

Standard readout, diploid not resolved
'normal' length allele (30)
FMR1n
FMR1i
'intermediate' length allele (45)
MIP bias for shorter targets
FMR1n
FMR1n
FMR1n
FMR1n
FMR1n
FMR1n
FMR1i
FMR1i
FMR1i
Readout: sum of alleles appears as 'long' end of normal length range, not intermediate het

FIG. 11B

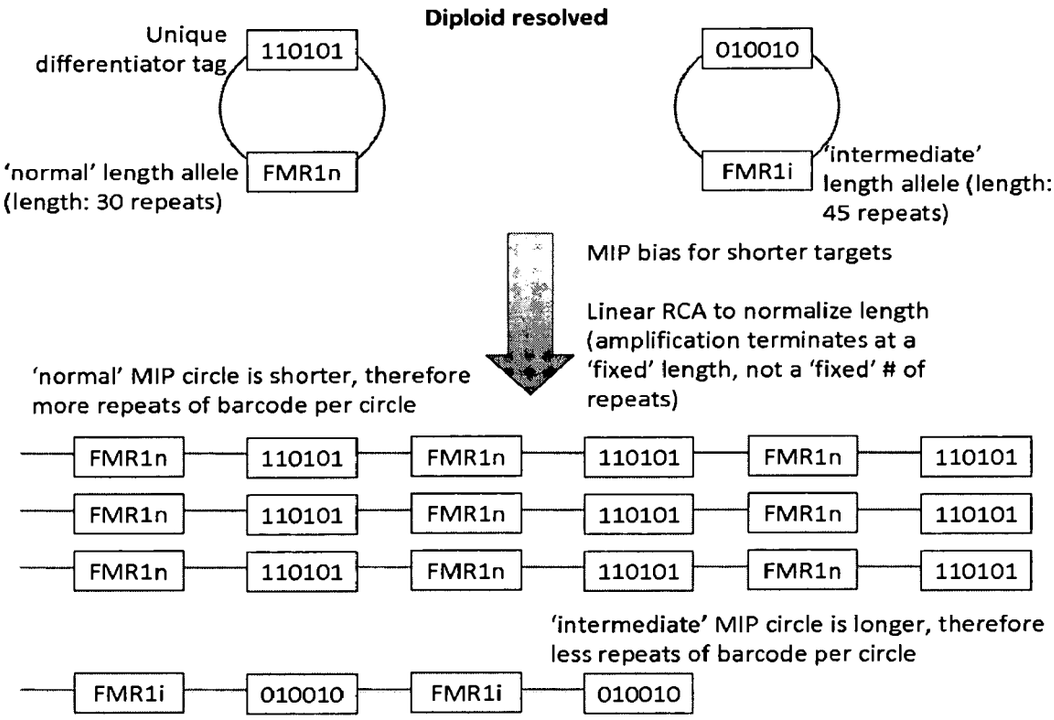

Diploid resolved
Unique differentiator tag
110101
FMR1n
'normal' length allele (length: 30 repeats)
010010
FMR1i
'intermediate' length allele (length: 45 repeats)
MIP bias for shorter targets
Linear RCA to normalize length (amplification terminates at a 'fixed' length, not a 'fixed' # of repeats)
'normal' MIP circle is shorter, therefore more repeats of barcode per circle
FMR1n 110101 FMR1n 110101 FMR1n 110101
FMR1n 110101 FMR1n 110101 FMR1n 110101
FMR1n 110101 FMR1n 110101 FMR1n 110101
'intermediate' MIP circle is longer, therefore less repeats of barcode per circle
FMR1i 010010 FMR1i 010010

FIG. 11C

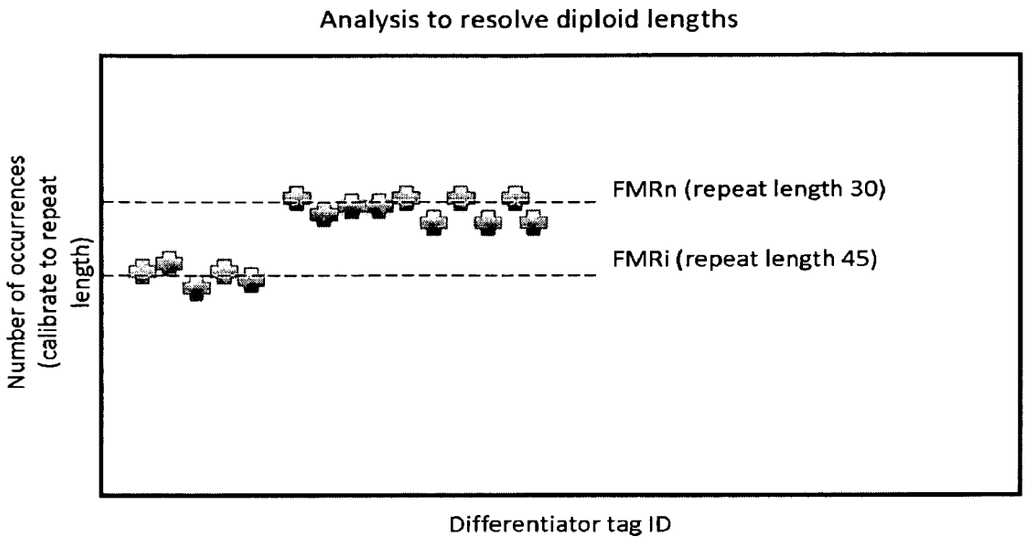

Analysis to resolve diploid lengths
Number of occurrences (calibrate to repeat length)
FMRn (repeat length 30)
FMRi (repeat length 45)
Differentiator tag ID

FIG. 12

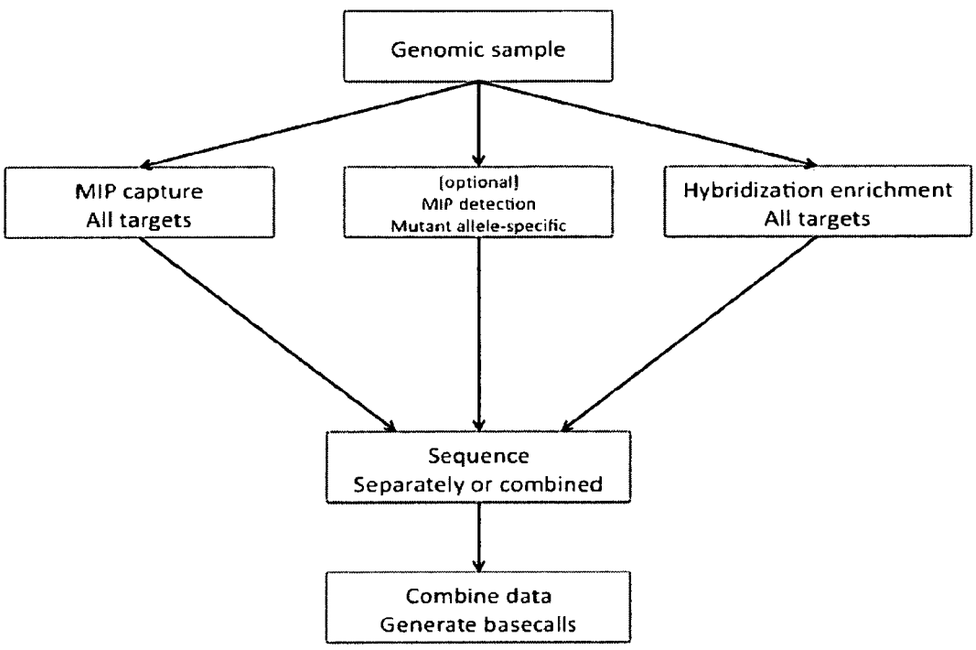

Genomic sample
MIP capture
All targets
(optional)
MIP detection
Mutant allele-specific
Hybridization enrichment
All targets
Sequence
Separately or combined
Combine data
Generate basecalls

FIG. 13

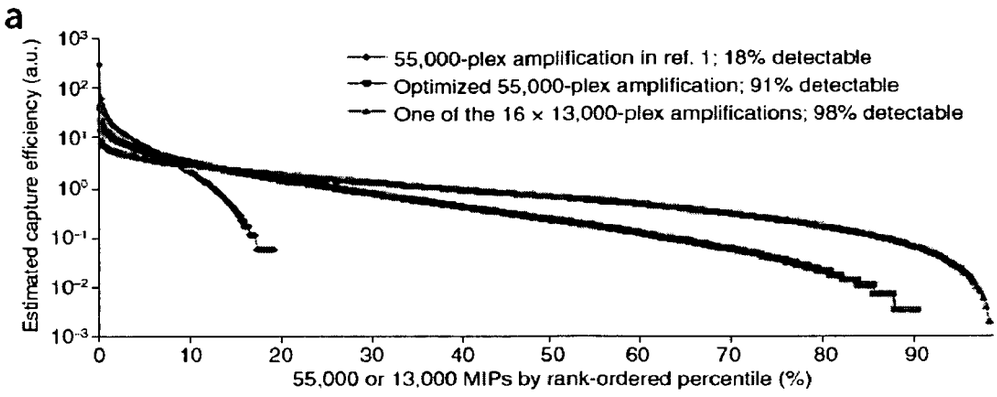
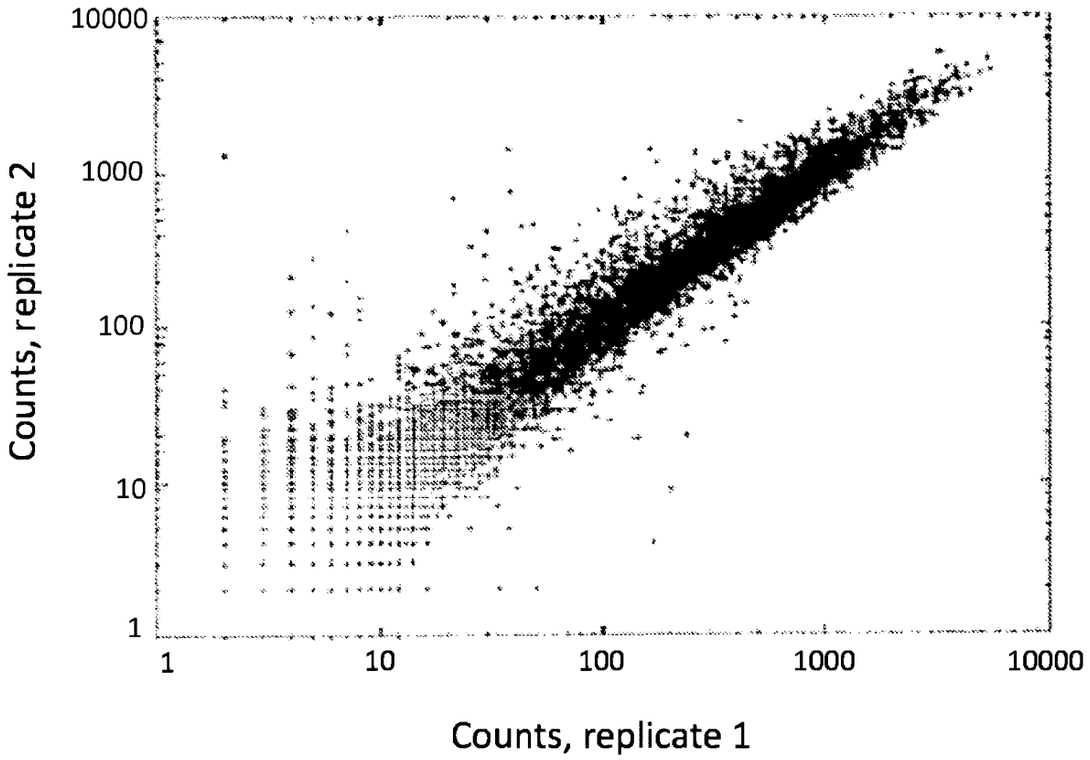

a
Estimated capture efficiency (a.u.)
10^3
10^2
10^1
10^0
10^{-1}
10^{-2}
10^{-3}
0
10
20
30
40
50
60
70
80
90
55,000 or 13,000 MIPs by rank-ordered percentile (%)
55,000-plex amplification in ref. 1; 18% detectable
Optimized 55,000-plex amplification; 91% detectable
One of the 16 × 13,000-plex amplifications; 98% detectable

FIG. 14


10000
1000
100
10
1
Counts, replicate 2
1
10
100
1000
10000
Counts, replicate 1

METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS

RELATED APPLICATIONS

This application is a continuation of U.S. application Ser. No. 15/231,687, filed Aug. 8, 2016, which is a continuation of U.S. application Ser. No. 13/266,862, filed Mar. 13, 2012, which is a national phase application that claims the benefit of, and priority to, international PCT Application No. PCT/US2010/001293, filed Apr. 30, 2010, which claims the benefit under 35 U.S.C. § 119(e) of U.S. Provisional Application No. 61/174,470, filed Apr. 30, 2009, U.S. Provisional Application No. 61/178,923, filed May 15, 2009, U.S. Provisional Application No. 61/179,358, filed May 18, 2009, and U.S. Provisional Application No. 61/182,089, filed May 28, 2009, the entire contents of each of which are incorporated herein by reference.

FIELD OF INVENTION

The invention relates to methods and compositions for determining genotypes in patient samples.

BACKGROUND OF THE INVENTION

Information about the genotype of a subject is becoming more important and relevant for a range of healthcare decisions as the genetic basis for many diseases, disorders, and physiological characteristics is further elucidated. Medical advice is increasingly personalized, with individual decisions and recommendations being based on specific genetic information. Information about the type and number of alleles at one or more genetic loci impacts disease risk, prognosis, therapeutic options, and genetic counseling amongst other healthcare considerations.

For cost-effective and reliable medical and reproductive counseling on a large scale, it is important to be able correctly and unambiguously identify the allelic status for many different genetic loci in many subjects.

Numerous technologies have been developed for detecting and analyzing nucleic acid sequences from biological samples. These technologies can be used to genotype subjects and determine the allelic status of any locus of interest. However, they are not sufficiently robust and cost-effective to be scaled up for reliable high throughput analysis of many genetic loci in large numbers of patients. The frequency of incorrect or ambiguous calls is too high for current technology to manage large numbers of patient samples without involving expensive and time-consuming steps to resolve uncertainties and provide confidence in the information output.

SUMMARY OF THE INVENTION

Aspects of the invention relate to preparative and analytical methods and compositions for evaluating genotypes, and in particular, for determining the allelic identity (or identities in a diploid organism) of one or more genetic loci in a subject.

Aspects of the invention are based, in part, on the identification of different sources of ambiguity and error in genetic analyses, and, in part, on the identification of one or more approaches to avoid, reduce, recognize, and/or resolve these errors and ambiguities at different stages in a genetic analysis.

According to aspects of the invention, certain types of genetic information can be under-represented or over-represented in a genetic analysis due to a combination of stochastic variation and systematic bias in any of the preparative stages (e.g., capture, amplification, etc.), determining stages (e.g., allele-specific detection, sequencing, etc.), data interpretation stages (e.g., determining whether the assay information is sufficient to identify a subject as homozygous or heterozygous), and/or other stages.

According to aspects of the invention, error or ambiguity may be apparent in a genetic analysis, but not readily resolved without running additional samples or more expensive assays (e.g., array-based assays may report no-calls due to noisy/low signal). According to further aspects of the invention, error or ambiguity may not be accounted for in a genetic analysis and incorrect base calls may be made even when the evidence for them is limited and/or not statistically significant (e.g., next-generation sequencing technologies may report base calls even if the evidence for them is not statistically significant). According to further aspects of the invention error or ambiguity may be problematic for a multi-step genetic analysis because it is apparent but not readily resolved in one or more steps of the analysis and not apparent or accounted for in other steps of the analysis.

In some embodiments, sources of error and ambiguity in one or more steps can be addressed by capturing and/or interrogating each target locus of interest with one or more sets of overlapping probes that are designed to overcome any systematic bias or stochastic effects that may impact the complexity and/or fidelity of the genetic information that is generated.

In some embodiments, sources of error and ambiguity in one or more steps can be addressed by capturing and/or interrogating each target locus of interest with at least one set of probes, wherein different probes are labeled with different identifiers that can be used to track the assay reactions and determine whether certain types of genetic information are under-represented or over-represented in the information that is generated.

In some embodiments, errors and ambiguities associated with the analysis of regions containing large numbers of sequence repeats are addressed by systematically analyzing frequencies of certain nucleic acids at particular stages in an assay (e.g., at a capture, sequencing, or detection stage). It should be appreciated that such techniques may be particularly useful in the context of a standardized protocol that is designed to allow many different loci to be evaluated in parallel without requiring different assay procedures for each locus. In some embodiments, the use of a single detection modality (e.g., sequencing) to assay multiple types of genetic lesions (e.g., point mutations, insertions/deletions, length polymorphisms) is advantageous in the clinical setting. In some embodiments of the invention, methods are provided that facilitate the use of multiple sample preparation steps in parallel, coupled with multiple analytical processes following sequence detection. Thus, in some embodiments of the invention, an improved workflow is provided that reduces error and uncertainty when simultaneously assaying different types of genetic lesions across multiple loci in multiple patients.

In some embodiments, aspects of the invention provide methods for overcoming preparative and/or analytical bias by combining two or more techniques, each having a different bias (e.g., a known bias towards under-representation or over-representation of one or more types of sequences), and using the resulting data to determine a genetic call for a subject with greater confidence.

It should be appreciated that in some embodiments, aspects of the invention relate to multiplex diagnostic methods. In some embodiments, multiplex diagnostic methods comprise capturing a plurality of genetic loci in parallel (e.g., one or more genetic loci from Table 1). In some embodiments, the genetic loci possess one or more polymorphisms (e.g., one or more polymorphisms from Table 2) the genotypes of which correspond to disease causing alleles. Accordingly, in some embodiments, the disclosure provides methods for assessing multiple heritable disorders in parallel. In some embodiments, methods are provided for diagnosing multiple heritable disorders in parallel at a pre-implantation, prenatal, perinatal, or postnatal stage. In some embodiments, the disclosure provides methods for analyzing multiple genetic loci (e.g., a plurality of target nucleic acids selected from Table 1) from a patient sample, such as a blood, pre-implantation embryo, chorionic villus or amniotic fluid sample, or other sample (e.g., other biological fluid or tissue sample such as a biopsy sample) as aspects of the invention are not limited in this respect.

Other samples may include tumor tissue or circulating tumor cells. In some embodiments, a patient sample (e.g., a tumor tissue or cell sample) is mosaic for one or more mutations of interest, and thus, may require higher sensitivity than is needed for a germline mutation analysis. In some embodiments, a sample comprises cells from a non-host organism (e.g., bacterial or viral infections in a human subject) or a sample for environmental monitoring (e.g., bacterial, viral, fungal composition of a soil, water, or air sample).

Accordingly, in some embodiments, aspects of the methods disclosed herein relate to genotyping a polymorphism of a target nucleic acid. In some embodiments, the genotyping may comprise determining that one or more alleles of the target nucleic acid are heterozygous or homozygous. In further embodiments, the genotyping may comprise determining the sequence of a polymorphism and comparing that sequence to a control sequence that is indicative of a disease risk. In some embodiments, the polymorphism is selected from a locus in Table 1 or Table 2. However, it should be appreciated that any locus associated with a disease or condition of interest may be used.

In some embodiments, a diagnosis, prognosis, or disease risk assessment is provided to a subject based on a genotype determined for that subject at one or more genetic loci (e.g., based on the analysis of a biological sample obtained from that subject). In some embodiments, an assessment is provided to a couple, based on their respective genotypes at one or more genetic loci, of the risk of their having one or more children having a genotype associated with a disease or condition (e.g., a homozygous or heterozygous genotype associated with a disease or condition). In some embodiments, a subject or a couple may seek genetic or reproductive counseling in connection with a genotype determined according to embodiments of the invention. In some embodiments, genetic information from a tumor or circulating tumor cells is used to determine prognosis and guide selection of appropriate drugs/treatments.

It should be appreciated that any of the methods or compositions described herein may be used in combination with any of the medical evaluations associated with one or more genetic loci as described herein.

In some embodiments, aspects of the invention provide effective methods for overcoming challenges associated with systematic errors (bias) and/or stochastic effects in multiplex genomic capture and/or analysis (including sequencing analysis). In some embodiments, aspects of the invention are useful to avoid, reduce and/or account for variability in one or more sampling and/or analytical steps. For example, in some embodiments, variability in target nucleic acid representation and unequal sampling of heterozygous alleles in pools of captured target nucleic acids can be overcome.

Accordingly, in some embodiments, the disclosure provides methods that reduce variability in the detection of target nucleic acids in multiplex capture methods. In other embodiments, methods improve allelic representation in a capture pool and, thus, improve variant detection outcomes. In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve the use of different sets of multiple probes (e.g., molecular inversion probes MIPs) that capture overlapping regions of a target nucleic acid to achieve a more uniform representation of the target nucleic acids in a capture pool compared with methods of the prior art. In other embodiments, methods reduce bias, or the risk of bias, associated with large scale parallel capture of genetic loci, e.g., for diagnostic purposes. In other embodiments, methods are provided for increasing reproducibility (e.g., by reducing the effect of polymorphisms on target nucleic acid capture) in the detection of a plurality of genetic loci in parallel. In further embodiments, methods are provided for reducing the effect of probe synthesis and/or probe amplification variability on the analysis of a plurality of genetic loci in parallel.

According to some aspects, methods of analyzing a plurality of genetic loci are provided. In some embodiments, the methods comprise contacting each of a plurality of target nucleic acids with a probe set, wherein each probe set comprises a plurality of different probes, each probe having a central region flanked by a 5' region and a 3' region that are complementary to nucleic acids flanking the same strand of one of a plurality of subregions of the target nucleic acid, wherein the subregions of the target nucleic acid are different, and wherein each subregion overlaps with at least one other subregion, isolating a plurality of nucleic acids each having a nucleic acid sequence of a different subregion for each of the plurality of target nucleic acids, and analyzing the isolated nucleic acids.

In other embodiments, methods comprise contacting each of a plurality of target nucleic acids with a probe set, wherein each probe set comprises a plurality of different probes, each probe having a central region flanked by a 5' region and a 3' region that are complementary to nucleic acids flanking the same strand of one of a plurality of subregions of the target nucleic acid, wherein the subregions of the target nucleic acid are different, and wherein a portion of the 5' region and a portion of the 3' region of a probe have, respectively, the sequence of the 5' region and the sequence of the 3' region of a different probe, isolating a plurality of nucleic acids each having a nucleic acid sequence of a different subregion for each of the plurality of target nucleic acids, and analyzing the isolated nucleic acids.

Aspects of the disclosure are based, in part, on the discovery of methods for overcoming problems associated with systematic and random errors (bias) in genome capture, amplification and sequencing methods, namely high variability in the capture and amplification of nucleic acids and disproportionate representation of heterozygous alleles in sequencing libraries. Accordingly, in some embodiments, the disclosure provides methods that reduce errors associated with the variability in the capture and amplification of nucleic acids. In other embodiments, the methods improve allelic representation in sequencing libraries and, thus,

improve variant detection outcomes. In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve the use of differentiator tag sequences to uniquely tag individual nucleic acid molecules. In some embodiments, the differentiator tag sequence permit the detection of bias based on the occurrence of combinations of differentiator tag and target sequences observed in a sequencing reaction. In other embodiments, the methods reduce errors caused by bias, or the risk of bias, associated with the capture, amplification and sequencing of genetic loci, e.g., for diagnostic purposes.

Aspects of the invention relate to providing sequence tags (referred to as differentiator tags) that are useful to determine whether target nucleic acid sequences identified in an assay are from independently isolated target nucleic acids or from multiple copies of the same target nucleic acid molecule (e.g., due to bias in a preparative step, for example, amplification). This information can be used to help analyze a threshold number of independently isolated target nucleic acids from a biological sample in order to obtain sequence information that is reliable and can be used to make a genotype conclusion (e.g., call) with a desired degree of confidence. This information also can be used to detect bias in one or more nucleic acid preparative steps.

In some embodiments, the methods disclosed herein are useful for any application where reduction of bias, e.g., associated with genomic isolation, amplification, sequencing, is important. For example, detection of cancer mutations in a heterogeneous tissue sample, detection of mutations in maternally-circulating fetal DNA, and detection of mutations in cells isolated during a preimplantation genetic diagnostic procedure.

Accordingly, in some aspects, methods of genotyping a subject are provided. In some embodiments, the methods comprise determining the sequence of at least a threshold number of independently isolated nucleic acids, wherein the sequence of each isolated nucleic acid comprises a target nucleic acid sequence and a differentiator tag sequence, wherein the threshold number is a number of unique combinations of target nucleic acid and differentiator tag sequences, wherein the isolated nucleic acids are identified as independently isolated if they comprise unique combinations of target nucleic acid and differentiator tag sequences, and wherein the target nucleic acid sequence is the sequence of a genomic locus of a subject.

In some embodiments, the isolated nucleic acids are products of a circularization selection-based preparative method, e.g., molecular inversion probe capture products. In other embodiments, the isolated nucleic acids are products of an amplification-based preparative methods. In other embodiments, the isolated nucleic acids are products of hybridization-based preparative methods.

Circularization selection-based preparative methods selectively convert regions of interest (target nucleic acids) into a covalently-closed circular molecule which is then isolated typically by removal (usually enzymatic, e.g. with exonuclease) of any non-circularized linear nucleic acid. Oligonucleotide probes (e.g., molecular inversion probes) are designed which have ends that flank the region of interest (target nucleic acid) and, optionally, primer sites, e.g., sequencing primer sites. The probes are allowed to hybridize to the genomic target, and enzymes are used to first (optionally) fill in any gap between probe ends and second ligate the probe closed. Following circularization, any remaining (non-target) linear nucleic acid is typically removed, resulting in isolation (capture) of target nucleic acid. Circularization selection-based preparative methods include molecular inversion probe capture reactions and 'selector' capture reactions. In some embodiments, molecular inversion probe capture of a target nucleic acid is indicative of the presence of a polymorphism in the target nucleic acid.

In amplification-based (e.g., PCR-based or LCR-based, etc.) preparative methods, genomic loci (target nucleic acids) are isolated directly by means of a polymerase chain reaction or ligase chain reaction (or other amplification method) that selectively amplifies each locus using one or more oligonucleotide primers. It is to be understood that primers will be sufficiently complementary to the target sequence to hybridize with and prime amplification of the target nucleic acid. Any one of a variety of art known methods may be utilized for primer design and synthesis. One or more of the primers may be perfectly complementary to the target sequence. Degenerate primers may also be used. Primers may also include additional nucleic acids that are not complementary to target sequences but that facilitate downstream applications, including for example restriction sites and differentiator tag sequences. Amplification-based methods include amplification of a single target nucleic acid and multiplex amplification (amplification of multiple target nucleic acids in parallel).

Hybridization-based preparative methods involve selectively immobilizing target nucleic acids for further manipulation. It is to be understood that one or more oligonucleotides (immobilization oligonucleotides), which comprise differentiator tag sequences, and which may be from 15 to 170 nucleotides in length, are used which hybridize along the length of a target region of a genetic locus to immobilize it. In some embodiments, immobilization oligonucleotides, are either immobilized before hybridization is performed (e.g., Roche/Nimblegen 'sequence capture'), or are prepared such that they include a moiety (e.g. biotin) which can be used to selectively immobilize the target nucleic acid after hybridization by binding to e.g., streptavidin-coated microbeads (e.g. Agilent 'SureSelect').

It should be appreciated that any of the circularization, amplification, and/or hybridization based methods described herein may be used in connection with one or more of the tiling/staggering, tagging, size-detection, and/or sensitivity enhancing algorithms described herein.

In some embodiments, the methods disclosed herein comprise determining the sequence of molecular inversion probe capture products, each comprising a molecular inversion probe and a target nucleic acid, wherein the sequence of the molecular inversion probe comprises a differentiator tag sequence and, optionally, a primer sequence, and wherein the target nucleic acid is a captured genomic locus of a subject, and genotyping the subject at the captured genomic locus based on the sequence of at least a threshold number of unique combinations of target nucleic acid and differentiator tag sequences of molecular inversion probe capture products.

In some embodiments, the methods disclosed herein comprise obtaining molecular inversion probe capture products, each comprising a molecular inversion probe and a target nucleic acid, wherein the sequence of the molecular inversion probe comprises a differentiator tag sequence and, optionally, a primer sequence, wherein the target nucleic acid is a captured genomic locus of the subject, amplifying the molecular inversion probe capture products, and genotyping the subject by determining, for each target nucleic acid, the sequence of at least a threshold number of unique combinations of target nucleic acid and differentiator tag sequence of molecular inversion probe capture products. In certain embodiments, obtaining comprises capturing target

nucleic acids from a genomic sample of the subject with molecular inversion probes, each comprising a unique differentiator tag sequence. In specific embodiments, capturing is performed under conditions wherein the likelihood of obtaining two or more molecular inversion probe capture products with identical combinations of target and differentiator tag sequences is equal to or less than a predetermined value, optionally wherein the predetermined value is about 0.05.

In one embodiment, the threshold number for a specific target nucleic acid sequence is selected based on a desired statistical confidence for the genotype. In some embodiments, the methods further comprising determining a statistical confidence for the genotype based on the number of unique combinations of target nucleic acid and differentiator tag sequences.

According to some aspects, methods of analyzing a plurality of genetic loci are provided. In some embodiments, the methods comprise obtaining a plurality of molecular inversion probe capture products each comprising a molecular inversion probe and a target nucleic acid, wherein the sequence of the molecular inversion probe comprises a differentiator tag sequence and, optionally, a primer sequence (e.g., a sequence that is complementary to the sequence of a nucleic acid that is used as a primer for sequencing or other extension reaction), amplifying the plurality of molecular inversion probe capture products, determining numbers of occurrence of combinations of target nucleic acid and differentiator tag sequence of molecular inversion probe capture products in the amplified plurality, and if the number of occurrence of a specific combination of target nucleic acid sequence and differentiator tag sequence exceeds a predetermined value, detecting bias in the amplification of the molecular inversion probe comprising the specific combination. In some embodiments, the methods further comprise genotyping target sequences in the plurality, wherein the genotyping comprises correcting for bias, if detected.

In some embodiments, the target nucleic acid is a gene (or portion thereof) selected from Table 1. In some embodiments, the genotyping comprises determining the sequence of a target nucleic acid (e.g., a polymorphic sequence) at one or more (both) alleles of a genome (a diploid genome) of a subject. In certain embodiments, the genotyping comprises determining the sequence of a target nucleic acid at both alleles of a diploid genome of a subject, wherein in the target nucleic acid comprises, or consists of, a sequence of Table 1, Table 2, or other locus of interest.

In some embodiments, aspects of the invention provide methods and compositions for identifying nucleic acid insertions or deletions in genomic regions of interest without determining the nucleotide sequences of these regions. Aspects of the invention are particularly useful for detecting nucleic acid insertions or deletions in genomic regions containing nucleic acid sequence repeats (e.g., di- or tri-nucleotide repeats). However, the invention is not limited to analyzing nucleic acid repeats and may be used to detect insertions or deletions in any target nucleic acid of interest. Aspects of the invention are particularly useful for analyzing multiple loci in a multiplex assay.

In some embodiments, aspects of the invention relate to determining whether an amount of target nucleic acid that is captured in a genomic capture assay is higher or lower than expected. In some embodiments, a statistically significant deviation from an expected amount (e.g., higher or lower) is indicative of the presence of a nucleic acid insertion or deletion in the genomic region of interest. In some embodiments, the amount is a number of nucleic acid molecules that are captured. In some embodiments, the amount is a number of independently captured nucleic acid molecules in a sample. It should be appreciated that the captured nucleic acids may be literally captured from a sample, or their sequences may be captured without actually capturing the original nucleic acids in the sample. For example, nucleic acid sequences may be captured in an assay that involves a template-based extension of nucleic acids having the region of interest, in the sample.

Aspects of the invention are based on the recognition that the efficiency of certain capture techniques is affected by the length of the nucleic acid being captured.

Accordingly, an increase or decrease in the length of a target nucleic acid (e.g., due to an insertion or deletion of a repeated sequence) can alter the capture efficiency of that nucleic acid. In some embodiments, a difference in the capture efficiency (e.g., a statistically significant difference in the capture efficiency) of a target nucleic acid is indicative of an insertion or deletion in the target nucleic acid. It should be appreciated that the capture efficiency for a target nucleic acid may be evaluated based on an amount of captured nucleic acid (e.g., number of captured nucleic acid molecules) relative to a control amount (e.g., based on an amount of control nucleic acid that is captured). However, the invention is not limited in this respect and other techniques for evaluating capture efficiency also may be used.

According to aspects of the invention, evaluating the capture efficiency as opposed to determining the sequence of the entire repeat region reduces errors associated with sequencing through repeat regions. Repeat sequences often give rise to stutters or skips in sequencing reactions that make it very difficult to accurately determine the number of repeats in a target region without running multiple sequencing reactions under different conditions and carefully analyzing the results. Such procedures are cumbersome and not readily scalable in a manner that is consistent with high throughput analyses of target nucleic acids. In some embodiments, repeat regions may be longer than the length of the individual sequence read, making length determination on the basis of a single read impossible. For example, when using next-generation sequencing the repeat regions may be longer than the length of the individual sequence read, making length determination on the basis of a single read impossible. Accordingly, aspects of the invention are useful to increase the sensitivity of detecting insertions or deletions in target regions, particularly target regions containing repeated sequences.

In some embodiments, aspects of the invention relate to capturing genomic nucleic acid sequences using a molecular inversion probe (e.g., MIP or Padlock probe) technique, and determining whether the amount (e.g., number) of captured sequences is higher or lower than expected. In some embodiments, the amount (e.g., number) of captured sequences is compared to an amount (e.g., number) of sequences captured in a control assay. The control assay may involve analyzing a control sample that contains a nucleic acid from the same genetic locus having a known sequence length (e.g., a known number of nucleic acid repeats). However, a control may involve analyzing a second (e.g., different) genetic locus that is not expected to contain any insertions or deletions. The second genetic locus may be analyzed in the same sample as the locus being interrogated or in a different sample where its length has been previously determined. The second genetic locus may be a locus that is not char-

acterized by the presence of nucleic acid repeats (and thus not expected to contain insertions or deletions of the repeat sequence).

In some embodiments, a target nucleic acid region that is being evaluated may be determined by the identity of the targeting arms of a probe that is designed to capture the target region (or sequence thereof). For example, the targeting arms of a MIP probe may be designed to be complementary (e.g., sufficiently complementary for selective hybridization and//or polymerase extension and/or ligation) to genomic regions flanking a target region suspected of containing an insertion or deletion. It should be appreciated that two targeting arms may be designed to be complementary (e.g., sufficiently complementary for selective hybridization and/or polymerase extension and/or ligation) to the two flanking regions that are immediately adjacent (e.g., immediately 5' and 3', respectively) to a region of a sequence repeat on one strand of a genomic nucleic acid. However, one or both targeting arms may be designed to hybridize several bases (e.g., 1-5, 5-10, 10-25, 25-50, or more) upstream or downstream from the repeat region in such a way that the captured sequence includes a region of unique genomic sequence that on one or both sides of the repeat region. This unique region can then be used to identify the captured target (e.g., based on sequence or hybridization information).

In some embodiments, two or more (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10 or more) different loci may be interrogated in parallel in a single assay (e.g., in a multiplex assay). In some embodiments, the ratio of captured nucleic acids for each locus may be used to determine whether a nucleic acid insertion or deletion is present in one locus relative to the other. For example, the ratio may be compared to a control ratio that is representative of the two loci when neither one has an insertion or deletion relative to control sequences (e.g., sequences that are normal or known to be associated with healthy phenotypes for those loci). However, the amount of captured nucleic acids may be compared to any suitable control as discussed herein.

The locus of a captured sequence may be identified by determining a portion of unique sequence 5' and/or 3' to the repeat region in the target nucleic acid suspected of containing a deletion or insertion. This does not require sequencing the captured repeat region itself. However, some or all of the repeat region also could be sequenced as aspects of the invention are not limited in this respect.

Aspects of the invention may be combined with one or more sequence-based assays (e.g., SNP detection assays), for example in a multiplex format, to determine the genotype of one or more regions of a subject.

In some embodiments, methods of detecting a polymorphism in a nucleic acid in a biological sample are provided. In some embodiments, the methods comprise evaluating the efficiency of capture at one or more loci and determining whether one or both alleles at that locus contain an insertion or deletion relative to a control locus (e.g., a locus indicative of a length of repeat sequence that is associated with a healthy phenotype).

Accordingly, aspects of the invention relate to methods for determining whether a target nucleic acid has an abnormal length by evaluating the capture efficiency of a target nucleic acid in a biological sample from a subject, wherein a capture efficiency that is different from a reference capture efficiency is indicative of the presence, in the biological sample, of a target nucleic acid having an abnormal length. It should be appreciated that the term "abnormal" is a relative term based on a comparison to a "normal" length. In some embodiments, a normal length is a length that is associated with a normal (e.g., healthy or non-carrier phenotype). Accordingly, an abnormal length is a length that is either shorter or longer than the normal length. In some embodiments, the presence of an abnormal length is indicative of an increased risk that the locus is associated with a disease or a disease carrier phenotype. In some embodiments, the abnormal length is indicative that the subject is either has a disease or condition or is a carrier of a disease or condition (e.g., associated with the locus). However, it should be appreciated that the description of embodiments relating to detecting the presence of an abnormal length also support detecting the presence of a length that is different from an expected or control length.

In some embodiments, aspects of the invention relate to estimating the length of a target nucleic acid (e.g., of a sub-target region within a target nucleic acid). In some embodiments, aspects of the invention relate to methods for estimating the length of a target nucleic acid by contacting the target nucleic acid with a plurality of detection probes under conditions that permit hybridization of the detection probes to the target nucleic acid, wherein each detection probe is a polynucleotide that comprises a first arm that hybridizes to a first region of the target nucleic acid and a second arm that hybridizes to a second region of the target nucleic acid, wherein the first and second regions are on a common strand of the target nucleic acid, and wherein the nucleotide sequence of the target between the 5' end of the first region and the 3' end of the second region is the nucleotide sequence of a sub-target nucleic acid; and capturing a plurality of sub-target nucleic acids that are hybridized with the plurality of detection probes; and measuring the frequency of occurrence of a sub-target nucleic acid in the plurality of sub-target nucleic acids, wherein the frequency of occurrence of the sub-target nucleic acid in the plurality of sub-target nucleic acids is indicative of the length of the sub-target nucleic acid. It should be appreciated that methods for estimating a nucleic acid length may involve comparing a capture efficiency for a target nucleic acid region to two or more reference efficiencies for known nucleic acid lengths in order to determine whether the target nucleic acid region is smaller, intermediate, or larger in size than the known control lengths. In some embodiments, a series of nucleic acids of known different lengths may be used to provide a calibration curve for evaluating the length of a target nucleic acid region of interest.

In some embodiments, the capture efficiency of a target region suspected of having a deletion or insertion is determined by comparing the capture efficiency to a reference indicative of a normal capture efficiency. In some embodiments, the capture efficiency is lower than the reference capture efficiency. In some embodiments, the subject is identified as having an insertion in the target region. In some embodiments, the capture efficiency is higher than the reference capture efficiency. In some embodiments, the subject is identified as having a deletion in the target region. In some embodiments, the subject is identified as being heterozygous for the insertion. In some embodiments, the subject is identified as being heterozygous for the deletion.

In some embodiments of any of the methods described herein (e.g., tiling/staggering, tagging, size-detection, and/or sensitivity enhancement) aspects of the invention relate to capturing a sub-target nucleic acid (or a sequence of a sub-target nucleic acid). In some embodiments, a molecular inversion probe technique is used. In some embodiments, a molecular inversion probe is a single linear strand of nucleic acid that comprises a first targeting arm at its 5' end and a

second targeting arm at its 3' end, wherein the first targeting arm is capable of specifically hybridizing to a first region flanking one end of the sub-target nucleic acid, and wherein the second targeting arm is capable of specifically hybridizing to a second region flanking the other end of the sub target nucleic acid on the same strand of the target nucleic acid. In some embodiments, the first and second targeting arms are between about 10 and about 100 nucleotides long. In some embodiments, the first and second targeting arms are about 10-20, 20-30, 30-40, or 40-50 nucleotides long. In some embodiments, the first and second targeting arms are about 20 nucleotides long. In some embodiments, the first and second targeting arms have the same length. In some embodiments, the first and second targeting arms have different lengths. In some embodiments, each pair of first and second targeting arms in a set of probes has the same length. Accordingly, if one of the targeting arms is longer, the other one is correspondingly shorter. This allows for a quality control step in some embodiments to confirm that all captured probe/target sequence products have the same length after a multiplexed plurality of capture reactions. In some embodiments, a set of probes may be designed to have the same length if the intervening region is varied to accommodate any differences in the length of either one or both of the first and second targeting arms.

In some embodiments, the hybridization Tms of the first and second targeting arms are similar. In some embodiments, the hybridization Tms of the first and second targeting arms are within 2-5° C. of each other. In some embodiments, the hybridization Tms of the first and second targeting arms are identical. In some embodiments, the hybridization Tms of the first and second targeting arms are close to empirically-determined optima but not necessarily identical.

In some embodiments, the first and second targeting arms of a molecular inversion probe have different Tms. For example, the Tm of the first targeting arm (at the 5' end of the molecular inversion probe) may be higher than the Tm of the second targeting arm (at the 3' end of the molecular inversion probe). According to aspects of the invention, and without wishing to be bound by theory, a relatively high Tm for the first targeting arm may help avoid or prevent the first targeting arm from being displaced after hybridization by the extension product of the 3' end of the second targeting arm. It should be appreciated that a reference to the Tm of a targeting arm as used herein relates to the Tm of hybridization of the targeting arm to a nucleic acid having the complementary sequence (e.g., the region of the target nucleic acid that has a sequence that is complementary to the sequence of the targeting arm). It also should be appreciated that the Tms of the targeting arms described herein may be calculated using any appropriate method. For example, in some embodiments an experimental method (e.g., a gel shift assay, a hybridization assay, a melting curve analysis, for example in a PCR machine with a SYBR dye by stepping through a temperature ramp while monitoring signal level from an intercalating dye, for example, bound to a double-stranded DNA, etc.) may be used to determine one or more Tms empirically. In some embodiments, an optimal Tm may be determined by evaluating the number of products formed (e.g., for each of a plurality of MIP probes), and determining the optimal Tm as the center point in a histogram of Tm for all targeting arms. In some embodiments, a predictive algorithm may be used to determine a Tm theoretically. In some embodiments, a relatively simple predictive algorithm may be used based on the number of G/C and A/T base pairs when the sequence is hybridized to its target and/or the length of the hybridized product (e.g., for example, 64.9+41*([G+C]−16.4)/(A+T+G+C), see for example, Wallace, R. B., Shaffer, J., Murphy, R. F., Bonner, J., Hirose, T., and Itakura, K. (1979) Nucleic Acids Res 6:3543-3557). In some embodiments, a more complex algorithm may be used to account for the effects of base stacking entropy and enthalpy, ion concentration, and primer concentration (see, for example, SantaLucia J (1998), Proc Natl Acad Sci USA, 95:1460-5). In some embodiments an algorithm may use modified parameters (e.g., nearest-neighbor parameters for basepair entropy/enthalpy values). It should be appreciated that any suitable algorithm may be used as aspects of the invention are not limited in this respect. However, it also should be appreciated that different methodologies may results in different calculated or predicted Tms for the same sequences. Accordingly, in some embodiments, the same empirical and/or theoretical method is used to determine the Tms of different sequences for a set of probes to avoid a negative impact of any systematic difference in the Tm determination or prediction when designing a set of probes with predetermined similarities or differences for different Tms.

In some embodiments, the Tm of the first targeting arm may be about 1° C., about 2° C., about 3° C., about 4° C., about 5° C., or more than about 5° C. higher than the Tm of the second targeting arm. In some embodiments, each probe in a plurality of probes (e.g., each probe in a set of 5-10, each probe in a set of at least 10, each probe in a set of 10-50, each probe in a set of 50-100, each probe in a set of 100-500, each probe in a set of 500-1,000, each probe in a set of 1,000-1,500, each probe in a set of 1,500-2,000, each probe in a set of 2,000-3,000, 3,000-5,000, 5,000-10,000 or each probe in a set of at least 5,000 different probes) has a unique first targeting arm (e.g., they all have different sequences) and a unique second targeting arm (e.g., they all have different sequences). In some embodiments, for at least 10% of the probes (e.g., at least 25%, 25%-50%, 50%-75%, 75%-90%, 90%-95% or over 95%, or all of the probes) the first targeting arm has a Tm for its complementary sequence that is higher (e.g., about 1° C., about 2° C., about 3° C., about 4° C., about 5° C., or more than about 5° C. higher) than the Tm of the second targeting arm for its complementary sequence. In some embodiments, each of the first targeting arms have similar or identical Tms for their respective complementary sequences and each of the second targeting arms have similar or identical Tms for their respective complementary sequences (and the first targeting arms have higher Tms than the second targeting arms). For example, in some embodiments, the Tm of the first arm(s) may be about 58° C. and the Tm of the second arm(s) may be about 56° C. In some embodiments, the Tm of the first arm(s) may be about 68° C., and the Tm of the second arm(s) may be about 65° C. It should be appreciated that in some embodiments the similarity (e.g., within a range of 1° C., 2° C., 3° C., 4° C., 5° C.) or identity of the Tms for the different targeting arms should be based either on empirical data for each arm or based on the same predictive algorithm for each arm (e.g., Wallace, R. B., Shaffer, J., Murphy, R. F., Bonner, J., Hirose, T., and Itakura, K. (1979) Nucleic Acids Res 6:3543-3557, SantaLucia J (1998), Proc Natl Acad Sci USA, 95:1460-5, or other algorithm).

In some embodiments, the Tm of the first targeting arm of a molecular inversion probe (at the 5' end of the molecular inversion probe) is selected to be sufficiently stable to prevent displacement of the first targeting arm from its complementary sequence on a target nucleic acid. In some embodiments, the Tm of the first targeting arm is 50-55° C.,

at least 55° C., 55-60° C., at least 60° C., 60-65° C., at least 65° C., at least 70° C., at least 75° C., or at least 80° C. As discussed above, it should be appreciated that the for a particular targeting arm may be determined empirically or theoretically. Different theoretical models may be used to determine a Tm and it should be appreciated that the predicted Tm for a particular sequence may be different depending on the algorithm used for the prediction. In some embodiments, each probe in a plurality of probes (e.g., each probe in a set of 5-10, each probe in a set of at least 10, each probe in a set of 10-50, each probe in a set of 50-100, each probe in a set of 100-500, or each probe in a set of at least 500 different probes) has a different first targeting arm (e.g., different sequences) but each different first targeting arm has a similar or identical Tm for its complementary sequence on a target nucleic acid. It should be appreciated that in some embodiments the similarity (e.g., within a range of I C, 2 C, 3 C, 4 C, 5 C) or identity of the Tms for the different targeting arms should be based either on empirical data for each arm or based on the same predictive algorithm for each arm (e.g., Wallace, R. B., Shaffer, J., Murphy, R. F., Bonner, J., Hirose, T., and Itakura, K. (1979) Nucleic Acids Res 6:3543-3557, SantaLucia J (1998), Proc Natl Acad Sci USA, 95:1460-5, or other algorithm).

In some embodiments, the sub-target nucleic acid contains a nucleic acid repeat. In some embodiments, the nucleic acid repeat is a dinucleotide or trinucleotide repeat. In some embodiments, the sub-target nucleic acid contains 10-100 copies of the nucleic acid repeat in the absence of an abnormal increase or decrease in nucleic acid repeats. In some embodiments, the sub-target nucleic acid is a region of the Fragile-X locus that contains a nucleic acid repeat. In some embodiments, one or both targeting arms hybridize to a region on the target nucleic acid that is immediately adjacent to a region of nucleic acid repeats. In some embodiments, one or both targeting arms hybridize to a region on the target nucleic acid that is separated from a region of nucleic acid repeats by a region that does not contain any nucleic acid repeats. In some embodiments, the molecular inversion probe further comprises a primer-binding region that can be used to sequence the captured sub-target nucleic acid and optionally the first and/or second targeting arm.

In some embodiments, aspects of the invention relate to evaluating the length of a plurality of different target nucleic acids in a biological sample. In some embodiments, the plurality of target nucleic acids are analyzed using a plurality of different molecular inversion probes. In some embodiments, each different molecular inversion probe comprises a different pair of first and second targeting arms at each of the 3' and 5' ends. In some embodiments, each different molecular inversion probe comprises the same primer-binding sequence.

In some embodiments, aspects of the invention relate to analyzing nucleic acid from a biological sample obtained from a subject. In some embodiments, the biological sample is a blood sample. In some embodiments, the biological sample is a tissue sample, specific cell population, tumor sample, circulating tumor cells, or environmental sample. In some embodiments, the biological sample is a single cell. In some embodiments, nucleic acids are analyzed in biological samples obtained from a plurality of different subjects. In some embodiments, nucleic acids from a biological sample are analyzed in multiplex reactions. It should be appreciated that a biological sample contains a plurality of copies of a genome derived from a plurality of cells in the sample. Accordingly, a sample may contain a plurality of independent copies of a target nucleic acid region of interest, the capture efficiency of which can be used to evaluate its size as described herein.

In some embodiments, aspects of the invention relate to evaluating a nucleic acid capture efficiency by determining an amount of target nucleic acid that is captured (e.g., an amount of sub-target nucleic acid sequences that are captured). In some embodiments, the amount of target nucleic acid that is captured is determined by determining a number of independently captured target nucleic acid molecules (e.g., the amount of independently captured molecules that have the sequence of the sub-target region). In some embodiments, the amount of target nucleic acid that is captured is compared to a reference amount of captured nucleic acid. In some embodiments, the reference amount is determined by determining a number of independently captured molecules of a reference nucleic acid. In some embodiments, the reference nucleic acid is a nucleic acid of a different locus in the biological sample that is not suspected of containing a deletion or insertion. In some embodiments, the reference nucleic acid is a nucleic acid of known size and amount that is added to the capture reaction. As described herein, a number of independently captured nucleic acid sequences can be determined by contacting a nucleic acid sample with a preparation of a probe (e.g., a MIP probe as described herein). It should be appreciated that the preparation may comprise a plurality of copies of the same probe and accordingly a plurality of independent copies of the target region may be captured by different probe molecules. The number of probe molecules that actually capture a sequence can be evaluated by determining an amount or number of captured molecules using any suitable technique. This number is a reflection of both the number of target molecules in the sample and the efficiency of capture of those target molecules, which in turn is related to the size of the target molecules as described herein. Accordingly, the capture efficiency can be evaluated by controlling for the abundance of the target nucleic acid, for example by comparing the number or amount of captured target molecules to an appropriate control (e.g., a known size and amount of control nucleic acid, or a different locus that should be present in the same amount in the biological sample and is not expected to contain any insertions or deletions). It should be appreciated that other factors may affect the capture efficiency of a particular target nucleic acid region (e.g., the sequence of the region, the GC content, the presence of secondary structures, etc.). However, these factors also can be accounted for by using appropriate controls (e.g., known sequences having similar properties, the same sequences, other genomic sequences expected to be present in the biological sample at the same frequency, etc., or any combination thereof).

In some embodiments, aspects of the invention relate to identifying a subject as having an insertion or deletion in one or more alleles of a genetic locus if the capture efficiency for that genetic locus is statistically significantly different than a reference capture efficiency.

It should be appreciated that hybridization conditions used for any of the capture techniques described herein (e.g., MIP capture techniques) can be based on known hybridization buffers and conditions.

In some embodiments, the methods disclosed herein are useful for any application where the detection of deletions or insertions is important.

In some embodiments, aspects of the invention relate to basing a nucleic acid sequence analysis on results from two or more different nucleic acid preparatory techniques that

have different systematic biases in the types of nucleic acids that they sample. According to the invention, different techniques have different sequence biases that are systematic and not simply due to stochastic effects during nucleic acid capture or amplification. Accordingly, the degree of oversampling required to overcome variations in nucleic acid preparation needs to be sufficient to overcome the biases (e.g., an oversampling of 2-5 fold, 5-10 fold, 5-15 fold, 15-20 fold, 20-30 fold, 30-50 fold, or intermediate to higher fold).

According to some embodiments, different techniques have different characteristic or systematic biases. For example, one technique may bias a sample analysis towards one particular allele at a genetic locus of interest, whereas a different technique would bias the sample analysis towards a different allele at the same locus. Accordingly, the same sample may be identified as being different depending on the type of technique that is used to prepare nucleic acid for sequence analysis. This effectively represents a sensitivity limitation, because each technique has different relative sensitivities for polymorphic sequences of interest.

According to aspects of the invention, the sensitivity of a nucleic acid analysis can be increased by combining the sequences from different nucleic acid preparative steps and using the combined sequence information for a diagnostic assay (e.g., for a making a call as to whether a subject is homozygous or heterozygous at a genetic locus of interest).

In some embodiments, the invention provides a method of increasing the sensitivity of a nucleic acid detection assay by obtaining a first preparation of a target nucleic acid using a first preparative method on a biological sample, obtaining a second preparation of a target nucleic acid using a second preparative method on the biological sample, assaying the sequences obtained in both first and second nucleic acid preparations, and using the sequence information from both first and second nucleic acid preparations to determine the genotype of the target nucleic acid in the biological sample, wherein the first and second preparative methods have different systematic sequence biases. In some embodiments, the first and second nucleic acid preparations are combined prior to performing a sequence assay. In some embodiments, separate sequence assays are performed on the first and second nucleic acid preparations and the sequence information from both assays are combined to determine the genotype of the target nucleic acid in the biological sample. In some embodiments, the first preparative method is an amplification-based, a hybridization-based, or a circular probe-based preparative method. In some embodiments, the second method is an amplification-based, a hybridization-based, or a circular probe-based preparative method. In some embodiments, the first and second methods are of different types (e.g., only one of them is an amplification-based, a hybridization-based, or a circular probe-based preparative method, and the other one is one or the other two types of method). Accordingly, in some embodiments the second preparative method is an amplification-based, a hybridization-based, or a circular probe-based preparative method, provided that the second method is different from the first method. However, in some embodiments, both methods may be of the same type, provided they are different methods (e.g., both are amplification based or hybridization-based, but are different types of amplification or hybridization methods, e.g., with different relative biases).

In amplification-based (e.g., PCR-based or LCR-based, etc.) preparative methods, genomic loci (target nucleic acids) are isolated directly by means of a polymerase chain reaction or ligase chain reaction (or other amplification method) that selectively amplifies each locus using a pair of oligonucleotide primers. It is to be understood that primers will be sufficiently complementary to the target sequence to hybridize with and prime amplification of the target nucleic acid. Any one of a variety of art known methods may be utilized for primer design and synthesis. One or both of the primers may be perfectly complementary to the target sequence. Degenerate primers may also be used. Primers may also include additional nucleic acids that are not complementary to target sequences but that facilitate downstream applications, including for example restriction sites and identifier sequences (e.g., source sequences). PCR based methods may include amplification of a single target nucleic acid and multiplex amplification (amplification of multiple target nucleic acids in parallel).

Hybridization-based preparative may methods involve selectively immobilizing target nucleic acids for further manipulation. It is to be understood that one or more oligonucleotides (immobilization oligonucleotides), which in some embodiments may be from 10 to 200 nucleotides in length, are used which hybridize along the length of a target region of a genetic locus to immobilize it. In some embodiments, immobilization oligonucleotides are either immobilized before hybridization is performed (e.g., Roche/Nimblegen 'sequence capture'), or are prepared such that they include a moiety (e.g., biotin) which can be used to selectively immobilize the target nucleic acid after hybridization by binding to e.g., streptavidin-coated microbeads (e.g., Agilent 'SureSelect').

Circularization selection-based preparative methods selectively convert each region of interest into a covalently-closed circular molecule which is then isolated by removal (usually enzymatic, e.g., with exonuclease) of any non-circularized linear nucleic acid. Oligonucleotide probes are designed which have ends that flank the region of interest. The probes are allowed to hybridize to the genomic target, and enzymes are used to first (optionally) fill in any gap between probe ends and second ligate the probe closed. In some embodiments, following circularization, any remaining (non-target) linear nucleic acid can be removed, resulting in isolation (capture) of target nucleic acid. Circularization selection-based preparative methods include molecular inversion probe capture reactions and 'selector' capture reactions. However, other techniques may be used as aspects of the invention are not limited in this respect. In some embodiments, molecular inversion probe capture of a target nucleic acid is indicative of the presence of a polymorphism in the target nucleic acid.

A variety of methods may be used to evaluate and compare bias profiles of each preparative technique. Next-generation sequencing may be used to quantitatively measure the abundance of each isolated target nucleic acid obtained from a certain preparative method. This abundance may be compared to a control abundance value (e.g., a known starting abundance of the target nucleic acid) and/or with an abundance determined through the use of an alternative preparative method. For example, a set of target nucleic acids may be isolated by one or more of the three preparative methods; the target nucleic acid may be observed x times using the amplification technique, y times using the hybridization enrichment technique, and z times using the circularization selection technique. A pairwise correlation coefficient may be computed between each abundance value (e.g., x and y, x and z, and y and z) to assess bias in nucleic acid isolation between pairs of preparative methods. Since the mechanisms of isolation are different in each

approach, the abundances will usually be different and largely uncorrelated with each other.

In some embodiments, the invention provides a method of obtaining a nucleic acid preparation that is representative of a target nucleic acid in a biological sample by obtaining a first preparation of a target nucleic acid using a first preparative method on a biological sample, obtaining a second preparation of a target nucleic acid using a second preparative method on the biological sample, and combining the first and second nucleic acid preparations to obtain a combined preparation that is representative of the target nucleic acid in the biological sample.

In some embodiments of any of the methods described herein, a third preparation of the target nucleic acid is obtained using a third preparative method that is different from the first and second preparative methods, wherein the first, second, and third preparative methods all have different systematic sequence biases. In some embodiments of any of the methods described herein, the different preparative methods are used for a plurality of different loci in the biological sample to increase the sensitivity of a multiplex nucleic acid analysis. In some embodiments, the target nucleic acid has a sequence of a gene selected from Table 1.

However, it should be appreciated that a genotyping method of the invention may include several steps, each of which independently may involve one or more different preparative techniques described herein. In some embodiments, a nucleic acid preparation may be obtained using one or more (e.g., 2, 3, 4, 5, or more) different techniques described herein (e.g., amplification, hybridization capture, circular probe capture, etc., or any combination thereof) and the nucleic acid preparation may be analyzed using one or more different techniques (e.g., amplification, hybridization capture, circular probe capture, etc., or any combination thereof) that are selected independently of the techniques used for the initial preparation.

In some embodiments, aspects of the invention also provide compositions, kits, devices, and analytical methods for increasing the sensitivity of nucleic acid assays. Aspects of the invention are particularly useful for increasing the confidence level of genotyping analyses. However, aspects of the invention may be used in the context of any suitable nucleic acid analysis, for example, but not limited to, a nucleic acid analysis that is designed to determine whether more than one sequence variant is present in a sample.

In some embodiments, aspects of the invention relate to a plurality of nucleic acid probes (e.g., 10-50, 50-100, 100-250, 250-500, 500-1,000, 1,000-2,000, 2,000-5,000, 5,000-7,500, 7,500-10,000, or lower, higher, or intermediate number of different probes).

In some embodiments, each probe or each of a subset of probes (e.g., 10-25%, 25-50%, 50-75%, 75-90%, or 90-99%) has a different first targeting arm. In some embodiments, each probe or each probe of a subset of probes (e.g., 10-25%, 25-50%, 50-75%, 75-90%, or 90-99%) has a different second targeting arm. In some embodiments, the first and second targeting arms are separated by the same intervening sequence. In some embodiments, the first and second targeting arms are complementary to target nucleic acid sequences that are separated by the same or a similar length (e.g., number of nucleic acids, for example, 0-25, 25-50, 50-100, 100-250, 250-500, 500-1,000, 1,000-2,500 or longer or intermediate number of nucleotides) on their respective target nucleic acids (e.g., genomic loci). In some embodiments, each probe or a subset of probes (e.g., 10-30 25%, 25-50%, 50-75%, 75-90%, or 90-99%) includes a first primer binding sequence. In some embodiments, the primer binding sequence is the same (e.g., it can be used to prime sequencing or other extension reaction). In some embodiments, each probe or a subset of probes (e.g., 10-25%, 25-50%, 50-75%, 75-90%, or 90-99%) includes a unique identifier sequence tag (e.g., that is predetermined and can be used to distinguish each probe).

In some embodiments, the methods disclosed herein are useful for any application where sensitivity is important. For example, detection of cancer mutations in a heterogenous tissue sample, detection of mutations in maternally-circulating fetal DNA, and detection of mutations in cells isolated during a preimplantation genetic diagnostic procedure.

According to some aspects of the invention, methods of detecting a polymorphism in a nucleic acid in a biological sample are provided. In some embodiments, the methods comprise obtaining a nucleic acid preparation using a preparative method (e.g., any of the preparative methods disclosed herein) on a biological sample, and performing a molecular inversion probe capture reaction on the nucleic acid preparation, wherein a molecular inversion probe capture (e.g., using a mutation-detection MIP) of a target nucleic acid of the nucleic acid preparation is indicative of the presence of a mutation (polymorphism) in the target nucleic acid, optionally wherein the polymorphism is selected from Table 2.

According to some aspects of the invention, methods of genotyping a nucleic acid in a biological sample are provided. In some embodiments, the methods comprise obtaining a nucleic acid preparation using a preparative method on a biological sample, sequencing a target nucleic acid of the nucleic acid preparation, and performing a molecular inversion probe capture reaction on the biological sample, wherein a molecular inversion probe capture of the target nucleic acid in the biological sample is indicative of the presence of a polymorphism in the target nucleic acid, genotyping the target nucleic acid based on the results of the sequencing and the capture reaction.

In some embodiments of the methods disclosed herein, the target nucleic acid has a sequence of a gene selected from Table 1.

It should be appreciated that any one or more embodiments described herein may be used for evaluating multiple genetic markers in parallel. Accordingly, in some embodiments, aspects of the invention relate to determining the presence of one or more markers (e.g., one or more alleles) at multiple different genetic loci in parallel.

Accordingly, the risk or presence of multiple heritable disorders may be evaluated in parallel. In some embodiments, the risk of having offspring with one or more heritable disorders may be evaluated. In some embodiments, an evaluation may be performed on a biological sample of a parent or a child (e.g., at a pre-implantation, prenatal, perinatal, or postnatal stage).

In some embodiments, the disclosure provides methods for analyzing multiple genetic loci (e.g., a plurality of target nucleic acids selected from Table 1 or 2) from a patient sample, such as a blood, pre-implantation embryo, chorionic villus or amniotic fluid sample. A patient or subject may be a human. However, aspects of the invention are not limited to humans and may be applied to other species (e.g., mammals, birds, reptiles, other vertebrates or invertebrates) as aspects of the invention are not limited in this respect. A subject or patient may be male or female. In some embodiments, in connection with reproductive genetic counseling, samples from a male and female member of a couple may be analyzed. In some embodiments, for example, in connection with an animal breeding program, samples from a plurality

of male and female subjects may be analyzed to determine compatible or optimal breeding partners or strategies for particular traits or to avoid one or more diseases or conditions. Accordingly, reproductive risks may be determined and/or reproductive recommendations may be provided based on information derived from one or more embodiments of the invention.

However, it should be appreciated that aspects of the invention may be used in connection with any medical evaluation where the presence of one or more alleles at a genetic locus of interest is relevant to a medical determination (e.g., risk or detection of disease, disease prognosis, therapy selection, therapy monitoring, etc.). Further aspects of the invention may be used in connection with detection, in tumor tissue or circulating tumor cells, of mutations in cellular pathways that cause cancer or predict efficacy of treatment regimens, or with detection and identification of pathogenic organisms in the environment or a sample obtained from a subject, e.g., a human subject.

These and other aspects of the invention are described in more detail in the following description and non-limiting examples and drawings.

BRIEF DESCRIPTION OF DRAWINGS

FIG. **1** illustrates a non-limiting embodiment of a tiled probe layout;

FIG. **2** illustrates a non-limiting embodiment of a staggered probe layout;

FIG. **3** illustrates a non-limiting embodiment of an alternating staggered probe layout;

FIGS. **4**A-C depict various non-limiting methods for combining differentiator tag sequence and target sequences (NNNN depicts a differentiator tag sequence);

FIG. **5** depicts a non-limiting method for genotyping based on target and 5 differentiator tag sequences;

FIG. **6** depicts non-limiting results of a simulation of a MIP capture reaction;

FIG. **7** depicts a non-limiting graph of sequencing coverage;

FIG. **8** illustrates that shorter sequences are captured with higher efficiency that longer sequences using MIPs;

FIG. **9** illustrates a non-limiting scheme of padlock (MIP) capture of a region that includes both repetitive regions (thick wavy line) and the adjacent unique sequence (thick strait line);

FIG. **10** illustrates a non-limiting hypothetical relationship between target gap size and the relative number of reads of the repetitive region;

FIG. **11**A depicts MIP capture of FMR1 repeat regions from a diploid genome;

FIG. **11**B depicts preparative methods for biallelic resolution of FMR I repeat region lengths in a diploid genome using MIP capture probes and unique differentiator tags;

FIG. **11**C depicts an analysis of FMR1 repeat region lengths in a diploid genome;

FIG. **12** is a schematic of an embodiment of an algorithm of the invention;

FIG. **13** illustrates a non-limiting example of a graph of per-target abundance with MIP capture; and,

FIG. **14** shows a non-limiting a graph of correlation between two MIP capture reactions.

DETAILED DESCRIPTION

Aspects of the invention relate to preparative and analytical methods and compositions for evaluating genotypes, and in particular, for determining the allelic identity (or identities in a diploid organism) of one or more genetic loci in a subject.

Aspects of the invention are based, in part, on the identification of different sources of ambiguity and error in genetic analyses, and, in part, on the identification of one or more approaches to avoid, reduce, recognize, and/or resolve these errors and ambiguities at different stages in a genetic analysis. Aspects of the invention relate to methods and compositions for addressing bias and/or stochastic variation associated with one or more preparative and/or analytical steps of a nucleic acid evaluation technology. In some embodiments, preparative methods can be adapted to avoid or reduce the risk of bias skewing the results of a genetic analysis. In some embodiments, analytical methods can be adapted to recognize and correct for data variations that may give rise to misinterpretation (e.g., incorrect calls such as homozygous when the subject is actually heterozygous or heterozygous when the subject is actually homozygous). Methods of the invention may be used for any type of mutation, for example a single base change (e.g., insertion, deletion, transversion or transition, etc.), a multiple base insertion, deletion, duplication, inversion, and/or any other change or combination thereof.

In some embodiments, additional or alternative techniques may be used to address loci characterized by multiple repeats of a core sequence where the length of the repeat is longer than a typical sequencing read thereby making it difficult to determine whether a deletion or duplication of one or more core sequence units has occurred based solely on a sequence read.

In some embodiments, increased confidence in an assay result may be obtained by i) selecting two or more different preparative and/or analytical techniques that have different biases (e.g., known to have different biases), ii) evaluating a patient sample using the two or more different techniques, iii) comparing the results from the two or more different techniques, and/or iv) determining whether the results are consistent for the two or more different techniques. In some embodiments, if determining in step (iv) indicates that the results are consistent (e.g., the same) then increased confidence in the assay result is obtained. In other embodiments, if determining in step (iv) indicates that the results are inconsistent (e.g., that the results are ambiguous) then one or more additional preparative and/or analytical techniques, which have a different bias (e.g., known to have a different bias) compared with the two or more different preparative and/or analytical techniques selected in step (i), are used to evaluate the patient sample, and the results of the one or more additional preparative and/or analytical techniques are compared with the results from step (ii) to resolve the inconsistency.

In some embodiments, two or more independent samples may be obtained from a subject and independently analyzed. In some embodiments, two or more independent samples are obtained at approximately the same time point. In some embodiments, two or more independent samples are obtained at multiple different time points. In some embodiments, the use of two or more independent sample facilitates the elimination, normalization, and/or quantification of stochastic measurement noise. It is to be appreciated that two or more independent samples may be obtained in connection with any of the methods disclosed herein, including, for example, methods for pathogen profiling in a human or other animal subjects, monitoring tumor progression/regression,

analyzing circulating tumor cells, analyzing fetal cells in maternal circulation, and analyzing/monitoring/profiling of environmental pathogens.

In some embodiments, one or more of the techniques described herein may be combined in a single assay protocol for evaluating multiple patient samples in parallel.

It should be appreciated that aspects of the invention may be useful for high throughput, cost-effective, yet reliable, genotyping of multiple patient samples (e.g., in parallel, for example in multiplex reactions). In some embodiments, aspects of the invention are useful to reduce the error frequency in a multiplex analysis. Certain embodiments may be particularly useful where multiple reactions (e.g., multiple loci and/or multiple patient samples) are being processed. For example, 10-25, 25-50, 50-75, 75-100 or more loci may be evaluated for each subject out of any number of subject samples that may be processed in parallel (e.g., 1-25, 25-50, 50-100, 100-500, 500-1,000, 1,000-2,500, 2,500-5,000 or more or intermediate numbers of patient samples). It should be appreciated that different embodiments of the invention may involve conducting two or more target capture reactions and/or two or more patient sample analyses in parallel in a single multiplex reaction. For example, in some embodiments a plurality of capture reactions (e.g., using different capture probes for different target loci) may be performed in a single multiplex reaction on a single patient sample. In some embodiments, a plurality of captured nucleic acids from each one of a plurality of patient samples may be combined in a single multiplex analysis reaction. In some embodiments, samples from different subjects are tagged with subject-specific (e.g., patient-specific) tags (e.g., unique sequence tags) so that the information from each product can be assigned to an identified subject. In some embodiments, each of the different capture probes used for each patient sample have a common patient-specific tag. In some embodiments, the capture probes do not have patient-specific tags, but the captured products from each subject may be amplified using one or a pair of amplification primers that are labeled with a patient-specific tag. Other techniques for associating a patient-specific tag with the captured product from a single patient sample may be used as aspects of the invention are not limited in this respect. It should be appreciated that patient-specific tags as used herein may refer to unique tags that are assigned to identified patients in a particular assay. The same tags may be used in a separate multiplex analysis with a different set of patient samples (e.g., from different patients) each of which is assigned one of the tags. In some embodiments, different sets of unique tags may be used in sequential (e.g., alternating) multiplex reactions in order to reduce the risk of contamination from one assay to the next and allow contamination to be detected on the basis of the presence of tags that are not expected to be present in a particular assay.

Embodiments of the invention may be used for any of a number of different settings: reproductive settings, disease screening, identifying subjects having cancer, identifying subjects having increased risk for a disease, stratifying a population of subjects according to one or more of a number of factors, for example responsiveness to a particular drug, lack or not of an adverse reaction (or risk therefore) to a particular drug, and/or providing information for medical records (e.g., homozygosity, heterozygosity at one or more loci). It should be appreciated that the invention is not limited to genomic analysis of patient samples. For example, aspects of the invention may be useful for high throughput genetic analysis of environment samples to detect pathogens.

In some embodiments, the methods disclosed herein are useful for diagnosis of one or more heritable disorders. In some embodiments, a heritable disorder that may be diagnosed with the methods disclosed herein is a genetic disorder that is prevalent in the Ashkenazi Jewish population. In some embodiments, the heritable disorders are selected from: 21-Hydroxylase-Defiocient Congenital Adrenal Hyperplasia; ABCC8-Related Hyperinsulinism; Alpha-Thalassemia, includes Constant Spring, & MR associated; Arylsulfatase A Deficiency-Metyachromatic Leukodystrophy; Biotinidase Deficiency Holocarboxylase Synthetase Deficiency; Bloom's Syndrome; Canavan Disease; CFTR-Related Disorders-cystic fibrosis; Citrullinemia Type 1; Combined MMA & Homocystinuria-db1C; Dystrophinopathies (DMD & BMD); Familial Dysautonomia; Fanconi Anemia-FANCC; Galactosemia-Classical: Galactokinase Defiency & Galactose Epimerase Deficiency; Gaucher Disease; GJB2-Related DFNB 1 Nonsyndromic Hearing Loss and Deafness; Glutaric acidemia Type 1; Hemoglobinopathies beta-chain disorders; Glycogen Storage Disease Type 1A; Maple Syrup Urine Disease; Types 1A, 1B, 2, 3; Medium Chain Acyl-Coenzyme A; Dehydrogenase Deficiency-MCADD; Methylmalonic Acidemia; Mucolipidosis IV; Nemaline Myopathy; Nieman-Pick Type A-Acid Sphingomyelinase Deficiency; Non-Ketotic Hyperglycinemia-Glycine Encephalopathy; Ornithine Transcarbamylase Deficiency; PKU Phenylalanine Hydroxylase Deficiency; Propionic Acidemia; Short Chain Acyl-CoA Dehydrogenase Deficiency-SCADD; Smith-Lemli-Opitz Syndrome; Spinal Muscular Atrophy (SMNI)-SMA; Tay Sachs-HexA Deficiency; Usher Synbdrome-Type I (Type IB, Type IC, Type ID, Type IF, Type IG); X-Linked Mental Retardation ARX-Related Disorders; X-Linked Mental Retardation with Cerebellar Cypoplasia and sistinctive Facial Appearance; X-Linked Mental Retardation; includes 9, 21, 30, 46, 58, 63, 88, 89; X linked mental retardation: FM1-Related Disorders—FRXA, Fragile X MR; X-linked SMR: Renpenning Syndrome 1; Zellweger Spectrum disorders—Peroxisomal Bifunctional Enzyme Deficiencies including Zellweger, NALD, and/or infantile Refsums. However, all of these, subsets of these, other genes, or combinations thereof may be used.

According to some aspects, the disclosure relates to multiplex diagnostic methods. In some embodiments, multiplex diagnostic methods comprise capturing a plurality of genetic loci in parallel (e.g., a genetic locus of Table 1). In some embodiments, genetic loci possess one or more polymorphisms (e.g., a polymorphism of Table 2) the genotypes of which correspond to disease causing alleles. Accordingly, in some embodiments, the disclosure provides methods for assessing multiple heritable disorders in parallel.

In some embodiments, methods are provided for diagnosing multiple heritable disorders in parallel at a pre-implantation, prenatal, perinatal, or postnatal stage. In some embodiments, the disclosure provides methods for analyzing multiple genetic loci (e.g., a plurality of target nucleic acids selected from Table 1) from a patient sample, such as a blood, pre-implantation embryo, chorionic villus or amniotic fluid sample. A patient or subject may be a human. However, aspects of the invention are not limited to humans and may be applied to other species (e.g., mammals, birds, reptiles, other vertebrates or invertebrates) as aspects of the invention are not limited in this respect. A subject or patient may be male or female. In some embodiments, in connection with reproductive genetic counseling, samples from a male and female member of a couple may be analyzed. In some embodiments, for example, in connection with an animal

breeding program, samples from a plurality of male and female subjects may be analyzed to determine compatible or optimal breeding partners or strategies for particular traits or to avoid one or more diseases or conditions.

However, it should be appreciated that any other diseases may be studied and/or risk factors for diseases or disorders including, but not limited to allergies, responsiveness to treatment, cancer tumor profiling for treatment and prognosis, monitoring and identification of patient infections, and monitoring of environmental pathogens.

1. Reducing Representational Bias in Multiplex Amplification Reactions:

In some embodiments, aspects of the invention relate to methods that reduce bias and increase reproducibility in multiplex detection of genetic loci, e.g., for diagnostic purposes.

Molecular inversion probe technology is used to detect or amplify particular nucleic acid sequences in potentially complex mixtures. Use of molecular inversion probes has been demonstrated for detection of single nucleotide polymorphisms (Hardenbol et al. 2005 Genome Res 15:269-75) and for preparative amplification of large sets of exons (Porreca et al. 2007 Nat Methods 4:931-6, Krishnakumar et al. 2008 Proc Natl Acad Sci USA 105:9296-301). One of the main benefits of the method is in its capacity for a high degree of multiplexing, because generally thousands of targets may be captured in a single reaction containing thousands of probes. However, challenges associated with, for example, amplification efficiency (See, e.g., Turner E H, et al., Nat Methods. 2009 Apr. 6:1-2.) have limited the practical utility of the method in research and diagnostic settings.

Aspects of the disclosure are based, in part, on the discovery of effective methods for overcoming challenges associated with systematic errors (bias) in multiplex genomic capture and sequencing methods, namely high variability in target nucleic acid representation and unequal sampling of heterozygous alleles in pools of captured target nucleic acids (e.g., isolated from a biological sample). Accordingly, in some embodiments, the disclosure provides methods that reduce variability in the detection of target nucleic acids in multiplex capture methods. In other embodiments, methods improve allelic representation in a capture pool and, thus, improve variant detection outcomes. In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve the use of different sets of multiple probes (e.g., molecular inversion probes MIPs) that capture overlapping regions of a target nucleic acid to achieve a more uniform representation of the target nucleic acids in a capture pool compared with methods of the prior art. In other embodiments, methods reduce bias, or the risk of bias, associated with large scale parallel capture of genetic loci, e.g., for diagnostic purposes. In other embodiments, methods are provided for increasing reproducibility (e.g., by reducing the effect of polymorphisms on target nucleic acid capture) in the detection of a plurality of genetic loci in parallel. In further embodiments, methods are provided for reducing the effect of probe synthesis and/or probe amplification variability on the analysis of a plurality of genetic loci in parallel.

In some aspects, the disclosure provides probe sets that comprise a plurality of different probes. As used herein, a 'probe' is a nucleic acid having a central region flanked by a 5' region and a 3' region that are complementary to nucleic acids flanking the same strand of a target nucleic acid or subregion thereof. An exemplary probe is a molecular inversion probe (MIP). A 'target nucleic acid' may be a genetic locus.

Exemplary genetic loci are disclosed herein in Table 1 (RefSeqGene Column).

While probes have been typically designed to meet certain constraints (e.g. melting temperature, G/C content, etc.) known to partially affect capture/amplification efficiency (Ball et al (2009) Nat Biotech 27:361-8 AND Deng et al (2009) Nat Biotech 27:353-60), a set of constraints which is sufficient to ensure either largely uniform or highly reproducible capture/amplification efficiency has not previously been achieved.

As disclosed herein, uniformity and reproducibility can be increased by designing multiple probes per target, such that each base in the target is captured by more than one probe. In some embodiments, the disclosure provides multiple MIPs per target to be captured, where each MIP in a set designed for a given target nucleic acid has a central region and a 5' region and 3' region ('targeting arms') which hybridize to (at least partially) different nucleic acids in the target nucleic acid (immediately flanking a subregion of the target nucleic acid). Thus, differences in efficiency between different targeting arms and fill-in sequences may be averaged across multiple MIPs for a single target, which results in more uniform and reproducible capture efficiency.

In some embodiments, the methods involve designing a single probe for each target (a target can be as small as a single base or as large as a kilobase or more of contiguous sequence).

It may be preferable, in some cases, to design probes to capture molecules (e.g., target nucleic acids or subregions thereof) having lengths in the range of 1-200 bp (as used herein, a bp refers to a base pair on a double-stranded nucleic acid—however, where lengths are indicated in bps, it should be appreciated that single-stranded nucleic acids having the same number of bases, as opposed to base pairs, in length also are contemplated by the invention). However, probe design is not so limited. For example, probes can be designed to capture targets having lengths in the range of up to 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 1000, or more bps, in some cases.

It is to be appreciated that the length of a capture molecule (e.g., a target nucleic acid or subregion thereof) is selected based upon multiple considerations. For example, where analysis of a target involves sequencing, e.g., with a next-generation sequencer, the target length should typically match the sequencing read-length so that shotgun library construction is not necessary. However, it should be appreciated that captured nucleic acids may be sequenced using any suitable sequencing technique as aspects of the invention are not limited in this respect.

It is also to be appreciated that some target nucleic acids are too large to be captured with one probe. Consequently, it may be necessary to capture multiple subregions of a target nucleic acid in order to analyze the full target.

In some embodiments, a subregion of a target nucleic acid is at least 1 bp. In other embodiments, a subregion of a target nucleic acid is at least 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 bp or more. In other embodiments, a subregion of a target nucleic acid has a length that is up to 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or more percent of a target nucleic acid length.

The skilled artisan will also appreciate that consideration is made, in the design of MIPs, for the relationship between probe length and target length. In some embodiments, MIPs

are designed such that they are several hundred basepairs (e.g., up to 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 bp or more) longer than corresponding target (e.g., subregion of a target nucleic acid, target nucleic acid).

In some embodiments, lengths of subregions of a target nucleic acid may differ.

For example, if a target nucleic acid contains regions for which probe hybridization is not possible or inefficient, it may be necessary to use probes that capture subregions of one or more different lengths in order to avoid hybridization with problematic nucleic acids and capture nucleic acids that encompass a complete target nucleic acid.

Aspects of the invention involve using multiple probes, e.g., MIPs, to amplify each target nucleic acid. In some embodiments, the set of probes for a given target can be designed to 'tile' across the target, capturing the target as a series of shorter sub targets. In some embodiments, where a set of probes for a given target is designed to 'tile' across the target, some probes in the set capture flanking non-target sequence). Alternately, the set can be designed to 'stagger' the exact positions of the hybridization regions flanking the target, capturing the full target (and in some cases capturing flanking non-target sequence) with multiple probes having different targeting arms, obviating the need for tiling. The particular approach chosen will depend on the nature of the target set. For example, if small regions are to be captured, a staggered-end approach might be appropriate, whereas if longer regions are desired, tiling might be chosen. In all cases, the amount of bias-tolerance for probes targeting pathological loci can be adjusted ('dialed in') by changing the number of different MIPs used to capture a given molecule.

In some embodiments, the 'coverage factor', or number of probes used to capture a basepair in a molecule, is an important parameter to specify. Different numbers of probes per target are indicated depending on whether one is using the tiling approach (see, e.g., FIG. **1**) or one of the staggered approaches (see, e.g., FIG. **2** or **3**).

FIG. **1** illustrates a non-limiting embodiment of a tiled probe layout showing ten captured sub-targets tiled across a single target. Each position in the target is covered by three sub-targets such that MIP performance per base pair is averaged across three probes.

FIG. **2** illustrates a non-limiting embodiment of a staggered probe layout showing the targets captured by a set of three MIPs. Each MIP captures the full target, shown in black, plus (in some cases) additional extra-target sequence, shown in gray, such that the targeting arms of each MIP fall on different sequence. Each position in the target is covered by three sub-targets such that MIP performance per basepair is averaged across three probes. Targeting arms land immediately adjacent to the black or gray regions shown. It should be appreciated that in some embodiments, the targeting arms (not shown) can be designed so that they do not overlap with each other.

FIG. **3** illustrates a non-limiting embodiment of an alternating staggered probe layout showing the targets captured by a set of three MIPs. Each MIP captures the full target, shown in black, plus (in some cases) additional extra-target sequence, shown in gray, such that the targeting arms of each MIP fall on different sequence. Each position in the target is covered by three sub-targets such that MIP performance per basepair is averaged across three probes. Targeting arms land immediately adjacent to the black or gray regions shown.

It should be appreciated that for any of the layouts, the targeting arms on adjacent tiled or staggered probes may be designed to either overlap, not overlap, or overlap for only a subset of the probes.

In certain embodiments for any of the layouts, a coverage factor of about 3 to about 10 is used. However, the methods are not so limited and coverage factors of up to 2, 3, 4, 5, 6, 7, 8, 9, 10, 20 or more may be used. It is to be appreciated that the coverage factor selected may depend the probe layout being employed. For example, in the tiling approach, for a desired coverage factor, the number of probes per target is typically a function of target length, sub-target length, and spacing between adjacent sub-target start locations (step size). For example, for a desired coverage factor of 3, a 200 bp target with a start-site separation of 20 bp and sub-target length of 60 bp may be encompassed with 12 MIPs (FIG. **1**). Thus, a specific coverage factor may be achieved by varying the number of probes per target nucleic acid and the length of the molecules captured. In the staggered approach, a fixed-length target nucleic acid is captured as several subregions or as 'super-targets', which are molecules comprising the target nucleic acid and additional flanking nucleic acids, which may be of varying lengths. For example, a target of 50 bp can be captured at a coverage factor of 3 with 3 probes in either a 'staggered' (FIG. **2**) or 'alternating staggered' configuration (FIG. **3**).

The coverage factor will be driven by the extent to which detection bias is tolerable. In some cases, where the bias tolerance is small, it may be desirable to target more subregions of target nucleic acid with, perhaps, higher coverage factors. In some embodiments, the coverage factor is up to 2, 3, 4, 5, 6, 7, 8, 9, 10 or more.

In some embodiments, when a tiled probe layout is used, when the target length is greater than 1 bp and when a step size (distance between the 5'-end of a target and the 5' end of its adjacent target) is less than the length of a target or subregion thereof, it is possible to compute probe number for a particular target based on target length (T), sub target length (S), and coverage factor (C), such that probe number=T/(S/C)+(C−1).

In some aspects, the disclosure provides methods to increase the uniformity of amplification efficiency when multiple molecules are amplified in parallel; methods to increase the reproducibility of amplification efficiency; methods to reduce the contribution of targeting probe variability to amplification efficiency; methods to reduce the effect on a given target nucleic acid of polymorphisms in probe hybridization regions; and/or methods to simplify downstream workflows when multiplex amplification by MIPs is used as a preparative step for analysis by nucleic acid sequencing.

Polymorphisms in the target nucleic acid under the regions flanking a target can interfere with hybridization, polymerase fill-in, and/or ligation. Furthermore, this may occur for only one allele, resulting in allelic drop-out, which ultimately decreases downstream sequencing accuracy. In some embodiments, using a set of MIPs having multiple hybridization sites for the capture of any given target, the probability of loss from polymorphism is substantially decreased because not all targeting arms in the set of MIPs will cover the location of the mutation.

Probes for MIP capture reactions may be synthesized on programmable microarrays because of the large number of sequences required. Because of the low synthesis yields of these methods, a subsequent amplification step is required to produce sufficient probe for the MIP amplification reaction. The combination of multiplex oligonucleotide synthesis and

pooled amplification results in uneven synthesis error rates and representational biases. By synthesizing multiple probes for each target, variation from these sources may be averaged out because not all probes for a given target will have the same error rates and biases.

Multiplex amplification strategies disclosed herein may be used analytically, as in detection of SNPs, or preparatively, often for next-generation sequencing or other sequencing techniques. In the preparative setting, the output of an amplification reaction is generally the input to a shotgun library protocol, which then becomes the input to the sequencing platform. The shotgun library is necessary in part because next-generation sequencing yields reads significantly shorter than amplicons such as exons. In addition to the bias-reduction afforded by the multi-tiled approach described here, tiling also obviates the need for shotgun library preparation. Since the length of the capture molecule can be specified when the probes, e.g., MIPs, are designed, it can be chosen to match the readlength of the sequencer. In this way, reads can 'walk' across an exon by virtue of the start position of each capture molecule in the probe set for that exon. Reducing analytical errors associated with bias in nucleic acid preparations: In some embodiments, aspects of the invention relate to preparative steps in DNA sequencing-related technologies that reduce bias and increase the reliability and accuracy of downstream quantitative applications.

There are currently many genomics assays that utilize next-generation (polony-based) sequencing to generate data, including genome resequencing, RNA-seq for gene expression, bisulphite sequencing for methylation, and Immune-seq, among others. In order to make quantitative measurements (including genotype calling), these methods utilize the counts of sequencing reads of a given genomic locus as a proxy for the representation of that sequence in the original sample of nucleic acids. The majority of these techniques require a preparative step to construct a high-complexity library of DNA molecules that is representative of a sample of interest. This may include chemical or biochemical treatment of the DNA (e.g., bisulphite treatment), capture of a specific subset of the genome (e.g., padlock probe capture, solution hybridization), and a variety of amplification techniques (e.g., polymerase chain reaction, whole genome amplification, rolling circle amplification).

Systematic and random errors are common problems associated with genome amplification and sequencing library construction techniques. For example, genomic sequencing library may contain an over- or under-representation of particular sequences from a source genome as a result of errors (bias) in the library construction process. Such bias can be particularly problematic when it results in target sequences from a genome being absent or undetectable in the sequencing libraries. For example, an under representation of particular allelic sequences (e.g., heterozygotic alleles) from a genome in a sequencing library can result in an apparent homozygous representation in a sequencing library. As most downstream sequencing library quantification techniques depend on stochastic counting processes, these problems have typically been addressed by sampling enough (over-sampling) to obtain a minimum number of observations necessary to make statistically significant decisions. However, the strategy of oversampling is generally limited to elimination of low-count Poisson noise, and the approach wastes resources and increases the expense required to perform such experiments. Moreover, oversampling can result in a reduced statistical confidence in certain conclusions (e.g., diagnostic calls) based on the data. Accordingly, new approaches are needed for overcoming bias in sequencing library preparatory methods.

Aspects of the disclosure are based, in part, on the discovery of methods for overcoming problems associated with systematic and random errors (bias) in genome capture, amplification and sequencing methods, namely high variability in the capture and amplification of nucleic acids and disproportionate representation of heterozygous alleles in sequencing libraries. Accordingly, in some embodiments, the disclosure provides methods that reduce variability in the capture and amplification of nucleic acids. In other embodiments, the methods improve allelic representation in sequencing libraries and, thus, improve variant detection outcomes. In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve the use of differentiator tag sequences to uniquely tag individual nucleic acid molecules. In some embodiments, the differentiator tag sequence permits the detection of bias based on the frequency with which pairs of differentiator tag and target sequences are observed in a sequencing reaction. In other embodiments, the methods reduce errors caused by bias, or the risk of bias, associated with the capture, amplification and sequencing of genetic loci, e.g., for diagnostic purposes.

Aspects of the invention relate to associating unique sequence tags (referred to as differentiator tag sequences) with individual target molecules that are independently captured and/or analyzed (e.g., prior to amplification or other process that may introduce bias). These tags are useful to distinguish independent target molecules from each other thereby allowing an analysis to be based on a known number of individual target molecules. For example, if each of a plurality of target molecule sequences obtained in an assay is associated with a different differentiator tag, then the target sequences can be considered to be independent of each other and a genotype likelihood can be determined based on this information. In contrast, if each of the plurality of target molecule sequences obtained in the assay is associated with the same differentiator tag, then they probably all originated from the same target molecule due to over-representation (e.g., due to biased amplification) of this target molecule in the assay. This provides less information than the situation where each nucleic acid was associated with a different differentiator tag. In some embodiments, a threshold number of independently isolated molecules (e.g., unique combinations of differentiator tag and target sequences) is analyzed to determine the genotype of a subject.

In some embodiments, the invention relates to compositions comprising pools (libraries) of preparative nucleic acids that each comprise "differentiator tag sequences" for detecting and reducing the effects of bias, and for genotyping target nucleic acid sequences. As used herein, a "differentiator tag sequence" is a sequence of a nucleic acid (a preparative nucleic acid), which in the context of a plurality of different isolated nucleic acids, identifies a unique, independently isolated nucleic acid. Typically, differentiator tag sequences are used to identify the origin of a target nucleic acid at one or more stages of a nucleic acid preparative method. For example, in the context of a multiplex nucleic acid capture reaction, differentiator tag sequences provide a basis for differentiating between multiple independent, target nucleic acid capture events. Also, in the context of a multiplex nucleic acid amplification reaction, differentiator tag sequences provide a basis for differentiating between multiple independent, primary amplicons of a target nucleic acid, for example. Thus, combinations of target nucleic acid

and differentiator tag sequence (target:differentiator tag sequences) of an isolated nucleic acid of a preparative method provide a basis for identifying unique, independently isolated target nucleic acids. FIG. **4**A-C depict various non-limiting examples of methods for combining differentiator tag sequence and target sequences.

It will be apparent to the skilled artisan that differentiator tags may be synthesized using any one of a number of different methods known in the art. For example, differentiator tags may be synthesized by random nucleotide addition.

Differentiator tag sequences are typically of a predefined length, which is selected to control the likelihood of producing unique target:differentiator tag sequences in a preparative reaction (e.g., amplification-based reaction, a circularization selection-based reaction, e.g., a MIP reaction). Differentiator tag sequences may be, up to 5, up to 6, up to 7 up to 8, up to 9, up to 10, up to 11, up to 12, up to 13, up to 14, up to 15, up to 16, up to 17, up to 18, up to 19, up to 20, up to 21, up to 22, up to 23, up to 24, up to 25, or more nucleotides in length. For purposes of genotyping, isolated nucleic acids are identified as independently isolated if they comprise unique combinations of target nucleic acid and differentiator tag sequences, and observance of threshold numbers of unique combinations of target nucleic acid and differentiator tag sequences provide a certain statistical confidence in the genotype.

During a library preparation process, each nucleic acid molecule may be tagged with a unique differentiator tag sequence in a configuration that permits the differentiator tag sequence to be sequenced along with the target nucleic acid sequence of interest (the nucleic acid sequence for which the library is being prepared, e.g., a polymorphic sequence). The incorporation of the nucleic acid comprising a differentiator tag sequence at a particular step allows the detection and correction of biases in subsequent steps of the protocol.

A large library of unique differentiator tag sequences may be created by using degenerate, random-sequence polynucleotides of defined length. The differentiator tag sequences of the polynucleotides may be read at the final stage of the sequencing. The observations of the differentiator tag sequences may be used to detect and correct biases in the final sequencing read-out of the library. For example, the total possible number of differentiator tag sequences, which may be produced, e.g., randomly, is 4N, where N is the length of the differentiator tag sequence. Thus, it is to be understood that the length of the differentiator tag sequence may be adjusted such that the size of the population of MIPs having unique differentiator tag sequences is sufficient to produce a library of MIP capture products in which identical independent combinations of target nucleic acid and differentiator tag sequence are rare. As used herein combinations of target nucleic acid and differentiator tag sequences, may also be referred to as "target:differentiator tag sequences".

In the final readout of a sequencing process, each read may have an additional unique differentiator tag sequence. In some embodiments, when differentiator tag sequences are distributed randomly in a library, all the unique differentiator tag sequences will be observed about an equal number of times. Accordingly, the number of occurrences of a differentiator tag sequence may follow a Poisson distribution.

In some embodiments, overrepresentation of target:differentiator tag sequences in a pool of preparative nucleic acids (e.g., amplified MIP capture products) is indicative of bias in the preparative process (e.g., bias in the amplification process). For example, target:differentiator tag sequence combinations that are statistically overrepresented are indicative of bias in the protocol at one or more steps between the incorporation of the differentiator tag sequences into MIPs and the actual sequencing of the MIP capture products.

The number of reads of a given target:differentiator tag sequence may be indicative (may serve as a proxy) of the amount of that target sequence present in the originating sample. In some embodiments, the numbers of occurrence of sequences in the originating sample is the quantity of interest. For example, using the methods disclosed herein, the occurrence of differentiator tag sequences in a pool of MIPs may be predetermined (e.g., may be the same for all differentiator tag sequences). Accordingly, changes in the occurrence of differentiator tag sequences after amplification and sequencing may be indicative of bias in the protocol. Bias may be corrected to provide an accurate representation of the composition of the original MIP pool, e.g., for diagnostic purposes.

According to some aspects, a library of preparative nucleic acid molecules (e.g., MIPs, each nucleic acid in the library having a unique differentiator tag sequence, may be constructed such that the number of nucleic acid molecules in the library is significantly larger than the number prospective target nucleic acid molecules to be captured using the library. This ensures that products of the preparative methods include only unique target:differentiator tag sequence; e.g., in a MIP reaction the capture step would undersample the total population of unique differentiator tag sequences in the MIP library. For example, an experiment utilizing 1 ug of genomic DNA will contain about ~150,000 copies of a diploid genome. For a MIP library, each MIP in the library comprising a randomly produced 12-mer differentiator tag sequence (~1.6 million possible unique differentiator tag sequences), there would be more than 100 unique differentiator tag sequences per genomic copy. For a MIP library, each MIP in the library comprising a randomly produced 15-mer differentiator tag sequence (~1 billion possible unique differentiator tag sequences), there would be more than 7000 unique differentiator tag sequences per genomic copy. Therefore, the probability of the same differentiator tag sequence being incorporated multiple times is incredibly small. Thus, it is to be appreciated that the length of the differentiator tag sequence is to be selected based on the amount of target sequence in a MIP capture reaction and the desired probability for having multiple, independent occurrences of target:differentiator tag sequence combinations.

FIG. **5** depicts a non-limiting method for genotyping based on target and differentiator tag sequences. Sequencing reads of target and differentiator tags sequences are collapsed to make diploid genotype calls. FIG. **6** depicts non-limiting results of a simulation of a MIP capture reaction in which MIP probes, each having a differentiator tag sequence of 15 nucleotides, are combined with 10000 target sequence copies (e.g., genome equivalents). In this simulated reaction, the probability of capturing one or more copies of a target sequence having the same differentiator tag sequence is 0.05. The Y axis reflects the number of observations. The X axis reflects the number of independent occurrences of target:differentiator tag combinations. FIG. **7** depicts a non-limiting graph of sequencing coverage, which can help ensure that alleles are sampled to sufficient depth (e.g., either 10× or 20× minimum sampling per allele, assuming 1000 targets). In this non-limiting example, the X axis is total per-target coverage required, and the Y axis is the probability that a given total coverage will result in at least 10× or 20× coverage for each allele.

The skilled artisan will appreciate that as part of a MIP library preparation process, adapters may be ligated onto the ends of the molecules of interest. Adapters often contain PCR primer sites (for amplification or emulsion PCR) and/or sequencing primer sites. In addition, barcodes may be included, for example, to uniquely identify individual samples (e.g., patient samples) that may be mixed together. (See, e.g., USPTO Publication Number US 2007/0020640 A1 (McCloskey et al.)

The actual incorporation of the random differentiator tag sequences can be performed through various methods known in the art. For example, nucleic acids comprising differentiator tag sequences may be incorporated by ligation. This is a flexible method, because molecules having differentiator tag sequence can be ligated to any blunt-ended nucleic acids. The sequencing primers must be incorporated subsequently such that they sequence both the differentiator tag sequence and the target sequence. Alternatively, the sequencing adaptors can be synthesized with the random differentiator tag sequences at their 3' end (as degenerate bases), so that only one ligation must be performed. Another method is to incorporate the differentiator tag sequence into a PCR primer, such that the primer structure is arranged with the common adaptor sequence followed by the random differentiator tag sequence followed by the PCR priming sequence (in 5' to 3' order). A differentiator tag sequence and adaptor sequence (which may contain the sequencing primer site) are incorporated as tags. Another method to incorporate the differentiator tag sequences is to synthesize them into a padlock probe prior to performing a gene capture reaction. The differentiator tag sequence is incorporated 3' to the targeting arm but 5' to the amplification primer that will be used downstream in the protocol. Another method to incorporate the differentiator tag sequences is as a tag on a gene-specific or poly-dT reverse-transcription primer. This allows the differentiator tag sequence to be incorporated directly at the cDNA level.

In some embodiments, at the incorporation step, the distribution of differentiator tag sequences can be assumed to be uniform. In this case, bias in any part of the protocol would change the uniformity of this distribution, which can be observed after sequencing. This allows the differentiator tag sequence to be used in any preparative process where the ultimate output is sequencing of many molecules in parallel.

Differentiator tag sequences may be incorporated into probes (e.g., MIPs) of a plurality when they are synthesized on-chip in parallel, such that degeneracy of the incorporated nucleotides is sufficient to ensure near-uniform distribution in the plurality of probes. It is to be appreciated that amplification of a pool of unique differentiator tag sequences may itself introduce bias in the initial pool. However, in most practical cases, the scale of synthesis (e.g., by column synthesis, chip based synthesis, etc.) is large enough that amplification of an initial pool of differentiator tag sequences is not necessary. By avoiding amplification or selection steps on the pool of unique differentiator tag sequences, potential bias may be minimized. One example of the use of the differentiator tag sequences is in genome re-sequencing.

Considering that the raw accuracy of most next-generation sequencing instruments is relatively low, it is crucial to oversample the genomic loci of interest.

Furthermore, since there are two alleles at every locus, it is important to sample enough to ensure that both alleles have been observed a sufficient number of times to determine with a sufficient degree of statistical confidence whether the sample is homozygous or heterozygous. Indeed, the sequencing is performed to sample the composition of molecules in the originating sample. However, after multiple reads have been collected for a given locus, it is possible that due to bias (e.g., caused by PCR amplification steps), a large fraction of the reads are derived from a single originating molecule. This would skew the population of target sequences observed, and would affect the outcome of the genotype call. For example, it is possible that a locus that is heterozygous is called as homozygous, because there are only a few observations of the second allele out of many observations of that locus. However, if information is available on differentiator tag sequences, this situation could be averted, because the over-represented allele would be seen to also have an over-represented differentiator tag sequence (i.e., the sequences with the overrepresented differentiator tag sequence all originated from the same single molecule). Therefore, the sequences and corresponding distribution of differentiator tag sequences can be used as an additional input to the genotype-calling algorithm to significantly improve the accuracy and confidence of the genotype calls.

In some aspects, the disclosure provides methods for analyzing a plurality of target sequences which are genetic loci or portions of genetic loci (e.g., a genetic locus of Table 1). The genetic loci may be analyzed by sequencing to obtain a genotype at one or more polymorphisms (e.g., SNPs). Exemplary polymorphisms are disclosed in Table 2. The skilled artisan will appreciate that other polymorphisms are known in the art and may be identified, for example, by querying the Entrez Single Nucleotide Polymorphism database, for example, by searching with a GeneID from Table 1.

TABLE 1

Target Nucleic Acids

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
CYP21 A2	1589	cytochrome P450, family 21, subfamily A, polypeptide 2	CAH1; CPS1; CA21H; CYP21; CYP21B; P450c21B; MGC150536; MGC150537; CYP21 A2	201910	NG_008337,1	6p21.3
ABCC8	6833	ATP-binding cassette, sub-family C (CFTR/MRP), member 8	HI; SUR; HHF1; MRP8; PHHI; SUR1; ABC36; HRINS; TNDM2; ABCC8	600509	NG_008867.1	11p15.1

TABLE 1-continued

Target Nucleic Acids

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
ATRX	546	alpha thalassemia/mental retardation syndrome X-linked (RAD54 homolog, *S. cerevisiae*)	SHS; XH2; XNP; ATR2; SFM1; RAD54; MRXHF1; RAD54L; ZNF-HX; MGC2094; ATRX	300032	NG_008838.1	Xq13.1-q21.1
ARSA	410	arylsulfatase A	MLD; ARSA	607574	NG_009260.1	22q13.31-qter; 22q13.33
PSAP	5660	Prosaposin	GLBA; SAP1; FLJ00245; MGC110993; PSAP	176801	NG_008835.1	10q21-q22
BTD	686	Biotinidase	BTD	609019	NG_008019.1	3p25
HLCS	3141	holocarboxylase synthetase (biotin-(proprionyl-Coenzyme A-carboxylase (ATP-hydrolysing)) ligase)	HCS; HLCS	609018	NC_000021.7	21q22.1; 21q22.13
BLM	641	Bloom syndrome, RecQ helicase-like	BS; RECQ2; RECQL2; RECQL3; MGC126616; MGC131618; MGC131620; BLM	604610	NG_007272.1	15q26.1
ASPA	443	aspartoacylase (Canavan disease)	ASP; ACY2; ASPA	608034	NG_008399.1	17pter-P13
CFTR	1080	cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7)	CF; MRP7; ABC35; ABCC7; CFTR/MRP; TNR-CFTR; dJ760C5.1; CFTR	602421	NC_000007.12	7q31.2
ASS1	445	argininosuccinate synthetase 1	ASS; CTLN1; ASS1	603470	NG_011542.1	9q34.1
MMACHC	25974	methylmalonic aciduria (cobalamin deficiency) cb1C type, with homocystinuria	cb1C; FLJ25671; DKFZp564I122; RP11-291L19.3; MMACWC	609831	NC_000001.9	1p34.1
IKBKAP	8518	inhibitor of kappa light polypeptide gene enhancer in B-cclls, kinase complex-associated protein	FD; DYS; ELP1; IKAP; IKI3; TOT1; FLJ12497; DKFZp781H1425; IKBKAP	603722	NG_008788.1	9q31
FANCC	2176	Fanconi anemia, complementation group C	FA3; FAC; FACC; FLJ14675; FANCC	227645	NG_011707.1	9q22.3
GALK1	2584	galactokinase 1	GK1; GALK; GALK1	604313	NG_008079.1	17q24
GALT	2592	galactose-1-phosphate uridylyltransferase	GALT	606999	NC_000009.10	9p13
GALE	2582	UDP-galactose-4-epimcrase	SDR1E1; FLJ95174; FLJ97302; GALE	606953	NG_007068.1	1p36-p35
GBA	2629	glucosidase, beta; acid (includes glucosylceramidase)	GCB; GBA1; GLUC; GBA	606463	NG_009783.1	1q21
GJB2	2706	gap junction protein, beta 2, 26 kDa	HID; KID; PPK; CX26; DFNA3; DFNB1; NSRD1; DFNA3A; DFNB1A; GJB2	121011	NG_008358.1	13q11-q12

TABLE 1-continued

Target Nucleic Acids

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
GCDH	2639	glutaryl-Cocnzyme A dehydrogenase	GCD; ACAD5; GCDH	608801	NG_009292.1	19p13.2
G6PC	2538	glucose-6-phosphatase, catalytic subunit	G6PT; GSD1; GSD1a; MGCI63350; G6PC	232200	NG_011808.1	17q21
HBB	3043	hemoglobin, beta	CD113t-C; beta-globin; HBB	141900	NG_000007.3	11p15.5
BCKDHA	593	branched chain keto acid dehydrogenase E1, alpha polypeptide	MSU; MSUD1; OVD1A; BCKDE1A; FLJ45695; BCKDHA	608348	NC_000019.8	19q13.1-q13.2
BCKDHB	594	branched chain keto acid dehydrogenase E1 , beta polypeptide	E1B; FLJ17880; dJ279A18.1; BCKDHB	248611	NG_009775.1	6q13-q15
DBT	1629	dihydrolipoamide branched chain transacylase E2	E2; E2B; BCATE2; MGC9061; DBT	248610	NG_011852.1	1p31
DLD	1738	dihydrolipoamide dehydrogenase	E3; LAD; DLDH; GCSL; PHE3; DLD	238331	NG_008045.1	7q31-q32
ACADM	34	acyl-Coenzyme A dehydrogenase, C-4 to C-12 straight chain	MCAD; ACAD1; MCADH; FLJ18227; FLJ93013; FLJ99884; ACADM	607008	NG_007045.1	1p31
MMAA	166785	methylmalonic aciduria (cobalamin deficiency) cblA type	cblA; MGC120010; MGC120011; MGC120012; MGC120013; MMAA	607481	NG_007536.1	4q31.22
MMAB	326625	methylmalonic aciduria (cobalamin deficiency) cblB type	ATR; cblB; MGC20496; MMAB	607568	NG_007096.1	12q24
MUT	4594	methylmalonyl Coenzyme A mutase	MCM; MUT	609058	NG_007100.1	6p12.3
MCOLN1	57192	mucolipin 1	ML4; MLIV; MST080; TRPML1; MSTP080; TRP-ML1; TRPM-L1 ; MCOLN1	605248	NC_000019.8	19p13.3-p13.2
ACTA1	58	actin, alpha 1, skeletal muscle	ACTA; ASMA; CFTD; MPFD; NEM1; NEM2; NEM3; CFTD1; CFTDM; ACTA1	102610	NG_006672.1	1q42.13
TPM3	7170	tropomyosin 3	TM3; TRK; NEM1; TM-5; TM30; TM30 nm; TPMsk3; hscp30; MGC3261; FLJ41118; MGC14582; MGC72094; OK/SW-c1.5; TPM3	191030	NG_008621.1	1q21.2
TNNT1	7138	troponin T type 1 (skeletal, slow)	ANM; TNT; STNT; TNTS; FLJ98147; MGC104241; TNNT1	191041	NG_011829.1	19q13.4
NEB	4703	nebulin	NEM2; NEB177D; FLJ11505; FLJ36536; FLJ39568; FLJ39584; DKFZp686C1456; NEB	161650	NG_009382.1	2q22
SMPD1	6609	sphingomyelin phosphodiesterase 1, acid lysosomal	ASM; NPD; SMPD1	607608	NG_011780.1	11p15.4-p15.1
GLDC	2731	glycine dehydrogenase (decarboxylating)	GCE; NKH; GCSP; HYGN1; MGC138198; MGC138200; GLDC	238300	NC_000009.10	9p22

TABLE 1-continued

Target Nucleic Acids

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
GCSH	2653	glycine cleavage system protein H (aminomethyl carrier)	GCE; NKH; GCSH	238330	NC_000016.8	16q23.2
AMT	275	aminomethyltransferase	GCE; NKH; GCST; AMT	238310	NC_000003.10	3p21.2-p21.1
OTC	5009	ornithine carbamoyltransferase	OCTD; MGC129967; MGC129968; MGC138856; OTC	300461	NG_008471.1	Xp21.1
PAH	5053	phenylalanine hydroxylase	PH; PKU; PKU1; PAH	612349	NG_008690.1	12g22-q24.2
DHPR	5860	quinoid dihydropteridine reductase	DHPR; PKU2; SDR33C1; FLJ42391; QDPR	612676	NG_008763.1	4p15.31
PTS	5805	6-pyruvoyltetrahydropterin synthase	PTPS; FLJ97081; PTS	261640	NG_008743.1	11q22.3-q23.3
PCCA	5095	propionyl Coenzyme A carboxylase, alpha polypeptide	PCCA	232000	NG_008768.1	13q32
PCCB	5096	propionyl Coenzyme A carboxylase, beta polypeptide	DKFZp451E113; PCCB	232050	NG_008939.1	3q21-q22
ACADS	35	acyl-Coenzyme A dehydrogenase, C-2 to C-3 short chain	SCAD; ACAD3; ACADS	606885	NG_007991.1	12g22-qter
DHCR7	1717	7-dehydrocholesterol reductase	SLOS; DHCR7	602858	NC_000011.8	11q13.2-q13.5
SMNT	6606	survival of motor neuron 1, telomeric	SMA; SMN; SMA1; SMA2; SMA3; SMA4; SMA@; SMNT; BCD541; T-BCD541; SMN1	600354	NG_008691.1	5q13
HEXA	3073	hexosaminidase A (alpha polypeptide)	TSD; MGC99608; HEXA	606869	NG_009017.1	15g23-q24
MYO7A	4647	myosin VIIA	DFNB2; MYU7A; NSRD2; USH1B; DFNA11; MYOVIIA; MYO7A	276903	NG_009086.1	11q13.5
USH1C	10083	Usher syndrome 1 C (autosomal recessive, severe)	PDZ73; AIE-75; DFNB18; PDZ-45; PDZ-73; NY-CO-37; NY-CO-38; ush1cpst; PDZ-73/NY-CO-38; USH1C	605242	NC_000011.8	11p15.1-p14
CDH23	64072	cadherin-like 23	USH1D; DFNB12; FLJ00233; FLJ36499; KIAA1774; KIAA1812; MGC102761; DKFZp434P2350; CDH23	605516	NG_008835.1	10g21-q22
PCDH15	65217	protocadherin 15	USH1F; DFNB23; DKFZp667A1711; PCDH15	605514	NG_009191.1	10q21.1
SANS	124590	Usher syndrome 1G (autosomal recessive)	SANS; ANKS4A; FLJ33924; USH1G	607696	NG_007882.1	17q25.1

TABLE 1-continued

Target Nucleic Acids

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
ARX	170302	aristaless related homeobox	ISSX; PRTS; MRX29; MRX32; MRX33; MRX36; MRX38; MRX43; MRX54; MRX76; MRX87; MRXS1; ARX	300382	NG_008281.1	Xp21
OPHN1	4983	oligophrenin 1	OPN1; MRX60; OPHN1	300127	NG_008960.1	Xq12
JAR1DIC	8242	lysine (K)-specific demethylase 5C	MRXJ; SMCX; MRXSJ; XE169; JARID1C; DXS1272E; KDM5C	314690	NG_008085.1	Xp11.22-p11.21
FTSJ1	24140	FtsJ homolog 1 (*E. coli*)	JM23; MRX9; SPB1; TRM7; CDLIV; MRX44; FTSJ1	300499	NG_008879.1	Xp11.23
SLC6A8	6535	solute carrier family 6 (neurotransmitter transporter, creatine), member 8	CRT; CT1; CRTR; MGC87396; SLC6A8	300036	NC_000023.9	Xq28
DLG3	1741	discs, large homolog 3 (*Drosophila*)	MRX; MRX90; NEDLG; NE-Dlg; SAP102; SAP-102; KIAA1232; DLG3	300189	NC_000023.9	Xq13.1
TM4SF2	7102	letraspanin 7	A15; MXS1; CD231; MRX58; CCG-B7; TM4SF2; TALLA-1; TM4SF2b; DXS1692E; TSPAN7	300096	NG_009160.1	Xp11.4
ZNF41	7592	zinc finger protein 41	MRX89; MGC8941; ZNF41	314995	NG_008238.1	Xp11.23
FACL4	2182	acyl-CoA synthetase long-chain family member 4	ACS4; FACL4; LACS4; MRX63; MRX68; ACSL4	300157	NG_008053.1	Xq22.3-q23
PQBP1	10084	polyglutamine binding protein 1	SHS; MRX55; MRXS3; MRXS8; NPW38; RENS1; PQBP1	300463	NC_000023.9	Xp11.23
PEX1	5189	peroxisomal biogenesis factor 1	ZWS1; PEX1	602136	NG_008341.1	7q21.2
PXMP3	5828	peroxisomal membrane protein 3, 35 kDa	PAF1; PEX2; PMP3; PAF-1; PMP35; RNF72; PXMP3	170993	NG_008371.1	8q21.1
PEX6	5190	peroxisomal biogenesis factor 6	PAF2; PAF-2; PXAAA1; PEX6	601498	NG_008370.1	6p21.1
PEX10	5192	peroxisomal biogenesis factor 10	NALD; RNF69; MGC1998; PEX10	602859	NG_008342.1	1p36.32
PEX12	5193	peroxisomal biogenesis factor 12	PAF-3; PEX12	601758	NG_008447.1	17q12

TABLE 1-continued

Target Nucleic Acids

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
PEX5	5830	peroxisomal biogenesis factor 5	PXR1; PTS1R; PTS1-BP; FLJ50634; FLJ50721; FLJ51948; PEX5	600414	NG_008448.1	12p13.31
PEX26	55670	peroxisomal biogenesis factor 26	FLJ20695; PEX26M1T; Pex26pM1T; PEX26	608666	NG_008339.1	22q11.21

The mutations listed in Table 2 are documented polymorphisms in several disease-associated genes (CFTR is mutated in cystic fibrosis, GBA is mutated in Gaucher disease, ASPA is mutated in Canavan disease, HEXA is mutated in Tay Sachs disease). The polymorphisms are of several types: insertion/deletion polymorphisms which will cause frameshifts (and thus generally interrupt protein function) unless the insertion/deletion length is a multiple of 3 bp, and substitutions which can alter the amino acid sequence of the protein and in some cases cause complete inactivation by introduction of a stop codon.

TABLE 2

Non-limiting examples of polymorphisms

Gene name	GeneID	SNP ID	Mutation	SEQ ID NO:
CFTR	1080	rs63500661	TCACATCACCAAGTTAAAAAAAAAAA[A/G]G GGGCGGGGGGGCAGAATGAAAATT	1
CFTR	1080	rs63107760	AAACAAGGATGAATTAAGTTTTTTTT[-/T] AAAAAAGAAACATTTGGTAAGGGGA	2
CFTR	1080	rs62469443	ATCACCAAGTTAAAAAAAAAAAAGGG[A/G]C GGGGGGGCAGAATGAAAATTGCAT	3
CFTR	1080	rs62469442	CTATTGAACCAGAACCAAACAGGAAT[A/G]C CATAGCATTTTGTAAACTAAACTG	4
CFTR	1080	rs62469441	CAGGAGTTCAAGACCAGCCTACTAAA[A/C]C ACACACACACACACACACACAC	5
CFTR	1080	rs62469439	GATTAAATAATAGTGTTTATGTACCC[C/G]GC TTATAGGAGAAGAGGGTGTGTGT	6
CFTR	1080	rs62469438	ATTGTTATCTTTTCATATAAGGTAAC[A/T]GA GGCCCAGAGAGATTAAATAACAT	7
CFTR	1080	rs62469437	TAATTTTAATTAAGTAAATTTAATTG[A/G]TA GATAAATAAGTAGATAAAAAATA	8
CFTR	1080	rs62469436	GTATAAAAAAAAAAAAAAAAAAAAGTT[A/T]G AATGTTTTCTTGCATTCAGAGCCT	9
CFTR	1080	rs62469435	ATACTAAAAATTTAAAGTTCTCTTGC[A/G]AT ATATTTTCTTAATATCTTACATC	10
CFTR	1080	rs62469434	TGCTGGGATTACAGGCGTGAGCCACC[A/G]C GCCTGGCCTGATGGGACATATTTT	11
CFTR	1080	rs62469433	CTACAATATAAGTATAGTATTGCAAA[A/C]CC ATCAGGAAGGGTGTTAACTATTT	12
CFTR	1080	rs61763210	GTTGTCTCCAAACTTTTTTTCAGGTG[-/AGA] AGGTGGCCAACCGAGCTTCGGAAAG	13
CFTR	1080	rs61720488	TTTTTTCATAAAAGATTATATAAAGG[A/C]TA TTGCTTTTGAATCACAAACACTA	14
CFTR	1080	rs61481156	ATCTAGTGAGCAGTCAGGAAAGAGAA[C/T]T TCCAGATCCTGGAAATCAGGGTTA	15
CFTR	1080	rs61443875	TAGAGTATAAAAAAAAAAAAAAAAAA[-/A] GTTTGAATGTTTTCTTGCATTCAGA	16
CFTR	1080	rs61312222	TGCAAATGCCAACTATCAAAGATATT[C/G]GA GTATACTGTCAATAAACTTCATA	17

TABLE 2-continued

Non-limiting examples of polymorphisms

Gene name	GeneID	SNP ID	Mutation	SEQ ID NO:
CFTR	1080	rs61159372	TCCTCAACAGTTAGAAACAATATTTT[C/G]AG TGATTTCCCATGCCAACTTTACT	18
CFTR	1080	rs61094145	TTTTTGGTATTGTTGTTAAATAAGTG[A/G]GA ATTCAATACAGTATAATGTCTGT	19
CFTR	1080	rs61086387	CTTGAAATCGGATATATATATATATA[-/T GTATATATATATATATATATATATATATAT ACATATATATATATA]GTATTATCCCTGTTTTC ACAGTTTT	20
CFTR	1080	rs60996744	AGAGGGGCTGTGAAGGACACCAAGGA[A/G]G AGACTAAGAGCCAGGAGGGAAAAC	21
CFTR	1080	rs60960860	TAGAGTTTATTAGCTTTTACTACTCT[A/G]CTT AGTTACTTTGTGTTACAGAATA	22
CFTR	1080	rs60923902	ACTAGTGATGATGAGCTTCTTTTCAT[-/AT] GTTTGTTGGCTGCATAAATGTCTTC	23
CFTR	1080	rs60912824	GCAGAGAAAAGAGGGGCTGTGAAGGA[C/G]A CCAAGGAGGAGACTAAGAGCCAGG	24
CFTR	1080	rs60887846	TTCAGAGGTCTACCACTGGTGCATAC[G/T]CT AATCACAGTGTCGAAAATTTTAC	25
CFTR	1080	rs60793174	AAGAAAGAGCAAAAGAGGGCAAACTT[C/T]T CATACATTTTTGATGTCGAAACCA	26
CFTR	1080	rs60788575	CCTAAAGTTTAAAAAGAAAAAAAAAA[-/A] GGAAGAAGGAATTAAAAATCCAAAG	27
CFTR	1080	rs60760741	GTGTGTGTGTGTATATATATATATAT[A/T]TA TATATTTTTTTTTTCCTGAGCCA	28
CFTR	1080	rs60456599	AAACTGTTGATGTTTTCATTTATTTA[C/G]ATC ATTGGAAAACTTTAGATTCTAG	29
CFTR	1080	rs60363249	TTTATCCATTCTTAACCAGAACAGAC[A/G]TT TTTTCAGAGCTGGTCCAGGAAAA	30
CFTR	1080	rs60355115	TTGAAATCGGATATATATATATATAT[A/G]TA TATATATATATATATATATATAT	31
CFTR	1080	rs60308689	TAGTTTTTTATTTCCTCATATTATTT[-/T] CAGTGGCTTTTTTCTTCCACATCTTT	32
CFTR	1080	rs60271242	ACATAGTTCTCAGTGGTACAACTACA[A/G]GT GATTTCTCTTTTCTTATTTCTGG	33
CFTR	1080	rs60010318	AGAGCAATGGCATCCCTTGTCTTGTG[C/T]TA TACAGGATGCAGCAATTTATAGG	34
CFTR	1080	rs59961323	TTCTGTCTACATAAGATGTCATACTA[A/G]AT TATCTTTTCCAGCATGCATTCAG	35
CFTR	1080	rs59961270	CAGGGTGGCATGTTAGGCAGTGCTTA[A/G]A ATAAATGAGTTGGTTATACAAGTA	36
CFTR	1080	rs59837506	AGGACACACACACACACACACACACA[-/CA] TGCACACACATTTAAATAGATGCAT	37
CFTR	1080	rs59572090	TAAAAAATTGGTATAATGAAATTGCA[C/T]TT GTAGTCTTTGGACATTTAAATCC	38
CFTR	1080	rs59548252	TTTCAATACTTAAGAGGTACGCAGAG[A/G]A AAGAGGGGCTGTGAAGGACACCAA	39
CFTR	1080	rs59519859	CAGCAATGAATATTTTGAGGCTGAGG[C/T]GC TGAGGGGTAAAATTGCAGCCTGG	40
CFTR	1080	rs59509837	TTATGGTTTATATTTTGTGTCTTCT[-/C TTT]AACACATCTTTTCTAGCAGAATTCA	41

TABLE 2-continued

Non-limiting examples of polymorphisms

Gene name	GeneID	SNP ID	Mutation	SEQ ID NO:
CFTR	1080	rs59417037	GTATTTTAGTTTTTTTTTTTTGTTTG[-/T] TTTGTTTTGTTTTGTTTTGTTTTTG	42
CFTR	1080	rs59159458	TGGGTGACTCCATTTTTACTTTTAGT[C/T]TGG TCTGTTGAGGCCTCGTGAGAGA	43
CFTR	1080	rs59048119	TATTTTCATGTATTTTAGTTTTTTTTTT[-/T TTT]GTTTGTTTTGTTTTGTTTTGTTTTG	44
CFTR	1080	rs58970500	GTGTGTGTGTATATATATATATATAT[A/T]TA TATTTTTTTTTTCCTGAGCCAAA	45
CFTR	1080	rs58942292	AACCTATTAGCATGTCTGGCAGAAAA[-/A] TAGATACTTAATAAATTTCTTAAAT	46
CFTR	1080	rs58917054	GAGGCTTAGACAGTTTAAGTAACTCA[A/G]G CATGGTTACACAACTAGCTAGGGC	47
CFTR	1080	rs58837484	GTGTGAGTATTATGAGACCATATGTT[A/G]GG AGATTTTATTTGGTATTGAGGAT	48
CFTR	1080	rs58829491	GAAACCCCACCCCTTCTATAGTTTTC[C/T]CTT TAATATTTACAATGGAACCATT	49
CFTR	1080	rs58805195	CATATATATATAGTGTGTGTGTGT[A/G]TA TATATATATATATATATTTTT	50
GBA	2629	rs60866785	CGAGCGAGAGAGAGAGAGAGAGAG[-/AG] GAGCCGGCGCGAGAACTACGCATGC	51
GBA	2629	rs60239603	GGCAGGTAATATCTAGTACCTTACTT[A/T]TA TTTCCTGAGCACATTCTACATTT	52
GBA	2629	rs56310840	GGCCAGGAATGGGAGTGCTTAGGTGC[A/G]G AGGTGGCACTGTTCCCGCAGCTGC	53
GBA	2629	rs41264927	GAAAACTCCATCCCCTCAGGGTCAT[C/T]AG ATGAAGAGAAGACCACAGGGGTT	54
GBA	2629	rs41264925	TGTAGGTAAGGGTCACATGTGGGAGA[C/G]G CAGCTGTGGGTAGGTCAGCCCTGT	55
GBA	2629	rs36024691	CCAAGAAGGCGCCATTACACTCCAGC[-/C] TGGGCGACAGGGCGAGACTCCCTCA	56
GBA	2629	rs36024092	TGCCACACCCAGCTAATTTGTGTGTG[-/G] TATGTGTGTGTATGTATGTGTGTGT	57
GBA	2629	rs35682967	GTTCCTCCAGTAATTTTTTTTTTTTT[-T/] GGTTTTGAGACAGAGTCTTGCCCTG	58
GBA	2629	rs35033592	ATCATGCCCAGATAATTTTTTTTTTT[-/T] GTATTTTAGTAGACACAGGGTTTCA	59
GBA	2629	rs34732744	CGAGCGAGAGAGAGAGAGAGAGAGAG[-/AG] GAGCCGGCGCGAGAACTACGCATGC	60
GBA	2629	rs34620635	CCTGTGAGGGGCACATTCCTTAGTAG[-/C] TAAGGAGTTGGGGGTGTGAAGATCC	61
GBA	2629	rs34302637	ACAGGCTACTGGCTGGGCCCAGGCAA[-/A] GGGGGCCTTGGCAGGAAAAGTTCCT	62
GBA	2629	rs33949225	GCGAGAGAGAGAGAGAGAGAGAGAGG[-/AG] AGCCGGCGCGAGAACTACGCATGCG	63
GBA	2629	rs28678003	AAGAAGAAAAATAAAAAGAAAGTGGG[C/T]C AGACCGAGAGAACAGGAAGCCTGA	64
GBA	2629	rs28559737	AAGGACAAAGGCAAAGAGACAAAGGC[G/T]C AACACTGGGGGTCCCCAGAGAGTG	65
GBA	2629	rs28373017	TACCTAGTCACTTCCTGCCTCCATGG[C/T]GC AAAAGGGGATGGGTGTGCCTCTT	66

TABLE 2-continued

Non-limiting examples of polymorphisms

Gene name	GeneID	SNP ID	Mutation	SEQ ID NO:
GBA	2629	rs12752133	CTCTTCCGAGGTTCCACCCTGAACAC[C/T]TT CCTGCTCCCTCGTGGTGTAGAGT	67
GBA	2629	rs12747811	TTCTGACTGGCAACCAGCCCCACTCT[C/T]TG GGAGCCCTCAGGAATGAACTTGC	68
GBA	2629	rs12743554	gctcagcctcccaggctggagtgcag[A/T]ggcgcgatc tcggctcaccgcaacc	69
GBA	2629	rs12041778	CATGAACCACATCAAATGAGATTTAG[C/T]GG GAGTGGCACACACAGTCATGACC	70
GBA	2629	rs12034326	AAGCAGCCCTGGGGAGTCGGGGCGGG[A/G]C CTGGATTGGAAAAGAGACGGTCAC	71
GBA	2629	rs11558184	CTCCAAGTTCTGGGAGCAGAGTGTGC[A/G]G CTAGGCTCCTGGGATCGAGGGATG	72
GBA	2629	rs11430678	GTTCCTCCAGTAAttttttttttttt[-/G/T] gttttgagacagagtcttgccctgt	73
GBA	2629	rs11264345	CTAGTACCTTACTTCCCTCAAGTTCA[A/T]TC ATCTCACAGATATTTCCTGAGCA	74
GBA	2629	rs10908459	aattagccgtgcgtggtggcgggtgc[C/T]tgtaatccc acgtacttgggaggct	75
GBA	2629	rs10796940	CCATGGCCAGCCGGGGAGGGGACGGG[A/C]A CACACAGACCCACACAGAGACTCA	76
GBA	2629	rs10668496	agcgagagagagagagagagagagag[-/AG] gagCCGGCGCGAGAACTACGCATGC	77
GBA	2629	rs7416991	CGTAGCAGTTAGCAGATGATAGGCGG[C/G/T] GAAATCTTATTTCACAGGGCATTAA	78
GBA	2629	rs4024049	CTGGCCCTGGTGACAGTGGGGCTGTG[C/T]GT GGGGCCAGAGCCTTCTCAGAGGT	79
GBA	2629	rs4024048	CAGATACTGGCCCTGGTGACAGTGGG[A/G]C TGTGCGTGGGGCCAGAGCCTTCTC	80
GBA	2629	rs4024047	GACAGATACTGGCCCTGGTGACAGTG[G/T]G GCTGTGCGTGGGGCCAGAGCCTTC	81
GBA	2629	rs3841430	GGCTCctctctctctctctctctc[-/TC] gctcgctctctcgctctctcgctct	82
GBA	2629	rs3754485	GTTTCAGACCAGCCTGGCCAACATAG[C/T]GA AACCCCATCTCTACTAAAAATAA	83
GBA	2629	rs3205619	AGTGGGCGATTGGATGGAGCTGAGTA[C/T]G GGGCCCATCCAGGCTAATCACACC	84
GBA	2629	rs2990227	CCGGGCTCCGTGAATGTTTGTCACAT[C/G]TC TGAAGAACGTATGAATTACATAA	85
GBA	2629	rs2990226	GAATCCCAACCCCGACGCTCGTCGCC[C/G]G GCTCCGTGAATGTTTGTCACATGT	86
GBA	2629	rs2990225	GCGAATCCCAACCCCGACGCTCGTCG[C/T]CG GGCTCCGTGAATGTTTGTCACAT	87
GBA	2629	rs2990224	TGGGCAGAAGTCAGGGTCCAAAGAAA[G/T]G GCAAAGAAAAGTGTcagtggctca	88
ASPA	443	rs63751297	TAAGAAAGACGTTTTTGATTTTTTTC[A/G]GA CTTCTCTGGCTCCACTACCCTGC	89
ASPA	443	rs62071301	CTGATTCCTGGCCAGGAGCGGTGGCT[C/T]AC GCCTGTAATCCCAGCGCTTTGGG	90
ASPA	443	rs62071300	TAAAAATGCTGATTCCTGGCCAGGAG[C/T]GG TGGCTCACGCCTGTAATCCCAGC	91

TABLE 2-continued

Non-limiting examples of polymorphisms

Gene name	GeneID	SNP ID	Mutation	SEQ ID NO:
ASPA	443	rs62071299	TTTAAAAATGCTGATTCCTGGCCAGG[A/C]GC GGTGGCTCACGCCTGTAATCCCA	92
ASPA	443	rs62071297	CAAGACCTGTCAAAGATCTGAGAAAT[A/T]TT ACCCGACTTACAAGCTAACCATT	93
ASPA	443	rs61697033	ACTGTAATAAGTGCTGTAAAAGAAAT[A/G]C ACAAAATAATATAGCAGAGGGTAT	94
ASPA	443	rs60743592	CTTGAGGTCAGGAGTTCAAGACCAGT[C/T]TG GGCAACATGGGGAAAACCTTGTC	95
ASPA	443	rs60666840	AGGTTGCAGTGAGCCGAGATCATGCC[A/G]TT GCACTCCAGCCGGGGCAACAAAA	96
ASPA	443	rs60147514	ACAAGTGTCTTGAAATTATCTGTGAT[C/T]TG CTATAGAGCAATACTTTTGTAAA	97
ASPA	443	rs59930743	GTGGGTATATGCAGCTCTATGCACTA[C/T]CT GCTCATTTATTTGGTAAATCTAA	98
ASPA	443	rs59690349	TGTGTGTGTGTGCGTGTGTGTGTGTG[-/T GTGTGTG]ATCATAAGAGTGGCTGCAGCAA ACT	99
ASPA	443	rs59676360	AGTCTGGAGTGCAATGGTGCAATCTC[A/G]GC TCACTGCAGCCTCCACCTCCGGG	100
ASPA	443	rs59335404	CTCCTAATGGATATTTCCTAAATTTT[G/T]CTG AACAGAATTTAACTTGAGCTGG	101
ASPA	443	rs58879097	ATTTAAAAATGGATTTCTAGAAAAAC[A/G]AT CACATACTTGAATATTTTAGCAA	102
ASPA	443	rs58686774	CTATAAATGGGTAGCATGAGGGATTC[A/G]A GGAGGTGGCTGAAAGAAGCACGTA	103
ASPA	443	rs57511162	AAGAAACCAAGCATAGTAGAGTGTTA[A/G]A AAACCAAAGCAACTAAACAACTGT	104
ASPA	443	rs55859596	CGGGGCTCAGAACTTGTAACAGAAAA[A/T]T AAAATATACTCCACTCAAGGGAAT	105
ASPA	443	rs55742972	TACTACACTTCACGGATACTGTACTT[-/G TACTT]TTTTTCCAAATTGAAGGTTTTTGGC	106
ASPA	443	rs55640436	TTGTTTTTGTTTTTGTTTTTGTTTTT[-/G TTTTTGTTTTT]TGAGATGGAGTCTCGCTCT GTCGCC	107
ASPA	443	rs36225687	TTTGCCTTACTACACTTCACGGATAC[-/T GTACT]TGTACTTTTTTTCCAAATTGAAGGT	108
ASPA	443	rs36051310	GAGGTGGCTGAAAGAAGCACGTATCC[-/C] TGATGGCATGGTTGCGGGTTATATG	109
ASPA	443	rs36034906	GAGAAAAGCAGTTCCTGGAACACCCC[-/C] ACCCCTTAACCCCTTATCTCTGCTT	110
ASPA	443	rs36033666	TTACATATGTATACATGTGCCATGTT[-/T] GGTGTGCCGCACCCATTAACTCGTC	111
ASPA	443	rs35730123	CTTTTTCCAGATTTTTTTTTTTTTTT[-/T] GAGACAGAGTTTCACTCTTGTTGCC	112
ASPA	443	rs35629100	TTTGGAAATCTTAAGCTTTTATTTGG[-/G] TGTCACAGAGAAACAGGATCTGTAT	113
ASPA	443	rs35614631	TACTTTAAGTTTTAGGGTACATGTGC[-/A] CCATGTGCAGGTTTGTTACATATGT	114

TABLE 2-continued

Non-limiting examples of polymorphisms

Gene name	GeneID	SNP ID	Mutation	SEQ ID NO:
ASPA	443	rs35225782	ATTCATGACCAGCCACATAAATGCAC[-/A] GTATTACTTCGCAAGCATGCCAATG	115
ASPA	443	rs35178659	GTGCACTAGAATTAGCTAAAGTGGGG[-/G] AAAAAAAGATGCATTTGATGGTCTA	116
ASPA	443	rs35095578	AACCTCCACCTCCCAGGTTCAAGAGA[-/A] TTCTCCTGCCTCAGCCTCCCAAGTA	117
ASPA	443	rs35002210	CCTCCCTGTGATCCGAAGTAGCAGAC[A/G]TA CTTAACTTCCATGGTGGATTGTT	118
ASPA	443	rs34744839	AAAACATTATTATATCTAGAAAAAAA[-/A] TGTATCTTAACCATTGTGGGAAGTG	119
ASPA	443	rs34680506	TTGAAGGTAAAATCATAGGGAGTTGG[-/G] AGCTGTCCTCTTGCGCTGAATCAGT	120
ASPA	443	rs34365618	ACTTGTGGCCTTTTTGGAGAGGTTAG[-/CA] ACTCTGAAAACTCTGTCCCTGGACC	121
ASPA	443	rs34275920	GAAGGAGAAAAAAGAGAGGAAATAAGT[-/T] AAAATAATAAACACAATTAATAAAG	122
ASPA	443	rs34109510	TGTATACATGTGCCATGTTGGTGTGC[C/T]GC ACCCATTAACTCGTCATTTAGCA	123
ASPA	443	rs34054576	TCACCTGTCACCTCCTATAGAACTTT[-/C] CCCTGACCCTCCTCTATAGCATTAA	124
ASPA	443	rs34015272	ATAAATGATCATCATTCACAGTAGGG[-/G] TTTTGTTTTGTTTTTTTTCTGGAAA	125
ASPA	443	rs34002091	ACAGACATATCTACAAACACACTTTT[-/T] CACATATTTGTGTAAGTCATTTATG	126
ASPA	443	rs28940574	AAAGACAACTAAACTAACGCTCAATG[A/C]A AAAAGTATTCGCTGCTGTTTACAT	127
ASPA	443	rs28940279	TACCGTGTACCCCGTGTTTGTGAATG[A/C]GG CCGCATATTACGAAAAGAAAGAA	128
ASPA	443	rs17850703	CAGGGCTGGAGGTAAAACCATTTATT[A/G]CT AACCCCAGAGCAGTGAAGAAGTG	129
ASPA	443	rs17222495	TTCTTCATTGCCTATTGAAGAGAGAG[C/T]GG AATGCTTTGGTTGCCAGATATGG	130
ASPA	443	rs17175228	CACAAGATCTCATTACTCAGGAGCTG[C/T]CC AAGTGTCTAATGTACTTAGTTAA	131
ASPA	443	rs16953074	TTCTGTGTAACATTTCATTTAAGCAA[A/G]GG ATTCGGCAAATCAAAAATTGTCA	132
ASPA	443	rs16953070	TAAAACGTATTGAAGGTATTATTGAC[G/T]CT GTTGAAGCAAAGAGAACAAAACA	133
HEXA	3073	rs62022858	ATCTGCTCTTCCAGTTGGATGACAAG[C/T]CT TGCTGTCTAACACCTGCTGCAGA	134
HEXA	3073	rs62022857	CCATTTTTTGTTGTATTTTTTTTTTC[C/T]TGAA TACTTTTTATCGCAGTTGGTT	135
HEXA	3073	rs62017872	CCCTGTCTCTAAAAGAAAAAAAAAAA[A/G]A AAAAAAAAAGAAAACAAAACCCAA	136
HEXA	3073	rs62017871	AGTGGCTCCAAAAAGGTCATGGAACC[C/T]CT TGAGGATGATGCAAATTGACTCT	137
HEXA	3073	rs61662730	TAAAGTTACTTTTCTTTTATTGACTT[C/T]CCC TTATTTTTTAACCTTATGCTTT	138
HEXA	3073	rs61329913	CAGAGTTAAAAAAAAAAAAAAAAAA[-/A] GGAAGTAGCAGCAACAGCTTGGAAA	139

TABLE 2-continued

Non-limiting examples of polymorphisms

Gene name	GeneID	SNP ID	Mutation	SEQ ID NO:
HEXA	3073	rs60920713	GTTGCCCAGGGTTGAGTGCAGAGGCA[C/T]AT TTGGCTCACAGCAACCTCTGCC	140
HEXA	3073	rs60783213	AAGGCTTTTTTTTTTTTTTTTTTTT[-/T TTT]GAGACAGAGTCTTGCTGTGTCACCC	141
HEXA	3073	rs60644867	GCCTACATTCTGCAAAGAGGAGGGAA[C/G]A TTCACAGCTCCATACTTGAACCCT	142
HEXA	3073	rs60288568	CCAAAGGAGAATAGCTCTAGGGGAGG[C/G]A GGTGGATGAGTATGCATGGGGGAG	143
HEXA	3073	rs59888548	GACTCCATCTCAAAAAAAAAAAAAAA[-/A] TGCAGTCTAATGGCAGAATTAGACT	144
HEXA	3073	rs59733856	TTATTTATTTATTTATTTATTTTTGA[A/G]ACA GGGTCTCTGTTGTCCAGGCTGG	145
HEXA	3073	rs59427837	TTTTGAGGCAGGGTCTCACTCTGTTG[C/T]CC AGGGTTGAGTGCAGAGGCACATC	146
HEXA	3073	rs59171976	CGCCTTGCGAAGGCCCCACAGCTTGC[C/T]TG TGACAAACGTTCATAGGCAAATG	147
HEXA	3073	rs58706602	GGAGGTCTGTACAAAGCACCACCTAC[C/T]TC ATGGGTCAGTTTCCACAGCAGAA	148
HEXA	3073	rs58696963	GAATCTTATAATTCACTGTGTACCTC[-/C CTC]TGTTTCATATTTTCGCAATTGAACT	149
HEXA	3073	rs58610850	AACATAGTATCTAATATAGCTTTACA[C/T]CC AAAGCCAAAATATGAATACACTG	150
HEXA	3073	rs58016062	TTGTTTTGTTTTGTTTGGGGGGGGGG[-/G] TTGTTTTTCTGAGAGGGAGTCTTGC	151
HEXA	3073	rs57733983	CATACCAAAGGGCAGCTGGAGGGATAC[C/T]A GACGGAAGTCATGTGGAGAGTGAA	152
HEXA	3073	rs57476645	CAGGTGTGAGCCACCACGACCACCAA[A/T]T TAGCTCTTTTTACTCCTTCCCTTC	153
HEXA	3073	rs56870003	AGTGGTAGCTGATTTTGCTTCTGGAT[A/C]CT TTGCCACCTTCCCACTCTTTAAT	154
HEXA	3073	rs56338339	AAAGACCTGTTTCTTAAAAAAAAAAA[-/A GAAAAAAAAAAA]GAAAGAAAAGAAAAG AAAAAAACAG	155
HEXA	3073	rs55995352	TAAAAAATCTTTCAATGAGGAGATGT[C/T]CC CAGAGCAAGACAGCTGTAGGATG	156
HEXA	3073	rs55860138	AAAAGAAAAAAAAAAAAAAAAAAAAA[-/A] GAAAACAAAACCCAAACCCATAAAG	157
HEXA	3073	rs55743646	CCTGTCTCTAAAAGAAAAAAAAAAAA[A/G]A AAAAAAAAGAAAACAAAACCCAAA	158
HEXA	3073	rs55665666	GTTATCATAGAAAAATATCACACTCT[-/GT] CTGTATCCCCACTTCCAGAAACTGT	159
HEXA	3073	rs36106892	CAGGAGCTCATAGAATTACATACAAT[-/C] TTTTTTTTTTTTTTTTGAGACAGCG	160
HEXA	3073	rs36091525	TTGAGAATCTTATAATTCACTGTGTA[-/C CTC]CCTCTGTTTCATATTTTCGCAATTG	161
HEXA	3073	rs35949555	CCACTACCACAGTGCCTAGAGAACAA[C/T]A TGTGTTTAATAATATTTAAATAAT	162
HEXA	3073	rs35827424	CCCTGTCTCTAAAAGAAAAAAAAAAA[-/A] AAAAAAAAAAGAAAACAAAACCCAA	163
HEXA	3073	rs35729578	CCATTATATCATTCATTTCCCACTCA[-/T] TTTCTTCATTCCAACCAAGATATAT	164

TABLE 2-continued

Non-limiting examples of polymorphisms

Gene name	GeneID	SNP ID	Mutation	SEQ ID NO:
HEXA	3073	rs35649102	TCCGTCTCAAAAAAAAAAAAAAAAAG[-/A] GAAAGGAATTATTCTCATGTATACA	165
HEXA	3073	rs35118677	CTGGGGCAGTTAAAAAGAAAAACAAA[-/C] CCCTGGTCCCTGCCCTTGAGGAGAT	166
HEXA	3073	rs35005352	CTCCAGGGTCCCATTCCAGGACCACA[-/C] GCCTGCTACCTCTGCAGCTCACTCA	167
HEXA	3073	rs34736306	GGATTGACATATACCAGTTAGACGGA[-/T] TTTTTTTTTCCATAAACCAGGCTCA	168
HEXA	3073	rs34607939	ACAAATAATTACTACATATCTACAAC[A/G]TT CCAGATACAGAAGAAATGGCCAA	169
HEXA	3073	rs34496117	TAAACACACTTGAAACATCATATAAA[-/A TG]ATATTACTACAAGACTTAACCGTAA	170
HEXA	3073	rs34300017	ACACAGGTAATCCATGTTTATTATAG[-/A] AAAATGCCACATTACTCTTTATTGA	171
HEXA	3073	rs34206496	AGTTATCATAGAAAAATATCACACTC[-/TG] TCTGTATCCCCACTTCCAGAAACTG	172
HEXA	3073	rs34110830	AATGAACTTACAGGAAGGTAATATAT[-/G] GGAAATAAACATCTTATTGAATTTA	173
HEXA	3073	rs34093438	GGACCCCTGAAAGGCACAAGACACCC[-/T] TTCAGGTTCACACTTCCTGAAAGCT	174
HEXA	3073	rs34085965	CCACCAATCACCAGAGCCTTCTGCTC[A/G]GG GGTACCTGAGGGAAAACAAGCAA	175
HEXA	3073	rs34004907	AAAGACTGAAAAAACATTCATAACTA[-/T] TTTTCTTGTTATCCTCGGAAATGTC	176
HEXA	3073	rs28942072	TATCTTCATCTTGGAGGAGATGAGGT[C/T]GA TTTCACCTGCTGGAAGTCCAACC	177
HEXA	3073	rs28942071	TTGCCTATGAACGTTTGTCACACTTC[C/T]GCT GTGAGTTGCTGAGGCGAGGTGT	178
HEXA	3073	rs28941771	GCTTGCTGTTGGATACATCTCGCCAT[C/T]AC CTGCCACTCTCTAGCATCCTGGA	179
HEXA	3073	rs28941770	CCGGGGCTTGCTGTTGGATACATCTC[G/T]CC ATTACCTGCCACTCTCTAGCATC	180

3. Nucleic Acid Target Length Evaluation:

In some embodiments, aspects of the invention relate to methods for detecting nucleic acid deletions or insertions in regions containing nucleic acid sequence repeats.

Genomic regions that contain nucleic acid sequence repeats are often the site of genetic instability due to the amplification or contraction of the number of sequence repeats (e.g., the insertion or deletion of one or more units of the repeated sequence). Instability in the length of genomic regions that contain high numbers of repeat sequences has been associated with a number of hereditary and non hereditary diseases and conditions.

For example, "Fragile X syndrome, or Martin-Bell syndrome, is a genetic syndrome which results in a spectrum of characteristic physical, intellectual, emotional and behavioral features which range from severe to mild in manifestation. The syndrome is associated with the expansion of a single trinucleotide gene sequence (CGG) on the X chromosome, and results in a failure to express the FMR-1 protein which is required for normal neural development. There are four generally accepted forms of Fragile X syndrome which relate to the length of the repeated CGG sequence; Normal (29-31 CGG repeats), Premutation (55-200 CGG repeats), Full Mutation (more than 200 CGG repeats), and Intermediate or Gray Zone Alleles (40-60 repeats)."

Other examples include cancer, which has been associated with microsatellite instability (MSI) involving an increase or decrease in the genomic copy number of nucleic acid repeats at one or more microsatellite loci (e.g., BAT-25 and/or BAT-26). The are currently many sequencing-based assays for determining the number of nucleic acid sequence repeats at a particular locus and identifying the presence of nucleic acid insertions or deletions. However, such techniques are not useful in a high throughput multiplex analysis where the entire length of a region may not be sequenced.

In contrast, in some embodiments, aspects of the invention relate to detecting the presence of an insertion or deletion at a genomic locus without requiring the locus to be sequenced (or without requiring the entire locus to be

sequenced). Aspects of the invention are particularly useful for detecting an insertion or deletion in a nucleic acid region that contains high levels of sequence repeats. The presence of sequence repeats at a genetic locus is often associated with relatively high levels of polymorphism in a population due to insertions or deletions of one or more of the sequence repeats at the locus. The polymorphisms can be associated with diseases or predisposition to diseases (e.g., certain polymorphic alleles are recessive alleles associated with a disease or condition). However, the presence of sequence repeats often complicates the analysis of a genetic locus and increases the risk of errors when using sequencing techniques to determine the precise sequence and number of repeats at that locus.

In some embodiments, aspects of the invention relate to determining the size of a genetic locus by evaluating the capture frequency of a portion of that locus suspected of containing an insertion or deletion (e.g., due to the presence of sequence repeats) using a nucleic acid capture technique (e.g., a nucleic acid sequence capture technique based on molecular inversion probe technology). According to aspects of the invention, a statistically significant difference in capture efficiency for a genetic locus of interest in different biological samples (e.g., from different subjects) is indicative of different relative lengths in those samples. It should be appreciated that the length differences may be at one or both alleles of the genetic locus. Accordingly, aspects of the invention may be used to identify polymorphisms regardless of whether biological samples being interrogated at heterozygous or homozygous for the polymorphisms. According to aspects of the invention, subjects that contain one or more loci with an insertion or deletion can be identified by analyzing capture efficiencies for nucleic acids obtained from one or more biological samples using appropriate controls (e.g., capture efficiencies for known nucleic acid sizes, capture efficiencies for other regions that are not suspected of containing an insertion or deletion in the biological sample(s), or predetermined reference capture efficiencies, or any combination thereof. However, it should be appreciated that aspects of the invention are not limited by the nature or presence of the control. In some embodiments, if a statistically significant variation in capture efficiency is detected, a subject may be identified as being at risk for a disease or condition associated with insertions or deletions at that genetic locus. In some embodiments, the subject may be analyzed in greater detail in order to determine the precise nature of the insertion or deletion and whether the subject is heterozygous or homozygous for one or more insertions or deletions. For example, gel electrophoresis of an amplification (e.g., PCR) product of the locus, or Southern blotting, or any combination thereof can be used as an orthogonal approach to verify the length of the locus. In some embodiments, a more exhaustive and detailed sequence analysis of the locus can be performed to identify the number and types of insertions and deletions. However, other techniques may be used to further analyze a locus identified as having an abnormal length according to aspects of the invention.

Accordingly, aspects of the invention relate to detecting abnormal nucleic acid lengths in genomic regions of interest. In some embodiments, the invention aims to estimate the size of genomic regions that are hard to be accessed, such as repetitive elements. However, it should be appreciated that methods of the invention do not require that the precise length be estimated. In some embodiments, it is sufficient to determine that one or more alleles with abnormal lengths are present at a locus of interest (e.g., based on the detection of abnormal capture efficiencies).

In a non-limiting example, fragile X can be used to illustrate aspects of the invention where the size of trinucleotide repeats (genotype) is linked to a symptom (phenotype). However, it should be appreciated that fragile X is a non-limiting example and similar analyses may be performed for other genetic loci (e.g., independently or simultaneously in multiplex analyses).

Use of molecular inversion probes (MIPs) has been demonstrated for detection of single nucleotide polymorphisms (Hardenbol et al. 2005 Genome Res 15:269-75) and for preparative amplification of large sets of exons (Porreca et al. 2007 Nat Methods 4:931-6, Krishnakumar et al. 2008 Proc Natl Acad Sci USA 105:9296-301). In both cases, oligonucleotide probes are designed which have ends ('targeting arms') that hybridize up-stream and down-stream of the locus that is to be amplified.

In some embodiments, aspects of the invention are based on the recognition that the effect of length on probe capturing efficiency can be used in the context of an assay (e.g., a high throughput and/or multiplex assay) to allow the length of sequences to be determined without requiring sequencing of the entire region being evaluated. This is particularly useful for repeat regions that are prone to changes in size. As illustrated in FIG. **8**, which is reproduced from Deng et al., Nature Biotech. 27:353-60, (see Supplemental FIG. **1**G of Deng et al.,) illustrates that shorter sequences are captured with higher efficiency that longer sequences using MIPs. The statistical package R and its effects module were used for this analysis. A linear model was used, and each individual factor was assumed to be independent. The dashed lines represent a 95% confidence interval. Shorter target sequences were captured with higher efficiency than long target sequences ($p<2\times10^{-16}$). However, the use of this differential capture efficiency for systematic sequence length analysis was not previously recognized.

In some embodiments, following probe hybridization, polymerase fill-in and ligation reactions are performed to convert the hybridized probe to a covalently-closed, circular molecule containing the desired target. PCR or rolling circle amplification plus exonuclease digestion of non-circularized material is performed to isolate and amplify the circular targets from the starting nucleic acid pool. Since one of the main benefits of the method is the potential for a high degree of multiplexing, generally thousands of targets are captured in a single reaction containing thousands of probes.

According to aspects of the invention, repetitive regions are surrounded by non repetitive unique sequences, which can be used to amplify the repeat-containing regions using, for example, PCR or padlock (MIP)-based method.

In addition to the repetitive regions, a probe (e.g., a MIP or padlock probe) can be designed to include at least a sequence that is sufficient to be uniquely identified in the genome (or target pool). After the probe is circularized and amplified, the amplicon can be end-sequenced so that the unique sequence can be identified and served as the "representative" of the repetitive region as illustrated in FIG. **9**. FIG. **9** illustrates a non-limiting scheme of padlock (MIP) capture of a region that includes both repetitive regions (thick wavy line) and the adjacent unique sequence (thick strait line). The regions of the probe are indicated with the targeting arms shown as regions "1" and "3." An intervening region that may be, or include, a sequencing primer binding site is shown as "2." After the padlock is circularized and

amplified, it can be end-sequenced to obtain the sequence of the unique sequence, which represents the repetitive region of interest.

Although capturing efficiency is overall negatively correlated with target length, different probe sequences may have unique features. Therefore, multiple probes could be designed and tested so that an optimal one is chosen to be sensitive enough to differentiate repetitive sizes of roughly 0-150 bp, 150-600 bp, and beyond, which represent normal, premutation and full mutation of fragile X syndrome, respectively. However, it should be appreciated that other probe sizes and sequences can be designed, and optionally optimized, to distinguish a range of repeat region size differences (e.g., length differences of about 3-30 bases, about 30-60 bases, about 60-90 bases, about 90-120 bases, about 120-150 bases, about 150-300 bases, about 300-600 bases, about 600-900 bases, or any intermediate or longer length difference). It should be appreciated that a length difference may be an increase in size or a decrease in size.

In some embodiments, an initial determination of an unexpected capture frequency is indicative of the presence of size difference. In some embodiments, an increase in capture frequency is indicative of a deletion. In some embodiments, a decrease in capture frequency is indicative of an insertion. However, it should be appreciated that depending on specific sequence parameters and the relative sizes of the capture probes, the target region, and the deletions or insertions, a change in capture frequency can be associated with either an increase or decrease in target region length. In some embodiments, the precise nature of the change can be determined using one or more additional techniques as described herein.

Accordingly, in some aspects a MIP probe includes a linear nucleic acid strand that contains two hybridization sequences or targeting arms, one at each end of the linear probe, wherein each of the hybridization sequences is complementary to a separate sequence on a the same strand of a target nucleic acid, and wherein these sequences on the target nucleic acid flank the two ends of the target nucleic acid sequence of interest. It should be appreciated that upon hybridization, the two ends of the probe are inverted with respect to each other in the sense that both 5' and 3' ends of the probe hybridize to the same strand to separate regions flanking the target region (as illustrated in FIG. **9** for example).

In some embodiments, the hybridization sequences are between about 10-100 nucleotides long, for example between about 10-30, about 30-60, about 60-90, or about 20, about 30, about 40, or about 50 nucleotides long. However, other lengths may be used depending on the application. In some embodiments, the hybridization Tms of both targeting arms of a probe are designed or selected to be similar. In some embodiments, the hybridization Tms of the targeting arms of a plurality of probes designed to capture different target regions are selected or designed to be similar so that they can be used together in a multiplex reaction. Accordingly, a typical size of a MIP probe prior to fill in is about 60-80 nucleotides long. However, other sizes can be used depending on the sizes of the targeting arms and any other sequences (e.g., primer binding or tag sequences) that are present in the MIP probe. In some embodiments, MIP probes are designed to avoid sequence-dependent secondary structures. In some embodiments, MIP probes are designed such that the targeting arms do not overlap with known polymorphic regions. In some embodiments, targeting arms that can be used for capturing the repeat region of the Fragile X locus can have the following sequences or complementary to these sequences depending on the strand that is captured.

left: CTCCUITYCGGTITCACTTC (SEQ ID NO: 181)

right: ATCTTCTCTTCAGCCCTGCT (SEQ ID NO: 182)

The typical captured size using these targeting arms is about 100 nucleotides in length (e.g., about 30 repeats of a tri-nucleotide repeat).

In some embodiments, the number of reads obtained for the "representative" of the repetitive region is not informative to estimate the target length because it is dependent on the total number of reads obtained. To overcome this, it is useful to include one or more probes that target other "control" regions where no or minimal polymorphism exists among populations. Because of the systematic consistency of capturing efficiency (see, e.g., FIG. **9**), the ratio of reads obtained for the repetitive "representative" to reads obtained for the control region(s) will be tuned using DNA with defined numbers of repeats. Ultimately, the ratio can serve as a measure of the repeat length as illustrated in FIG. **10**. FIG. **10** illustrates a non-limiting hypothetical relationship between target gap size and the relative number of reads of the repetitive region, which is measured by the ratio of the repeat "representative" reads vs. the "control" region reads. The unit of y-axis is arbitrary.

In some embodiments, to better tell targets with similar size range apart, the whole repetitive region can be sequenced by making a shotgun library (e.g., by making a shotgun library from a captured sequence, for example a sequence captured using a MIP probe). The longer the repeat is, the more short reads of repeats will be obtained. Therefore, the target length will contribute twice to the relative number of "repetitive" reads, which will gain better resolution of differentiating targets. In some embodiments, the expectation is that the number of reads from any given repeat will be a direct function of the number of repeats present. However, in some embodiments, a Poisson sampling induced spread may need to be considered and in some embodiments may be sufficiently large to limit the resolution.

When a precise measurement of the length of both alleles from a diploid sample is desired, further manipulations may be required. This is because the capture efficiency measured will actually be the average efficiency of the two alleles. To effectively achieve separate measurements for each allele, barcodes (e.g., sequence tags) can be used that allow the efficiency of individual capture events (from individual genomic loci) to be followed. FIG. **11**A-C shows the approach. For a given locus, MIPs are synthesized to contain one of a large number differentiator tags in their backbone such that the probability of any two MIPs in a reaction having the same differentiator tag sequence is low. MIP capture is performed on the sample; the reaction will be biased for shorter target lengths, and therefore the reaction product will be comprised of more 'short' circles than 'long' circles. Each circle should bear a unique differentiator tag sequence. Then, linear RCA (1RCA) is performed on the circles. In the 1RCA reaction, circles are converted into long, linear concatemers of themselves. The 1RCA reaction for a given circle stops when the concatemer has reached a 'fixed' length (based on the processivity/error rate of the polymerase). Concatemers derived from smaller circles will therefore contain more copies of the differentiator tag, and concatemers derived from larger circles will contain fewer copies of the differentiator tag. The number of each differentiator tag sequence is counted, for example, by next-generation sequencing.

When number of occurrences is plotted against differentiator tag ID, the data will naturally cluster into two groups reflecting the lengths of the two alleles in the diploid sample. The allele lengths can therefore be read directly off this graph, after absolute length calibration using known standards. In some embodiments, a sequencing technique (e.g., a next-generation sequencing technique) is used to sequence part of one or more captured targets (e.g., or amplicons thereof) and the sequences are used to count the number of different barcodes that are present. Accordingly, in some embodiments, aspects of the invention relate to a highly-multiplexed qPCR reaction.

Other non-limiting examples of loci at which insertions or deletions or repeat sequences may be associated with a disease or condition are provided in Tables 3 and 4. It should be appreciated that the presence of an abnormal length at any one or more of these loci may be evaluated according to aspects of the invention. In some embodiments, two or more of these loci or other loci may be evaluated in a single multiplex reaction using different probes designed to hybridize under the same reaction conditions to different target nucleic acid in a biological sample.

TABLE 3

Polyglutamine (PolyQ) Diseases

Type	Gene	Normal/ wildtype	Pathogenic
DRPLA (Dentatorubropallidoluysian atrophy)	ATN1 or DRPLA	6-35	49-88
HD (Huntington's disease)	HTT (Huntingtin)	10-35	35+
SBMA (Spinobulbar muscular atrophy or Kennedy disease)	Androgen receptor on the X chromosome.	9-36	38-62
SCA1 (Spinocerebellar ataxia Type 1)	ATXN1	6-35	49-88
SCA2 (Spinocerebellar ataxia Type 2)	ATXN2	14-32	33-77
SCA3 (Spinocerebellar ataxia Type 3 or Machado-Joseph disease)	ATXN3	12-40	55-86
SCA6 (Spinocerebellar ataxia Type 6)	CACNA1A	4-18	21-30
SCA7 (Spinocerebellar ataxia Type 7)	ATXN7	7-17	38-120
SCA17 (Spinocerebellar ataxia Type 17)	TBP	25-42	47-63

TABLE 4

Non-Polyglutamine Diseases

Type	Gene	Codon	Normal/ wildtype	Pathogenic
FRAXA (Fragile X syndrome)	FMR1, on the X-chromosome	CGG	6-53	230+
FXTAS (Fragile X-associated tremor/ataxia syndrome)	FMR1, on the X-chromosome	CGG	6-53	55-200
FRAXE (Fragile XE mental retardation)	AFF2 or FMR2, on the X-chromosome	GCC	6-35	200+
FRDA (Friedreich's ataxia)	FXN or X25, (frataxin)	GAA	7-34	100+

TABLE 4-continued

Non-Polyglutamine Diseases

Type	Gene	Codon	Normal/ wildtype	Pathogenic
DM (Myotonic dystrophy)	DMPK	CTG	5-37	50+
SCA8 (Spinocerebellar ataxia Type 8)	OSCA or SCA8	CTG	16-37	110-250
SCA12 (Spinocerebellar ataxia Type 12)	PPP2R2B or SCA12	CAG On 5' end	7-28	66-78

The following examples illustrate aspects and embodiments of the invention and are not intended to be limiting or restrictive. Many variations of the invention will become apparent to those skilled in the art upon review of this specification. The full scope of the invention should be determined by reference to the claims, along with their full scope of equivalents, and the specification, along with such variations.

4. Increasing Detection Sensitivity:

In some embodiments, aspects of the invention relate to methods for increasing the sensitivity of nucleic acid detection assays.

There are currently many genomic assays that utilize next-generation (e.g., polony-based) sequencing to generate data, including genome resequencing, RNA-seq for gene expression, bisulphite sequencing for methylation, and Immune-seq, among others. In order to make quantitative measurements (including genotype calling), these methods utilize the counts of sequencing reads of a given genomic locus as a proxy for the representation of that sequence in the original sample of nucleic acids. The majority of these techniques require a preparative step to construct a high-complexity library of DNA molecules that is representative of a sample of interest. Current assays use one of several alternative nucleic acid preparative techniques (e.g., amplification, for example PCR-based amplification; sequence-specific capture, for example, using immobilized capture probes; or target capture into a circularized probe followed by a sequence analysis step. In order to reduce errors associated with the unpredictability (stochastic nature) of nucleic acid isolation and sequence analysis techniques, current methods involve oversampling a target nucleic acid preparation in order to increase the likelihood that all sequences that are present in the original nucleic acid sample will be represented in the final sequence data. For example, a genomic sequencing library may contain an over- or under-representation of particular sequences from a source nucleic acid sample (e.g., genome preparation) as a result of stochastic variations in the library construction process. Such variations can be particularly problematic when they result in target sequences from a genome being absent or undetectable in a sequencing library. For example, an under-representation of particular allelic sequences (e.g., heterozygotic alleles) from a genome in a sequencing library can result in an apparent homozygous representation in a sequencing library.

In contrast, aspects of the invention relate to basing a nucleic acid sequence analysis on results from two or more different nucleic acid preparatory techniques that have different systematic biases in the types of nucleic acids that they sample rather than simply oversampling the target nucleic acid. According to some embodiments, different techniques have different sequence biases that are systematic and not simply due to stochastic effects during nucleic acid capture or amplification. Accordingly, in some embodiments, the degree of oversampling required to overcome variations in nucleic acid preparation needs to be sufficient to overcome the biases. In some embodiments, the invention provides methods that reduce the need for oversampling by combining nucleic acid and/or sequence results obtained from two or more different nucleic acid preparative techniques that have different biases.

According to the invention, different techniques have different characteristic or systematic biases. For example, one technique may bias a sample analysis towards one particular allele at a genetic locus of interest, whereas a different technique would bias the sample analysis towards a different allele at the same locus. Accordingly, the same sample may be identified as being different depending on the type of technique that is used to prepare nucleic acid for sequence analysis. This effectively represents a sensitivity issue, because each technique has a different relative sensitivities for polymorphic sequences of interest.

According to aspects of the invention, the sensitivity of a nucleic acid analysis can be increased by combining the sequences from different nucleic acid preparative steps and using the combined sequence information for a diagnostic assay (e.g., for a making a call as to whether a subject is homozygous or heterozygous at a genetic locus of interest).

Currently, the ability of DNA sequencing to detect mutations is limited by the ability of the upstream sample isolation (e.g., by amplification, immobilization enrichment, circularization capture, etc.) methods to reliably isolate the locus of interest.

If one wishes to make heterozygote base-calls for a diploid genome (e.g. a human sample presented for molecular diagnostic sequencing), it is important in some embodiments that the isolation method produces near- or perfectly-uniform amounts of the two alleles to be sequenced (at least sufficiently uniform to be "called" unambiguously as a heterozygote or a homozygote for a locus of interest).

Sample preparative methods may fall into three classes: 1) single- or several target amplification (e.g., uniplex PCR, 'multiplex' PCR), 2) multi-target hybridization enrichment (e.g., Agilent SureSelect 'hybrid capture' [Gnirke et al 2009, *Nature methods* 27:182-9], Roche/Nimblegen 'sequence capture' [Hodges et al 2007, *Nature genetics* 39:1522-7], and 3) multi-target circularization selection (e.g. molecular inversion probes or padlock probes, [Porreca et al 2007, *Nature methods* 4:931-6, Turner et al 2009, *Nature methods*

6:315-6], 'selectors' [Dahl et al 2005, *Nucleic acids research* 33:e71]). Each of these methods can result in a pool of isolated product that does not adequately represent the input abundance distribution. For example, the two alleles at a heterozygous position can become skewed far from their input 50:50 ratio to something that results in a missed basecall during downstream sequencing. For example, if the ratio was skewed from 50:50 to 10:90, and the sample was sequenced to 10×average coverage, there is a high probability that one of the two alleles would not be observed once in the ten sequencing reads. This would reduce the sensitivity of the sequencing method by converting a heterozygous position to homozygous (where potentially the 'mutant' allele was the one not observed). In some embodiments, a skewed ratio is a particular issue that decreases the sensitivity of detecting mutations present in a heterogeneous tumor tissue. For example, if only 10% of the cells analyzed in a heterogeneous sample harbored a heterozygous mutation, the mutation would be expected to be present in 5% of sequence reads, not 50%. In this scenario, the need for robust, sensitive detection may be even more acute.

The methods disclosed herein are based, in part, on the discovery that certain classes of isolation methods have different modes of bias. The disclosure provide methods for increasing the sensitivity of the downstream sequencing by using a combination of multiple isolation methods (e.g., one or more from at least two of the classes disclosed herein) for a sample. This is particularly important in molecular diagnostics where high sensitivity is required to minimize the chances of 'missing' a disease-associated mutation. For example, given a nominal false-negative error rate of 1×10^{-3} for sequencing following circularization selection, and a false-negative error rate of 1×10^{-3} for sequencing following hybridization enrichment, one can achieve a final false-negative rate of 1×10^{-6} by performing both techniques on the sample (assuming failures in each method are fully independent). For a recessive disease with carrier frequency of 0.1, caused by a single fully-penetrant mutant allele, the number of missed carrier diagnoses would decrease from 1000 per million patients tested to 1 per million patients tested. Furthermore, if the testing was used in the context of prenatal carrier screening, the number of affected children born as a result of missing the carrier call in one parent would decrease from 25 per million to 25 per billion born.

Additionally, the disclosure provides combinations of preparative methods to effectively increase sequencing coverage in regions containing disease-associated alleles. Since heterozygote error rate is largely tied to both deviations from 50:50 allele representation, and in the case of next-generation DNA sequencing deviations from average abundance (such that less abundant isolated targets are more likely to be undersampled at one or both alleles), selectively increasing coverage in these regions will also selectively increase sensitivity. Furthermore, MIPs that detect presence or absence of specific known disease-associated mutations can be used to increase sensitivity selectively. In some embodiments, these MIPs would have a targeting arm whose 3'-most region is complementary to the expected mutation, and has a fill-in length of 0 or more bp. Thus, the MIP will form only if the mutation is present, and its presence will be detected by sequencing.

Additionally, algorithms disclosed herein may be used to determine base identity with varying levels of stringency depending on whether the given position has any known disease-associated alleles. Stringency can be reduced in such positions by decreasing the minimum number of observed mutant reads necessary to make a consensus base-call. This will effectively increase sensitivity for mutant allele detection at the cost of decreased specificity.

An embodiment of the invention combines MIPs plus hybridization enrichment, plus optionally extra MIPs targeted to specific known, common disease-associated loci, e.g., to detect the presence of a polymorphism in a target nucleic acid. A non-limiting example is illustrated in FIG. **12** that illustrates a schematic using MIPs plus hybridization enrichment, plus optionally extra MIPs targeted to specific known, common disease-associated loci, e.g., to detect the presence of a polymorphism in a target nucleic acid.

FIGS. **13** and **14** illustrate different capture efficiencies for MIP-based captures. FIG. **13** shows a graph of per-target abundance with MIP capture. In this graph, bias largely drives the heterozygote error rate, since targets which are less abundant here are less likely to be covered in sufficient depth during sequencing to adequately sample both alleles. This is from Turner et al 2009, Nature methods 6:315-6. Hybridization enrichment results in a qualitatively similar abundance distribution, but the abundance of a given target is likely not correlated between the two methods. FIG. **14** shows a graph of correlation between two MIP capture reactions from Ball et al 2009, Nature biotechnology 27:361-8. Each point represents the target abundance in replicate 1 and replicate 2. Pearson correlation r=0.956. This indicates that MIP capture reproducibly biases targets to specific abundances. Hybridization enrichment is similarly correlated from one capture to the next.

According to aspects of the invention, such biases can be detected or overcome by systematically combining different capture and/or analytical techniques in an assay that interrogates a plurality of loci in a plurality of subject samples.

Accordingly, it should be appreciated that in any of the embodiments described herein (e.g., tiling/staggering, tagging, size-detection, sensitivity enhancing algorithms, or any combination thereof), aspects of the invention involve preparing genomic nucleic acid and/or contacting them with one or more different probes (e.g., capture probes, hybridization probes, MIPs, others etc.). In some embodiments, the amount of genomic nucleic acid used per subject ranges from 1 ng to 10 micrograms (e.g., 500 ng to 5 micrograms). However, higher or lower amounts (e.g., less than 1 ng, more than 10 micrograms, 10-50 micrograms, 50-100 micrograms or more) may be used. In some embodiments, for each locus of interest, the amount of probe used per assay may be optimized for a particular application. In some embodiments, the ratio (molar ratio, for example measured as a concentration ratio) of probe to genome equivalent (e.g., haploid or diploid genome equivalent, for example for each allele or for both alleles of a nucleic acid target or locus of interest) ranges from 1/100, 1/10, 1/1, 10/1, 100/1, 1000/1. However, lower, higher, or intermediate ratios may be used.

In some embodiments, the amount of target nucleic acid and probe used for each reaction is normalized to avoid any observed differences being caused by differences in concentrations or ratios. In some embodiments, in order to normalize genomic DNA and probe, the genomic DNA concentration is read using a standard spectrophotometer or by fluorescence (e.g., using a fluorescent intercalating dye). The probe concentration may be determined experimentally or using information specified by the probe manufacturer.

Similarly, once a locus has been captured (e.g., on a MIP or other probe or in another form), it may be amplified and/or sequenced in a reaction involving one or more primers. The amount of primer added for each reaction can range from 0.1 pmol to 1 nmol, 0.15 pmol to 1.5 nmol (for

example around 1.5 pmol). However, other amounts (e.g., lower, higher, or intermediate amounts) may be used.

In some embodiments, it should be appreciated that one or more intervening sequences (e.g., sequence between the first and second targeting arms on a MIP capture probe), identifier or tag sequences, or other probe sequences that are not designed to hybridize to a target sequence (e.g., a genomic target sequence) should be designed to avoid excessive complementarity (to avoid cross-hybridization) to target sequences or other sequences (e.g., other genomic sequences) that may be in a biological sample. For example, these sequences may be designed have a sufficient number of mismatches with any genomic sequence (e.g., at least 5, 10, 15, or more mismatches out of 30 bases) or as having a Tm (e.g., a mismatch Tm) that is lower (e.g., at least 5, 10, 15, 20, or more degrees C. lower) than the hybridization reaction temperature.

It should be appreciated that a targeting arm as used herein may be designed to hybridize (e.g., be complementary) to either strand of a genetic locus of interest if the nucleic acid being analyzed is DNA (e.g., genomic DNA). However, in the context of MIP probes, whichever strand is selected for one targeting arm will be used for the other one. However, in the context of RNA analysis, it should be appreciated that a targeting arm should be designed to hybridize to the transcribed RNA. It also should be appreciated that MIP probes referred to herein as "capturing" a target sequence are actually capturing it by template-based synthesis rather than by capturing the actual target molecule (other than for example in the initial stage when the arms hybridize to it or in the sense that the target molecule can remain bound to the extended MIP product until it is denatured or otherwise removed).

It should be appreciated that in some embodiments a targeting arm may include a sequence that is complementary to one allele or mutation (e.g., a SNP or other polymorphism, a mutation, etc.) so that the probe will preferentially hybridize (and capture) target nucleic acids having that allele or mutation. However, in many embodiments, each targeting arm is designed to hybridize (e.g., be complementary) to a sequence that is not polymorphic in the subjects of a population that is being evaluated. This allows target sequences to be captured and/or sequenced for all alleles and then the differences between subjects (e.g., calls of heterozygous or homozygous for one or more loci) can be based on the sequence information and/or the frequency as described herein.

It should be appreciated that sequence tags (also referred to as barcodes) may be designed to be unique in that they do not appear at other positions within a probe or a family of probes and they also do not appear within the sequences being targeted. Thus they can be used to uniquely identify (e.g., by sequencing or hybridization properties) particular probes having other characteristics (e.g., for particular subjects and/or for particular loci).

It also should be appreciated that in some embodiments probes or regions of probes or other nucleic acids are described herein as comprising or including certain sequences or sequence characteristics (e.g., length, other properties, etc.). However, it should be appreciated that in some embodiments, any of the probes or regions of probes or other nucleic acids consist of those regions (e.g., arms, central regions, tags, primer sites, etc., or any combination thereof) of consist of those sequences or have sequences with characteristics that consist of one or more characteristics (e.g., length, or other properties, etc.) as described herein in the context of any of the embodiments (e.g., for tiled or staggered probes, tagged probes, length detection, sensitivity enhancing algorithms or any combination thereof).

It should be appreciated that probes, primers, and other nucleic acids designed or used herein may be synthetic, natural, or a combination thereof. Accordingly, as used herein, the term "nucleic acid" refers to multiple linked nucleotides (i.e., molecules comprising a sugar (e.g., ribose or deoxyribose) linked to an exchangeable organic base, which is either a pyrimidine (e.g., cytosine (C), thymidine (T) or uracil (U)) or a purine (e.g., adenine (A) or guanine (G)). "Nucleic acid" and "nucleic acid molecule" may be used interchangeably and refer to oligoribonucleotides as well as oligodeoxyribonucleotides. The terms shall also include polynucleosides (i.e., a polynucleotide minus a phosphate) and any other organic base containing nucleic acid.

The organic bases include adenine, uracil, guanine, thymine, cytosine and inosine. Unless otherwise stated, nucleic acids may be single or double stranded. The nucleic acid may be naturally or non-naturally occurring. Nucleic acids can be obtained from natural sources, or can be synthesized using a nucleic acid synthesizer (i.e., synthetic).

Harvest and isolation of nucleic acids are routinely performed in the art and suitable methods can be found in standard molecular biology textbooks. (See, for example, Maniatis' Handbook of Molecular Biology.) The nucleic acid may be DNA or RNA, such as genomic DNA, mitochondrial DNA, mRNA, cDNA, rRNA, miRNA, or a combination thereof. Non-naturally occurring nucleic acids such as bacterial artificial chromosomes (BACs) and yeast artificial chromosomes (YACs) can also be used.

The invention also contemplates the use of nucleic acid derivatives. As will be described herein, the use of certain nucleic acid derivatives may increase the stability of the nucleic acids of the invention by preventing their digestion, particularly when they are exposed to biological samples that may contain nucleases. As used herein, a nucleic acid derivative is a non-naturally occurring nucleic acid or a unit thereof. Nucleic acid derivatives may contain non-naturally occurring elements such as non-naturally occurring nucleotides and non-naturally occurring backbone linkages.

Nucleic acid derivatives may contain backbone modifications such as but not limited to phosphorothioate linkages, phosphodiester modified nucleic acids, phosphorothiolate modifications, combinations of phosphodiester and phosphorothioate nucleic acid, methylphosphonate, alkylphosphonates, phosphate esters, alkylphosphonothioates, phosphoramidates, carbamates, carbonates, phosphate triesters, acetamidates, carboxymethyl esters, methylphosphorothioate, phosphorodithioate, p-ethoxy, and combinations thereof. The backbone composition of the nucleic acids may be homogeneous or heterogeneous.

Nucleic acid derivatives may contain substitutions or modifications in the sugars and/or bases. For example, they include nucleic acids having backbone sugars which are covalently attached to low molecular weight organic groups other than a hydroxyl group at the 3' position and other than a phosphate group at the 5' position (e.g., an 2'-0-alkylated ribose group). Nucleic acid derivatives may include non-ribose sugars such as arabinose. Nucleic acid derivatives may contain substituted purines and pyrimidines such as C-5 propyne modified bases, 5-methylcytosine, 2-aminopurine, 2-amino-6-chloropurine, 2,6-diaminopurine, hypoxanthine, 2-thiouracil and pseudoisocytosine. In some embodiments, substitution(s) may include one or more substitutions/modifications in the sugars/bases, groups attached to the base,

including biotin, fluorescent groups (fluorescein, cyanine, rhodamine, etc), chemically-reactive groups including carboxyl, NHS, thiol, etc., or any combination thereof.

A nucleic acid may be a peptide nucleic acid (PNA), locked nucleic acid (LNA), DNA, RNA, or co-nucleic acids of the same such as DNA-LNA co-nucleic acids. PNA are DNA analogs having their phosphate backbone replaced with 2-aminoethyl glycine residues linked to nucleotide bases through glycine amino nitrogen and methylenecarbonyl linkers. PNA can bind to both DNA and RNA targets by Watson-Crick base pairing, and in so doing form stronger hybrids than would be possible with DNA or RNA based oligonucleotides in some cases.

PNA are synthesized from monomers connected by a peptide bond (Nielsen, P. E. et al. *Peptide Nucleic Acids, Protocols and Applications*, Norfolk: Horizon Scientific Press, p. 1-19 (1999)). They can be built with standard solid phase peptide synthesis technology. PNA chemistry and synthesis allows for inclusion of amino acids and polypeptide sequences in the PNA design. For example, lysine residues can be used to introduce positive charges in the PNA backbone. All chemical approaches available for the modifications of amino acid side chains are directly applicable to PNA. Several types of PNA designs exist, and these include single strand PNA (ssPNA), bisPNA and pseudocomplementary PNA (pcPNA).

The structure of PNA/DNA complex depends on the particular PNA and its sequence. ssPNA binds to single stranded DNA (ssDNA) preferably in antiparallel orientation (i.e., with the N-terminus of the ssPNA aligned with the 3' terminus of the ssDNA) and with a Watson-Crick pairing. PNA also can bind to DNA with a Hoogsteen base pairing, and thereby forms triplexes with double stranded DNA (dsDNA) (Wittung, P. et al., *Biochemistry* 36:7973 (1997)).

A locked nucleic acid (LNA) is a modified RNA nucleotide. An LNA form hybrids with DNA, which are at least as stable as PNA/DNA hybrids (Braasch, D. A. et al., *Chem & Biol.* 8(1):1-7(2001)). Therefore, LNA can be used just as PNA molecules would be. LNA binding efficiency can be increased in some embodiments by adding positive charges to it. LNAs have been reported to have increased binding affinity inherently.

Commercial nucleic acid synthesizers and standard phosphoramidite chemistry are used to make LNAs. Therefore, production of mixed LNA/DNA sequences is as simple as that of mixed PNA/peptide sequences. The stabilization effect of LNA monomers is not an additive effect. The monomer influences conformation of sugar rings of neighboring deoxynucleotides shifting them to more stable configurations (Nielsen, P. E. et al. *Peptide Nucleic Acids, Protocols and Applications*, Norfolk: Horizon Scientific Press, p. 1-19 (1999)). Also, lesser number of LNA residues in the sequence dramatically improves accuracy of the synthesis. Most of biochemical approaches for nucleic acid conjugations are applicable to LNA/DNA constructs.

These and other aspects of the invention are illustrated by the following non-limiting examples.

EXAMPLES

The following examples illustrate non-limiting embodiments of the invention.

Example 1: Design a Set of Capture Probes for a Human Target Exon

All targets are captured as a set of partially-overlapping subtargets. For example, in the tiling approach, a 200 bp target exon might be captured as a set of 12 subtargets, each 60 bp in length (FIG. **1**). Each subtarget is chosen such that it partially overlaps two or three other targets.

In some embodiments, all probes are composed of three regions: 1) a 20 bp 'targeting arm' comprised of sequence which hybridizes immediately upstream from the sub-target, 2) a 30 bp 'constant region' comprised of sequence used as a pair of amplification priming sites, and 3) a second 20 bp 'targeting arm' comprised of sequence which hybridizes immediately downstream from the sub-target. Targeting arm sequences will be different for each capture probe in a set, while constant region sequence will be the same for all probes in the set, allowing all captured targets to be amplified with a single set of primers. Targeting arm sequences should be designed such that any given pair of 20 bp sequences is unique in the target genome (to prevent spurious capture of undesired sites). Additionally, melting temperatures should be matched for all probes in the set such that hybridization efficiency is uniform for all probes at a constant temperature (e.g., 60 C). Targeting arm sequences should be computationally screened to ensure they do not form strong secondary structure that would impair their ability to basepair with the genomic target.

Hybridize Capture Probes to Human Genomic Sample

Assemble Hybridization Reaction:

1.0 ul capture probe mix (~2.5 μmol)

2.0 ul 10× Ampligase buffer (Epicentre)

6.0 ul 500 ng/ul human genomic DNA (~16.7 fmol)

11 ul dH20

In a thermal cycler, heat reaction to 95 C for 5 min to denature genomic DNA, then cool to 60 C. Allow to incubate at 60 C for 40 hours.

Convert Hybridized Probes into Covalently-Closed Circular Products Containing Subtargets

Prepare Fill-In/Ligation Reaction Mixture:

0.25 ul 2 mM dNTP mix (Invitrogen)

2.5 ul 10×Ampligase buffer (Epicentre)

5.0 ul 5 U/ul Taq Stoffel fragment (Applied Biosystems)

12.5 ul 5 U/ul Ampligase (Epicentre)

4.75 ul dH20

Add 1.0 ul of this mix to the hybridized probe reaction, and incubate at 60 C for 10 hours.

Purify Circularized Probe/Subtarget Products from Un-Reacted Probes and Genomic DNA

Prepare Exonuclease Reaction Mixture:

21 ul fill-in/ligation reaction product

2.0 ul 10×exonuclease I buffer (New England Biolabs)

2.0 ul 20 U/ul exonuclease I (New England Biolabs)

2.0 ul 100 U/ul exonuclease III (New England Biolabs)

Incubate at 37 C for 60 min, then heat-inactivate by incubating at 80 C for 15 min. Immediately cool to 4 C for storage.

Amplify Circular Material by PCR Using Primers Specific to the 'Constant Region' of the Probes

Prepare PCR Mixture:

5.0 ul 10×Accuprime reaction buffer (Invitrogen)

1.5 ul 10 uM CP-2-FA (5'-GCACGATCCGACGGTAGTGT-3') (SEQ ID NO:183)

1.5 ul 10 uM CP-2-RA (5'-CCGTAATCGGGAAGCTGAAG-3') (SEQ ID NO:184)

0.4 ul 25 mM dNTP mix (Invitrogen)

2.0 ul heat-inactivated exonuclease reaction mix

1.5 ul 10×SybrGreen (Invitrogen)

0.4 ul 2.5 U/ul Accuprime Pfx polymerase (Invitrogen)

37.7 ul dH20

Thermal cycle in real-time thermal cycler according to the following protocol, but stop cycling before amplification yield plateaus (generally 8-12 cycles):

1. 95 C for 5 min
2. 95 C for 30 sec
3. 58 C for 60 sec
4. 72 C for 60 sec
5. goto 2, N more times

Prepare a Shotgun Next-Generation Sequencing Library for Analysis

Purify desired amplicon population from non-specific amplification products by gel extraction.

Concatemerize amplicons into high-molecular weight products suitable for shearing

Mechanically shear, using either a nebulizer, BioRuptor, Hydroshear, Covaris, or similar instrument. DNA should be sheared into fragments several hundred base-pairs in length.

Ligate adapters required for amplification by the sequencing platform used. If necessary, purify ligated product from unligated product and adapters.

Example 2: Use of Differentiator Tag Sequences to Detect and Correct Bias in a MIP-Capture Reaction of a Set of Exon Targets

The first step in performing the detection/correction is to determine how many differentiator tag sequences are necessary for the given sample. In this example, 1000 genomic targets corresponding to 1000 exons were captured. Since the differentiator tag sequence is part of the probe, it will measure/report biases that occur from the earliest protocol steps. Also, being located in the backbone, the differentiator tag sequence can easily be sequenced from a separate priming site, and therefore not impact the total achievable read-length for the target sequence.

MIP probes are synthesized using standard column-based oligonucleotide synthesis by any number of vendors (e.g. IDT), and differentiator tag sequences are introduced as 'degenerate' positions in the backbone. Each degenerate position increases the total number of differentiator tag sequences synthesized by a factor of 4, so a 10 nt degenerate region implies a differentiator tag sequence complexity of ~1e6 species.

Hybridize Capture Probes to Human Genomic Sample

Assemble Hybridization Reaction:

1.0 ul capture probe mix (~2.5 pmol)

2.0 ul 10×Ampligase buffer (Epicentre)

6.0 ul 500 ng/ul human genomic DNA (~16.7 fmol)

11 ul dH20

In a thermal cycler, heat reaction to 95 C for 5 min to denature genomic DNA, then cool to 60 C. Allow to incubate at 60 C for 40 hours.

Convert Hybridized Probes into Covalently-Closed Circular Products Containing Subtargets

Prepare Fill-In/Ligation Reaction Mixture:

0.25 ul 2 mM dNTP mix (Invitrogen)

2.5 ul 10×Ampligase buffer (Epicentre)

5.0 ul 5 U/ul Taq Stoffel fragment (Applied Biosystems)

12.5 ul 5 U/ul Ampligase (Epicentre)

4.75 ul dH20

Add 1.0 ul of this mix to the hybridized probe reaction, and incubate at 60 C for 10 hours.

Purify Circularized Probe/Subtarget Products from Un-Reacted Probes and Genomic DNA

Prepare Exonuclease Reaction Mixture:

21 ul fill-in/ligation reaction product

2.0 ul 10×exonuclease I buffer (New England Biolabs)

2.0 ul 20 U/ul exonuclease I (New England Biolabs)

2.0 ul 100 U/ul exonuclease III (New England Biolabs)

Incubate at 37 C for 60 min, then heat-inactivate by incubating at 80 C for 15 min. Immediately cool to 4 C for storage.

Amplify Circular Material by PCR Using Primers Specific to the 'Constant Region' of the Probes

Prepare PCR Mixture:

5.0 ul 10×Accuprime reaction buffer (Invitrogen)

1.5 ul 10 uM CP-2-FA (5'-GCACGATCCGACGGTAGTGT-3') (SEQ ID NO: 183)

1.5 ul 10 uM CP-2-RA (5'-CCGTAATCGGGAAGCTGAAG-3') (SEQ ID NO: 184)

0.4 ul 25 mM dNTP mix (Invitrogen)

2.0 ul heat-inactivated exonuclease reaction mix

1.5 ul 10×SybrGreen (Invitrogen)

0.4 ul 2.5 U/ul Accuprime Pfx polymerase (Invitrogen)

37.7 ul dH20

Thermal cycle in real-time thermal cycler according to the following protocol, but stop cycling before amplification yield plateaus (generally 8-12 cycles):

6. 95 C for 5 min
7. 95 C for 30 sec
8. 58 C for 60 sec
9. 72 C for 60 sec
10. goto 2, N more times

Prepare a Shotgun Next-Generation Sequencing Library for Analysis

Purify desired amplicon population from non-specific amplification products by gel extraction.

Concatemerize amplicons into high-molecular weight products suitable for shearing

Mechanically shear, using either a nebulizer, BioRuptor, Hydroshear,

Covaris, or similar instrument. DNA should be sheared into fragments several hundred basepairs in length.

Ligate adapters required for amplification by the sequencing platform used. If necessary, purify ligated product from unligated product and adapters.

Perform Sequencing of Library According to Manufacturer's Directions (e.g. Illumina, ABI, Etc), Reading Both the Target Sequence and the Differentiator Tag Sequence.

Analyze Data by Correcting for any Biases Detected by Quantitation of Differentiator Tag Sequence Abundance

Construct a table of target: differentiator tag abundances from the read data, e.g.:

Target ID	Differentiator Tag sequence ID	Count
1	3547	1
2	4762	1
1	9637	1
1	1078	5
3	4762	1
1	2984	1

All 'count' entries should be '1', since any particular target:differentiator tag mapping will not occur more than once by chance, and therefore will only be observed if bias was present somewhere in the sample preparation process. For any target:differentiator tag combination observed more

than once, all such reads are 'collapsed' into a single read before consensus basecalls are determined. This will cancel the effect of bias on consensus basecall accuracy. FIG. **5** depicts a method for making diploid genotype calls in which repeat target:differentiator tag combination are collapsed.

Example 3: Differentiator Tag Sequence Design for MIP Capture Reactions

For a set of targets, the number of differentiator tag sequences necessary to be confident (within some statistical bounds) that a certain differentiator tag sequence will not be observed more than once by chance in combination with a certain target sequence was determined. The total number of unique differentiator tag sequences for a certain differentiator tag sequence length is determined as $4^{(Length\ in\ nucleotides\ of\ the\ differentiator\ tag\ sequence)}$. For a molecular inversion probe capture reaction that uses MIP probes having differentiator tag sequences, the probability of performing the capture reaction and capturing one or more copies of a target sequence having the same differentiator tag sequence is calculated as: p=1−[N!/(N−M)!]/[NAM], wherein N is the total number of possible unique differentiator tag sequences and M is the number of target sequence copies in the capture reaction. Thus, by varying the differentiator tag sequence length it is possible to perform a MIP capture reaction in which the probability of capturing one or more copies of a target sequence having the same differentiator tag sequence is set at a predetermined probability value.

For example, for a differentiator tag sequence of 15 nucleotides in length, there are 1,073,741,824 possible differentiator tag sequences. A MIP capture reaction in which MIP probes, each having a differentiator tag sequence of 15 nucleotides, are combined with 10000 target sequence copies (e.g., genome equivalents), the probability of capturing one or more copies of a target sequence having the same differentiator tag sequence is 0.05. In this example, the MIP reaction will produce very few (usually 0, but occasionally 1 or more) targets where multiple copies are tagged with the same differentiator tag sequence. FIG. **6** depicts results of a simulation for 100000 capture reactions having 15 nucleotide differentiator tag sequences and 10000 target sequences.

Example 4: Assessment of the Probability for Obtaining Enough Sequencing Reads to Make Accurate Base-Calls at Multiple Independent Loci, as a Function of Sequencing Coverage

Monte Carlo simulations were performed to determine sequencing coverage requirements. The simulations assume 10000 genomic copies of a given locus (target) half mom alleles and half dad alleles. The simulations further assume 1% efficiency of capture for the MIP reaction. The simulation samples from a capture mix 100 times without replacement to create a set of 100 capture products. The simulation then samples from the set of 100 capture products with replacement (assuming unbiased amplification) to generate 'reads' from either mom or dad. The number of reads sampled depends on the coverage. The number of independent reads from both mom and dad necessary to make a high-quality base-call (assumed to be 10 or 20 reads) were then determined. The process was repeated 1000 times for each coverage level, and the fraction of times that enough reads from both parents were successfully obtained was determined. This fraction was raised to the power 1000, assuming we have 1000 independent loci that must obtain successful base-calls, plotted (See FIG. **7**). Result show that roughly 50× coverage is required to capture each allele >=10× with >0.95 probability.

Example 5: MIP Capture of 'Target' Locus and 'Control' Loci

In some embodiments, to accurately quantify the efficiency of target locus capture, at least three sets of control loci are captured in parallel that have a priori been shown to serve as proxies for various lengths of target locus. For example, if the target locus is expected to have a length between 50 and 1000 bp, then sets of control loci having lengths of 50, 250, and 1000 bp could be captured (e.g. 20 loci per set should provide adequate protection from outliers), and their abundance digitally measured by sequencing. These loci should be chosen such that minimal variation in efficiency between samples and on multiple runs of the same sample is observed (and are therefore 'efficiency invariant'). These will serve as 'reference' points that define the shape of the curve of abundance-vs-length. Determining the length of the target is then simply a matter of 'reading' the length from the appropriate point on the calibration curve.

In some embodiments, the statistical confidence one has in the estimate of target length from this method is driven largely by three factors: 1) reproducibility/variation of the abundance data used to generate the calibration curve; 2) goodness of fit of the regression to the 'control' datapoints; 3) reproducibility of abundance data for the target locus being measured. Statistical bounds on 1) and 2) will be known in advance, having been measured during development of the assay. Additionally, statistical bounds on 3) will be known in general in advance, since assay development should include adequate population sampling and measure of technical reproducibility. Standard statistical methods should be used to combine these three measures into a single P value for any given experimental measure of target abundance.

In some embodiments, given the set of calibration observations, and a linear regression fit to that data, the regression can be used to predict the length value for n observations of the target locus whose length is unknown. First, choose an acceptable range for the confidence interval of the length estimate. For example, in the case of distinguishing "normal" (87-93 bp) from "premutation" (165-600 bp) potential cases of Fragile X, the goal is to measure length to sufficient precision to distinguish 93 bp from 165 bp. The predicted response value, computed when n observations is substituted into the equation for the regressed line, will have arbitrary precision. However, if for example a 95% confidence level is desired, that 95% confidence interval must be sufficiently short that it does not overlap both the "normal" and "premutation" length ranges. Continuing the example, if one calculates a length of 190 from n=400 MIP observations, and based on the regression from calibration data, the 95% confidence interval is 190+/−20 bp, one can conclude the sample represents a "premutation" length with 95% certainty. Conversely, if the calibration data were less robust, error estimates of the regression would be higher, leading to larger confidence intervals on the predicted response value. In some embodiments, if the 95% CI were calculated as 190+/−100 bp from n=400, one could not determine whether the predicted response value corresponds to a "normal" or "premutation" length.

In some embodiments, the confidence interval for a predicted response is calculated as:

The estimate for the response $\hat{y}$ is identical to the estimate for the mean of the reponse: $\hat{y}=b_0+b_1x^*$. The confidence interval for the predicted value is given by $\hat{y}\pm t^*s_{\hat{y}}$, where $\hat{y}$ is the fitted value corresponding to x^*. The value t^* is the upper $(1-C)/2$ critical value for the $t(n-2)$ distribution.

In some embodiments, a technique for analyzing a locus of interest can involve the following steps.

Convert Hybridized Probes into Covalently-Closed Circular Products Containing Subtargets

Prepare Fill-In/Ligation Reaction Mixture:

0.25 ul 2 mM dNTP mix (Invitrogen)

2.5 ul 10×Ampligase buffer (Epicentre)

5.0 ul 5 U/ul Taq Stoffel fragment (Applied Biosystems)

12.5 ul 5 U/ul Ampligase (Epicentre)

4.75 ul dH20

Add 1.0 ul of this mix to the hybridized probe reaction, and incubate at 60 C for 10 hours.

Purify Circularized Probe/Subtarget Products from Un-Reacted Probes and 30 Genomic DNA

Prepare Exonuclease Reaction Mixture:

21 ul fill-in/ligation reaction product

2.0 ul 10×exonuclease I buffer (New England Biolabs)

2.0 ul 20 U/ul exonuclease I (New England Biolabs)

2.0 ul 100 U/ul exonuclease III (New England Biolabs)

Incubate at 37 C for 60 min, then heat-inactivate by incubating at 80 C for 15 min. Immediately cool to 4 C for storage.

Amplify Circular Material by PCR Using Primers Specific to the 'Constant Region' of the Probes

Prepare PCR Mixture:

5.0 ul 10×Accuprime reaction buffer (Invitrogen)

1.5 ul 10 uM CP-2-FA-Ilmn (platform-specific amplification sequence plus 'circle constant region'-specific sequence)

1.5 ul 10 uM CP-2-RA-Ilmn (platform-specific amplification sequence plus 'circle constant region'-specific sequence)

0.4 ul 25 mM dNTP mix (Invitrogen)

2.0 ul heat-inactivated exonuclease reaction mix

1.5 ul 10×SybrGreen (Invitrogen)

0.4 ul 2.5 U/ul Accuprime Pfx polymerase (Invitrogen)

37.7 ul dH20

Thermal cycle in real-time thermal cycler according to the following protocol, but stop cycling before amplification yield plateaus (generally 8-12 cycles):

11. 95 C for 5 min

12. 95 C for 30 sec

13. 58 C for 60 sec

14. 72 C for 60 sec

15. goto 2, N more times

Perform Sequencing (e.g., Next-Generation Sequencing) on Sample for Digital Quantitation According to Manufacturer's Instructions (e.g., Illumina, ABI)

Example 6: MIP-capture reaction of a set of exon target nucleic acids

MIP probes are synthesized using standard column-based oligonucleotide synthesis by any number of vendors (e.g. IDT).

Hybridize Capture Probes to Human Genomic Sample

Assemble Hybridization Reaction:

1.0 ul capture probe mix (~2.5 pmol)

2.0 ul 10×Ampligase buffer (Epicentre)

6.0 ul 500 ng/ul human genomic DNA (~16.7 fmol)

11 ul dH20

In a thermal cycler, heat reaction to 95 C for 5 min to denature genomic DNA, then cool to 60 C. Allow to incubate at 60 C for 40 hours.

Convert Hybridized Probes into Covalently-Closed Circular Products Containing Target Nucleic Acids

Prepare Fill-In/Ligation Reaction Mixture:

0.25 ul 2 mM dNTP mix (Invitrogen)

2.5 ul 10×Ampligase buffer (Epicentre)

5.0 ul 5 U/ul Taq Stoffel fragment (Applied Biosystems)

12.5 ul 5 U/ul Ampligase (Epicentre)

4.75 ul dH20

Add 1.0 ul of this mix to the hybridized probe reaction, and incubate at 60 C for 10 hours.

Purify, Circularized Probe/Target Nucleic Acid Products from Un-Reacted Probes and Genomic DNA

Prepare Exonuclease Reaction Mixture:

21 ul fill-in/ligation reaction product

2.0 ul 10×exonuclease I buffer (New England Biolabs)

2.0 ul 20 U/ul exonuclease I (New England Biolabs)

2.0 ul 100 U/ul exonuclease III (New England Biolabs)

Incubate at 37 C for 60 min, then heat-inactivate by incubating at 80 C for 15 min. Immediately cool to 4 C for storage.

Amplify Circular Material by PCR Using Primers Specific to the 'Constant Region' Of the Probes

Prepare PCR Mixture:

5.0 ul 10×Accuprime reaction buffer (Invitrogen)

1.5 ul 10 uM CP-2-FA (5'-GCACGATCCGACGGTAGTGT-3') (SEQ ID NO: 183)

1.5 ul 10 uM CP-2-RA (5'-CCGTAATCGGGAAGCTGAAG-3') (SEQ ID NO: 184)

0.4 ul 25 mM dNTP mix (Invitrogen)

2.0 ul heat-inactivated exonuclease reaction mix

1.5 ul 10×SybrGreen (Invitrogen)

0.4 ul 2.5 U/ul Accuprime Pfx polymerase (Invitrogen)

37.7 ul dH20

Thermal cycle in real-time thermal cycler according to the following protocol, but stop cycling before amplification yield plateaus (generally 8-12 cycles):

16. 95 C for 5 min

17. 95 C for 30 sec

18. 58 C for 60 sec

19. 72 C for 60 sec

20. goto 2, N more times

Prepare a Shotgun Next-Generation Sequencing Library for Analysis

Purify desired amplicon population from non-specific amplification products by gel extraction.

Concatemerize amplicons into high-molecular weight products suitable for shearing

Mechanically shear, using either a nebulizer, BioRuptor, Hydroshear, Covaris, or similar instrument. DNA should be sheared into fragments several hundred base-pairs in length.

Ligate adapters required for amplification by the sequencing platform used. If necessary, purify ligated product from unligated product and adapters.

Perform Sequencing of Library According to Manufacturer's Directions (e.g. Illumina, ABI, Etc), Reading the Target Sequence to Determine Abundance of the Target Nucleic Acid.

Example 7: Use of MIPs, Hybridization, and Mutation-Detection MIPs to Genotype a Set of 1000 Targets

MIPs, hybridization, and mutation-detection MIPs are used to genotype a set of 1000 targets. The protocol permits detection of any of 50 specific known point mutations

First, separate MIP, hybridization, and mutation-detection MIP reactions are performed on a biological sample. A MIP capture reaction is performed essentially as described in Turner et al 2009, Nature methods 6:315-6. A set of MIPs is designed such that each probe in the set flanks one of the 1000 targets. Separately, a hybridization enrichment reaction is performed using the Agilent SureSelect procedure. Prior to selection, the genomic DNA to be enriched is converted into a shotgun sequencing library using Illumina's 'Fragment Library' kit and protocol. Agilent's web interface is used to design a set of probes which will hybridize to the target nucleic acids. Separately, a set of probes are designed (mutation-detection MIPs) which will form MIPs only if mutations (e.g., specific polymorphisms) are present. Each mutation-detection MIP has a 3'-most base identity that is specific for a single known mutation. A reaction with this set of mutation-detection MIPs is performed to selectively detect the presence of any mutant alleles.

Once all three reactions have been performed, the two MIP reactions are combined (e.g., at potentially non-equimolar ratios to further increase sensitivity of mutation detection) into a single tube, and run as one sample on the next-generation DNA sequencing instrument. The hybridization-enriched reaction is run as a separate sample on the next-generation DNA sequencing instrument. Reads from each 'sample' are combined by a software algorithm which forms a consensus diploid genotype at each position in the target set by evaluating the total coverage at each position, the origin of each read in that total coverage, the quality score of each individual read, and the presence (or absence) of any reads derived from mutation-specific MIPs overlapping the region.

It should be appreciated that the preceding examples are non-limiting and aspects of the invention may be implemented as described herein using alternative techniques and/or protocols that are available to one or ordinary skill in the art.

It will be clear that the methods may be practiced other than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present disclosure are possible in light of the above teachings and, therefore, are within the scope of the claims. Preferred features of each aspect of the Disclosure are as for each of the other aspects mutatis mutandis. The documents including patents, patent applications, journal articles, or other disclosures mentioned herein are hereby incorporated by reference in their entirety. In the event of conflict, the disclosure of present application controls, other than in the event of clear error.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 184

<210> SEQ ID NO 1
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 1

tcacatcacc aagttaaaaa aaaaaarggg gcgggggggc agaatgaaaa tt 52

<210> SEQ ID NO 2
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 2

aaacaaggat gaattaagtt ttttttaaa aaagaaacat ttggtaaggg ga 52

<210> SEQ ID NO 3
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 3

atcaccaagt taaaaaaaaa aaagggrcgg gggggcagaa tgaaaattgc at 52

<210> SEQ ID NO 4
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 4

ctattgaacc agaaccaaac aggaatrcca tagcattttg taaactaaac tg 52

-continued

<210> SEQ ID NO 5
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 5

caggagttca agaccagcct actaaamcac acacacacac acacacacac ac 52

<210> SEQ ID NO 6
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 6

gattaaataa tagtgtttat gtacccsgct tataggagaa gagggtgtgt gt 52

<210> SEQ ID NO 7
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 7

attgttatct tttcatataa ggtaacwgag gcccagagag attaaataac at 52

<210> SEQ ID NO 8
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 8

taattttaat taagtaaatt taattgrtag ataaataagt agataaaaaa ta 52

<210> SEQ ID NO 9
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 9

gtataaaaaa aaaaaaaaaa aaagttwgaa tgttttcttg cattcagagc ct 52

<210> SEQ ID NO 10
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 10

atactaaaaa tttaaagttc tcttgcrata tattttctta atatcttaca tc 52

<210> SEQ ID NO 11
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 11

tgctgggatt acaggcgtga gccaccrcgc ctggcctgat gggacatatt tt 52

<210> SEQ ID NO 12
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 12

ctacaatata agtatagtat tgcaaamcca tcaggaaggg tgttaactat tt 52

-continued

<210> SEQ ID NO 13
<211> LENGTH: 54
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(29)
<223> OTHER INFORMATION: This region may or may not be present in its entirety

<400> SEQUENCE: 13

gttgtctcca aacttttttt caggtgagaa ggtggccaac cgagcttcgg aaag 54

<210> SEQ ID NO 14
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 14

ttttttcata aaagattata taaaggmtat tgcttttgaa tcacaaacac ta 52

<210> SEQ ID NO 15
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 15

atctagtgag cagtcaggaa agagaayttc cagatcctgg aaatcagggt ta 52

<210> SEQ ID NO 16
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 16

tagagtataa aaaaaaaaaa aaaaaaagtt tgaatgtttt cttgcattca ga 52

<210> SEQ ID NO 17
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 17

tgcaaatgcc aactatcaaa gatattsgag tatactgtca ataaacttca ta 52

<210> SEQ ID NO 18
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 18

tcctcaacag ttagaaacaa tattttsagt gatttcccat gccaacttta ct 52

<210> SEQ ID NO 19
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 19

tttttggtat tgttgttaaa taagtgrgaa ttcaatacag tataatgtct gt 52

-continued

<210> SEQ ID NO 20
<211> LENGTH: 87
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(72)
<223> OTHER INFORMATION: This region may or may not be present in its entirety

<400> SEQUENCE: 20

cttgaaatcg gatatatata tatatatgta tatatatata tatatatata tatatatata 60

tagtattatc cctgttttca cagtttt 87

<210> SEQ ID NO 21
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 21

agaggggctg tgaaggacac caaggargag actaagagcc aggagggaaa ac 52

<210> SEQ ID NO 22
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 22

tagagtttat tagcttttac tactctrctt agttactttg tgttacagaa ta 52

<210> SEQ ID NO 23
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(28)
<223> OTHER INFORMATION: This region may or may not be present in its entirety

<400> SEQUENCE: 23

actagtgatg atgagcttct tttcatatgt ttgttggctg cataaatgtc ttc 53

<210> SEQ ID NO 24
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 24

gcagagaaaa gaggggctgt gaaggasacc aaggaggaga ctaagagcca gg 52

<210> SEQ ID NO 25
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 25

ttcagaggtc taccactggt gcatackcta atcacagtgt cgaaaatttt ac 52

<210> SEQ ID NO 26
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 26

-continued

aagaaagagc aaaagagggc aaacttytca tacatttttg atgtcgaaac ca 52

<210> SEQ ID NO 27
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 27

cctaaagttt aaaaagaaaa aaaaaaagga agaaggaatt aaaaatccaa ag 52

<210> SEQ ID NO 28
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 28

gtgtgtgtgt gtatatatat atatatwtat atatttttt tttcctgagc ca 52

<210> SEQ ID NO 29
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 29

aaactgttga tgttttcatt tatttasatc attggaaaac tttagattct ag 52

<210> SEQ ID NO 30
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 30

tttatccatt cttaaccaga acagacrttt tttcagagct ggtccaggaa aa 52

<210> SEQ ID NO 31
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 31

ttgaaatcgg atatatatat atatatrtat atatatatat atatatatat at 52

<210> SEQ ID NO 32
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 32

tagtttttta tttcctcata ttattttcag tggctttttc ttccacatct tt 52

<210> SEQ ID NO 33
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 33

-continued

acatagttct cagtggtaca actacargtg atttctcttt tcttatttct gg 52

<210> SEQ ID NO 34
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 34

agagcaatgg catcccttgt cttgtgytat acaggatgca gcaatttata gg 52

<210> SEQ ID NO 35
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 35

ttctgtctac ataagatgtc atactaratt atcttttcca gcatgcattc ag 52

<210> SEQ ID NO 36
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 36

cagggtggca tgttaggcag tgcttaraat aaatgagttg gttatacaag ta 52

<210> SEQ ID NO 37
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(28)
<223> OTHER INFORMATION: This region may or may not be present in its entirety

<400> SEQUENCE: 37

aggacacaca cacacacaca cacacacatg cacacacatt taaatagatg cat 53

<210> SEQ ID NO 38
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 38

taaaaaattg gtataatgaa attgcayttg tagtctttgg acatttaaat cc 52

<210> SEQ ID NO 39
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 39

tttcaatact taagaggtac gcagagraaa gaggggctgt gaaggacacc aa 52

<210> SEQ ID NO 40
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 40

cagcaatgaa tattttgagg ctgaggygct gaggggtaaa attgcagcct gg 52

<210> SEQ ID NO 41

-continued

<211> LENGTH: 55
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(30)
<223> OTHER INFORMATION: This region may or may not be present in its entirety

<400> SEQUENCE: 41

ttatggttta tatttttgtg tcttctcttt aacacatctt ttctagcaga attca 55

<210> SEQ ID NO 42
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 42

gtattttagt ttttttttt tgtttgtttt gttttgtttt gttttgtttt tg 52

<210> SEQ ID NO 43
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 43

tgggtgactc catttttact tttagtytgg tctgttgagg cctcgtgaga ga 52

<210> SEQ ID NO 44
<211> LENGTH: 55
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(30)
<223> OTHER INFORMATION: This region may or may not be present in its entirety

<400> SEQUENCE: 44

tattttcatg tattttagtt ttttttttt gtttgttttg ttttgttttg ttttg 55

<210> SEQ ID NO 45
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 45

gtgtgtgtgt atatatatat atatatwtat attttttttt tcctgagcca aa 52

<210> SEQ ID NO 46
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 46

aacctattag catgtctggc agaaaaatag atacttaata aatttcttaa at 52

<210> SEQ ID NO 47
<211> LENGTH: 52

-continued

<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 47

gaggcttaga cagtttaagt aactcargca tggttacaca actagctagg gc 52

<210> SEQ ID NO 48
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 48

gtgtgagtat tatgagacca tatgttrgga gattttattt ggtattgagg at 52

<210> SEQ ID NO 49
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 49

gaaaccccac cccttctata gttttcyctt taatatttac aatggaacca tt 52

<210> SEQ ID NO 50
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 50

catatatata tagtgtgtgt gtgtgtrtat atatatatat atatatattt tt 52

<210> SEQ ID NO 51
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 51

cgagcgagag agagagagag agagagagga gccggcgcga gaactacgca tgc 53

<210> SEQ ID NO 52
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 52

ggcaggtaat atctagtacc ttacttwtat ttcctgagca cattctacat tt 52

<210> SEQ ID NO 53
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 53

ggccaggaat gggagtgctt aggtgcrgag gtggcactgt tcccgcagct gc 52

<210> SEQ ID NO 54
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 54

tgaaaactcc atcccctcag ggtcatyaga tgaagagaag accacagggg tt 52

<210> SEQ ID NO 55

-continued

<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 55

tgtaggtaag ggtcacatgt gggagasgca gctgtgggta ggtcagccct gt 52

<210> SEQ ID NO 56
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 56

ccaagaaggc gccattacac tccagcctgg gcgacagggc gagactccct ca 52

<210> SEQ ID NO 57
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 57

tgccacaccc agctaatttg tgtgtggtat gtgtgtgtat gtatgtgtgt gt 52

<210> SEQ ID NO 58
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 58

gttcctccag taatttttttt ttttttttggt tttgagacag agtcttgccc tg 52

<210> SEQ ID NO 59
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 59

atcatgccca gataattttt ttttttttgta ttttagtaga cacagggttt ca 52

<210> SEQ ID NO 60
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(28)
<223> OTHER INFORMATION: This region may or may not be present in its entirety

<400> SEQUENCE: 60

cgagcgagag agagagagag agagagagga gccggcgcga gaactacgca tgc 53

-continued

<210> SEQ ID NO 61
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 61

cctgtgaggg gcacattcct tagtagctaa ggagttgggg gtgtgaagat cc 52

<210> SEQ ID NO 62
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 62

acaggctact ggctgggccc aggcaaaggg ggccttggca ggaaaagttc ct 52

<210> SEQ ID NO 63
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(28)
<223> OTHER INFORMATION: This region may or may not be present in its entirety

<400> SEQUENCE: 63

gcgagagaga gagagagaga gagaggagag ccggcgcgag aactacgcat gcg 53

<210> SEQ ID NO 64
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 64

aagaagaaaa ataaaaagaa agtgggycag accgagagaa caggaagcct ga 52

<210> SEQ ID NO 65
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 65

aaggacaaag gcaaagagac aaaggckcaa cactgggggt ccccagagag tg 52

<210> SEQ ID NO 66
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 66

tacctagtca cttcctgcct ccatggygca aaaggggatg ggtgtgcctc tt 52

<210> SEQ ID NO 67
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

-continued

<400> SEQUENCE: 67

ctcttccgag gttccaccct gaacacyttc ctgctccctc gtggtgtaga gt 52

<210> SEQ ID NO 68
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 68

ttctgactgg caaccagccc cactctytgg gagccctcag gaatgaactt gc 52

<210> SEQ ID NO 69
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 69

gctcagcctc ccaggctgga gtgcagwggc gcgatctcgg ctcaccgcaa cc 52

<210> SEQ ID NO 70
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 70

catgaaccac atcaaatgag atttagyggg agtggcacac acagtcatga cc 52

<210> SEQ ID NO 71
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 71

aagcagccct ggggagtcgg ggcgggrcct ggattggaaa agagacggtc ac 52

<210> SEQ ID NO 72
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 72

ctccaagttc tgggagcaga gtgtgcrgct aggctcctgg gatcgaggga tg 52

<210> SEQ ID NO 73
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 73

gttcctccag taatttttttt ttttttkgtt ttgagacaga gtcttgccct gt 52

<210> SEQ ID NO 74
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 74

ctagtacctt acttccctca agttcawtca tctcacagat atttcctgag ca 52

-continued

<210> SEQ ID NO 75
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 75

aattagccgt gcgtggtggc gggtgcytgt aatcccacgt acttgggagg ct 52

<210> SEQ ID NO 76
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 76

ccatggccag ccggggaggg gacgggmaca cacagaccca cacagagact ca 52

<210> SEQ ID NO 77
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (77)..(77)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 77

agcgagagag agagagagag agagagrgag ccggcgcgag aactacgcat gc 52

<210> SEQ ID NO 78
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 78

cgtagcagtt agcagatgat aggcggbgaa atcttatttc acagggcatt aa 52

<210> SEQ ID NO 79
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 79

ctggccctgg tgacagtggg gctgtgygtg gggccagagc cttctcagag gt 52

<210> SEQ ID NO 80
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 80

cagatactgg ccctggtgac agtgggrctg tgcgtggggc cagagccttc tc 52

<210> SEQ ID NO 81
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 81

gacagatact ggccctggtg acagtgkggc tgtgcgtggg gccagagcct tc 52

<210> SEQ ID NO 82
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:

-continued

<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(28)
<223> OTHER INFORMATION: This region may or may not be present in its entirety

<400> SEQUENCE: 82

ggctcctctc tctctctctc tctctctcgc tcgctctctc gctctctcgc tct 53

<210> SEQ ID NO 83
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 83

gtttcagacc agcctggcca acatagygaa accccatctc tactaaaaat aa 52

<210> SEQ ID NO 84
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 84

agtgggcgat tggatggagc tgagtayggg gcccatccag gctaatcaca cc 52

<210> SEQ ID NO 85
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 85

ccgggctccg tgaatgtttg tcacatstct gaagaacgta tgaattacat aa 52

<210> SEQ ID NO 86
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 86

gaatcccaac cccgacgctc gtcgccsggc tccgtgaatg tttgtcacat gt 52

<210> SEQ ID NO 87
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 87

gcgaatccca accccgacgc tcgtcgycgg gctccgtgaa tgtttgtcac at 52

<210> SEQ ID NO 88
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 88

tgggcagaag tcagggtcca aagaaakggc aaagaaaagt gtcagtggct ca 52

<210> SEQ ID NO 89
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 89

taagaaagac gtttttgatt tttttcrgac ttctctggct ccactaccct gc 52

-continued

<210> SEQ ID NO 90
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 90

ctgattcctg gccaggagcg gtggctyacg cctgtaatcc cagcgctttg gg 52

<210> SEQ ID NO 91
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 91

taaaaatgct gattcctggc caggagyggt ggctcacgcc tgtaatccca gc 52

<210> SEQ ID NO 92
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 92

tttaaaaatg ctgattcctg gccaggmgcg gtggctcacg cctgtaatcc ca 52

<210> SEQ ID NO 93
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 93

caagacctgt caaagatctg agaaatwtta cccgacttac aagctaacca tt 52

<210> SEQ ID NO 94
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 94

actgtaataa gtgctgtaaa agaaatrcac aaaataatat agcagagggt at 52

<210> SEQ ID NO 95
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 95

cttgaggtca ggagttcaag accagtytgg gcaacatggg gaaaaccttg tc 52

<210> SEQ ID NO 96
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 96

aggttgcagt gagccgagat catgccrttg cactccagcc ggggcaacaa aa 52

<210> SEQ ID NO 97
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 97

acaagtgtct tgaaattatc tgtgatytgc tatagagcaa tacttttgta aa 52

-continued

<210> SEQ ID NO 98
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 98

gtgggtatat gcagctctat gcactayctg ctcatttatt tggtaaatct aa 52

<210> SEQ ID NO 99
<211> LENGTH: 59
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(34)
<223> OTHER INFORMATION: This region may or may not be present in its entirety

<400> SEQUENCE: 99

tgtgtgtgtg tgcgtgtgtg tgtgtgtgtg tgtgatcata agagtggctg cagcaaact 59

<210> SEQ ID NO 100
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 100

agtctggagt gcaatggtgc aatctcrgct cactgcagcc tccacctccg gg 52

<210> SEQ ID NO 101
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 101

ctcctaatgg atatttccta aattttkctg aacagaattt aacttgagct gg 52

<210> SEQ ID NO 102
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 102

atttaaaaat ggatttctag aaaaacratc acatacttga atattttagc aa 52

<210> SEQ ID NO 103
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 103

ctataaatgg gtagcatgag ggattcragg aggtggctga aagaagcacg ta 52

<210> SEQ ID NO 104
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 104

aagaaaccaa gcatagtaga gtgttaraaa accaaagcaa ctaaacaact gt 52

<210> SEQ ID NO 105
<211> LENGTH: 52

-continued

<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 105

cggggctcag aacttgtaac agaaaawtaa aatatactcc actcaaggga at 52

<210> SEQ ID NO 106
<211> LENGTH: 57
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(32)
<223> OTHER INFORMATION: This region may or may not be present in its entirety

<400> SEQUENCE: 106

tactacactt cacggatact gtacttgtac ttttttttcca aattgaaggt ttttggc 57

<210> SEQ ID NO 107
<211> LENGTH: 63
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(38)
<223> OTHER INFORMATION: This region may or may not be present in its entirety

<400> SEQUENCE: 107

ttgtttttgt ttttgttttt gtttttgttt ttgtttttttg agatggagtc tcgctctgtc 60

gcc 63

<210> SEQ ID NO 108
<211> LENGTH: 57
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(32)
<223> OTHER INFORMATION: This region may or may not be present in its entirety

<400> SEQUENCE: 108

tttgccttac tacacttcac ggatactgta cttgtacttt ttttccaaat tgaaggt 57

<210> SEQ ID NO 109
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 109

gaggtggctg aaagaagcac gtatccctga tggcatggtt gcgggttata tg 52

<210> SEQ ID NO 110
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 110

-continued

gagaaaagca gttcctggaa cacccccacc ccttaacccc ttatctctgc tt 52

<210> SEQ ID NO 111
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 111

ttacatatgt atacatgtgc catgtttggt gtgccgcacc cattaactcg tc 52

<210> SEQ ID NO 112
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 112

ctttttccag attttttttt ttttttgag acagagtttc actcttgttg cc 52

<210> SEQ ID NO 113
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 113

tttggaaatc ttaagctttt atttgggtgt cacagagaaa caggatctgt at 52

<210> SEQ ID NO 114
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 114

tactttaagt tttagggtac atgtgcacca tgtgcaggtt tgttacatat gt 52

<210> SEQ ID NO 115
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 115

attcatgacc agccacataa atgcacagta ttacttcgca agcatgccaa tg 52

<210> SEQ ID NO 116
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:

-continued

<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 116

gtgcactaga attagctaaa gtgggggaaa aaaagatgca tttgatggtc ta 52

<210> SEQ ID NO 117
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 117

aacctccacc tcccaggttc aagagaattc tcctgcctca gcctcccaag ta 52

<210> SEQ ID NO 118
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 118

cctccctgtg atccgaagta gcagacrtac ttaacttcca tggtggattg tt 52

<210> SEQ ID NO 119
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 119

aaaacattat tatatctaga aaaaaaatgt atcttaacca ttgtgggaag tg 52

<210> SEQ ID NO 120
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 120

ttgaaggtaa aatcataggg agttgggagc tgtcctcttg cgctgaatca gt 52

<210> SEQ ID NO 121
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(28)
<223> OTHER INFORMATION: This region may or may not be present in its entirety

<400> SEQUENCE: 121

acttgtggcc tttttggaga ggttagcaac tctgaaaact ctgtccctgg acc 53

<210> SEQ ID NO 122
<211> LENGTH: 52
<212> TYPE: DNA

-continued

<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 122

gaaggagaaa aagagaggaa ataagttaaa ataataaaca caattaataa ag 52

<210> SEQ ID NO 123
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 123

tgtatacatg tgccatgttg gtgtgcygca cccattaact cgtcatttag ca 52

<210> SEQ ID NO 124
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 124

tcacctgtca cctcctatag aactttcccc tgaccctcct ctatagcatt aa 52

<210> SEQ ID NO 125
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 125

ataaatgatc atcattcaca gtaggggttt tgttttgttt tttttctgga aa 52

<210> SEQ ID NO 126
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 126

acagacatat ctacaaacac acttttttcac atatttgtgt aagtcattta tg 52

<210> SEQ ID NO 127
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 127

aaagacaact aaactaacgc tcaatgmaaa aagtattcgc tgctgtttac at 52

<210> SEQ ID NO 128
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 128

-continued

taccgtgtac cccgtgtttg tgaatgmggc cgcatattac gaaaagaaag aa 52

<210> SEQ ID NO 129
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 129

cagggctgga ggtaaaacca tttattrcta accccagagc agtgaagaag tg 52

<210> SEQ ID NO 130
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 130

ttcttcattg cctattgaag agagagygga atgctttggt tgccagatat gg 52

<210> SEQ ID NO 131
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 131

cacaagatct cattactcag gagctgycca agtgtctaat gtacttagtt aa 52

<210> SEQ ID NO 132
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 132

ttctgtgtaa catttcattt aagcaargga ttcggcaaat caaaaattgt ca 52

<210> SEQ ID NO 133
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 133

taaaacgtat tgaaggtatt attgackctg ttgaagcaaa gagaacaaaa ca 52

<210> SEQ ID NO 134
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 134

atctgctctt ccagttggat gacaagyctt gctgtctaac acctgctgca ga 52

<210> SEQ ID NO 135
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 135

ccatttttg ttgtattttt tttttcytga atacttttta tcgcagttgg tt 52

<210> SEQ ID NO 136
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

-continued

<400> SEQUENCE: 136

ccctgtctct aaaagaaaaa aaaaaaraaa aaaaaaagaa aacaaaaccc aa 52

<210> SEQ ID NO 137
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 137

agtggctcca aaaaggtcat ggaaccyctt gaggatgatg caaattgact ct 52

<210> SEQ ID NO 138
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 138

taaagttact tttcttttat tgactttccc ttatttttta accttatgct tt 52

<210> SEQ ID NO 139
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 139

cagagttaaa aaaaaaaaaa aaaaaaagga agtagcagca acagcttgga aa 52

<210> SEQ ID NO 140
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 140

gttgcccagg gttgagtgca gaggcayatc ttggctcaca gcaacctctg cc 52

<210> SEQ ID NO 141
<211> LENGTH: 55
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(30)
<223> OTHER INFORMATION: This region may or may not be present in its entirety

<400> SEQUENCE: 141

aaggcttttt ttttttttttt ttttttttttt gagacagagt cttgctgtgt cacccc 55

<210> SEQ ID NO 142
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 142

gcctacattc tgcaaagagg agggaasatt cacagctcca tacttgaacc ct 52

<210> SEQ ID NO 143
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

-continued

<400> SEQUENCE: 143

ccaaaggaga atagctctag gggaggsagg tggatgagta tgcatggggg ag 52

<210> SEQ ID NO 144
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 144

gactccatct caaaaaaaaa aaaaaaatgc agtctaatgg cagaattaga ct 52

<210> SEQ ID NO 145
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 145

ttatttattt atttatttat ttttgaraca gggtctctgt tgtccaggct gg 52

<210> SEQ ID NO 146
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 146

ttttgaggca gggtctcact ctgttgycca gggttgagtg cagaggcaca tc 52

<210> SEQ ID NO 147
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 147

cgccttgcga aggccccaca gcttgcytgt gacaaacgtt cataggcaaa tg 52

<210> SEQ ID NO 148
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 148

ggaggtctgt acaaagcacc acctacytca tgggtcagtt tccacagcag aa 52

<210> SEQ ID NO 149
<211> LENGTH: 55
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(30)
<223> OTHER INFORMATION: This region may or may not be present in its entirety

<400> SEQUENCE: 149

gaatcttata attcactgtg tacctccctc tgtttcatat tttcgcaatt gaact 55

<210> SEQ ID NO 150
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

-continued

<400> SEQUENCE: 150

aacatagtat ctaatatagc tttacaycca aagccaaaat atgaatacac tg 52

<210> SEQ ID NO 151
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 151

ttgttttgtt ttgtttgggg ggggggggttg tttttctgag agggagtctt gc 52

<210> SEQ ID NO 152
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 152

cataccaaag ggcagctgga gggatayaga cggaagtcat gtggagagtg aa 52

<210> SEQ ID NO 153
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 153

caggtgtgag ccaccacgac caccaawtta gctcttttta ctccttccct tc 52

<210> SEQ ID NO 154
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 154

agtggtagct gattttgctt ctggatmctt tgccaccttc ccactcttta at 52

<210> SEQ ID NO 155
<211> LENGTH: 64
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(39)
<223> OTHER INFORMATION: This region may or may not be present in its entirety

<400> SEQUENCE: 155

aaagacctgt ttcttaaaaa aaaaaaagaa aaaaaaaaag aaagaaaaga aaagaaaaaa 60

acag 64

<210> SEQ ID NO 156
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 156

taaaaaatct ttcaatgagg agatgtyccc agagcaagac agctgtagga tg 52

<210> SEQ ID NO 157
<211> LENGTH: 52
<212> TYPE: DNA

-continued

<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 157

aaaagaaaaa aaaaaaaaaa aaaaaaagaa aacaaaaccc aaacccataa ag 52

<210> SEQ ID NO 158
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 158

cctgtctcta aaagaaaaaa aaaaaaraaa aaaaaagaaa acaaaaccca aa 52

<210> SEQ ID NO 159
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(28)
<223> OTHER INFORMATION: This region may or may not be present in its entirety

<400> SEQUENCE: 159

gttatcatag aaaaatatca cactctgtct gtatccccac ttccagaaac tgt 53

<210> SEQ ID NO 160
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 160

caggagctca tagaattaca tacaatcttt ttttttttt tttgagacag cg 52

<210> SEQ ID NO 161
<211> LENGTH: 55
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(30)
<223> OTHER INFORMATION: This region may or may not be present in its entirety

<400> SEQUENCE: 161

ttgagaatct tataattcac tgtgtacctc cctctgtttc atattttcgc aattg 55

<210> SEQ ID NO 162
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 162

ccactaccac agtgcctaga gaacaayatg tgtttaataa tatttaaata at 52

<210> SEQ ID NO 163
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

-continued

<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 163

ccctgtctct aaaagaaaaa aaaaaaaaaa aaaaaaagaa aacaaaaccc aa 52

<210> SEQ ID NO 164
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 164

ccattatatc attcatttcc cactcatttt cttcattcca accaagatat at 52

<210> SEQ ID NO 165
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 165

tccgtctcaa aaaaaaaaaa aaaaagagaa aggaattatt ctcatgtata ca 52

<210> SEQ ID NO 166
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 166

ctggggcagt taaaaagaaa aacaaacccc tggtccctgc ccttgaggag at 52

<210> SEQ ID NO 167
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 167

ctccagggtc ccattccagg accacacgcc tgctacctct gcagctcact ca 52

<210> SEQ ID NO 168
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 168

ggattgacat ataccagtta gacggatttt ttttttccat aaaccaggct ca 52

-continued

<210> SEQ ID NO 169
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 169

acaaataatt actacatatc tacaacrttc cagatacaga agaaatggcc aa 52

<210> SEQ ID NO 170
<211> LENGTH: 54
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(29)
<223> OTHER INFORMATION: This region may or may not be present in its entirety

<400> SEQUENCE: 170

taaacacact tgaaacatca tataaaatga tattactaca agacttaacc gtaa 54

<210> SEQ ID NO 171
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 171

acacaggtaa tccatgttta ttatagaaaa atgccacatt actctttatt ga 52

<210> SEQ ID NO 172
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(28)
<223> OTHER INFORMATION: This region may or may not be present in its entirety

<400> SEQUENCE: 172

agttatcata gaaaaatatc acactctgtc tgtatcccca cttccagaaa ctg 53

<210> SEQ ID NO 173
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 173

aatgaactta caggaaggta atatatggga aataaacatc ttattgaatt ta 52

<210> SEQ ID NO 174
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 174

-continued

ggacccctga aaggcacaag acacctttc aggttcacac ttcctgaaag ct 52

<210> SEQ ID NO 175
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 175

ccaccaatca ccagagcctt ctgctcrggg gtacctgagg gaaaacaagc aa 52

<210> SEQ ID NO 176
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 176

aaagactgaa aaaacattca taactatttt tcttgttatc ctcggaaatg tc 52

<210> SEQ ID NO 177
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 177

tatcttcatc ttggaggaga tgaggtygat ttcacctgct ggaagtccaa cc 52

<210> SEQ ID NO 178
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 178

ttgcctatga acgtttgtca cacttcygct gtgagttgct gaggcgaggt gt 52

<210> SEQ ID NO 179
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 179

gcttgctgtt ggatacatct cgccatyacc tgccactctc tagcatcctg ga 52

<210> SEQ ID NO 180
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 180

ccggggcttg ctgttggata catctckcca ttacctgcca ctctctagca tc 52

<210> SEQ ID NO 181
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 181

ctccgtttcg gtttcacttc 20

<210> SEQ ID NO 182

-continued

<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 182

atcttctctt cagccctgct 20

<210> SEQ ID NO 183
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 183

gcacgatccg acggtagtgt 20

<210> SEQ ID NO 184
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 184

ccgtaatcgg gaagctgaag 20

What is claimed is:

1. A method for correcting for errors or bias introduced during nucleic acid analysis workflow, the method comprising the steps of:

obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin;

introducing a set of differentiator tags, wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag;

amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons;

sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and

correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count.

2. The method of claim **1**, wherein said correcting step further comprises collapsing all sequence reads comprising a same differentiator tag into a single read before determining a consensus base call for the target nucleic acid molecule sequences.

3. The method of claim **2**, further comprising the step of determining the presence of a sequencing or amplification error based upon a number of said differentiator tag sequences.

4. The method of claim **1**, wherein said biological sample is selected from the group consisting of blood and biopsy tissue.

5. The method of claim **1**, wherein said plurality of target nucleic acid molecules comprises bacterial or viral nucleic acids isolated from said biological sample.

6. The method of claim **1**, wherein said plurality of target nucleic acid molecules are circulating tumor nucleic acid molecules.

7. The method of claim **1**, further comprising the step of determining the copy number of a genomic region or transcript in a patient from whom the sample is obtained.

8. The method of claim **1**, wherein said plurality of target nucleic acid molecules are maternally-circulating fetal or placental nucleic acid molecules.

9. A method for correcting for errors or bias introduced during nucleic acid analysis workflow, the method comprising the steps of:

obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin;

introducing a set of unique differentiator tags, wherein members of said set of unique differentiator tags are randomly associated with members of said plurality of target nucleic acid molecules, such that any given tag is uniquely associated with said locus of origin;

amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons;

sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and

correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule.

* * * * *

Bibliographic Data

Application No: 16/952,764

Foreign Priority claimed: ○ Yes ◉ No

35 USC 119 (a-d) conditions met: ☐ Yes ☑ No ☐ Met After Allowance

Verified and Acknowledged: /Steven Pohnert/ (Examiner's Signature) | scp (Initials)

Title: METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS

FILING or 371(c) DATE	CLASS	GROUP ART UNIT	ATTORNEY DOCKET NO.
11/19/2020	506	1634	MLB-002/06US 34468/40
RULE			

APPLICANTS

Molecular Loop Biosciences, Inc., Cambridge, MA, UNITED STATES

INVENTORS

Gregory Porreca, Cambridge, MA, UNITED STATES

Uri Laserson, Boston, MA, UNITED STATES

CONTINUING DATA

This application is a CON of 15231687 08/08/2016ABN

15231687 is a CON of 13266862 03/13/2012

13266862 is a 371 of PCT/US10/01293 04/30/2010

PCT/US10/01293 has PRO of 61182089 05/28/2009

PCT/US10/01293 has PRO of 61179358 05/18/2009

PCT/US10/01293 has PRO of 61178923 05/15/2009

PCT/US10/01293 has PRO of 61174470 04/30/2009

FOREIGN APPLICATIONS

IF REQUIRED, FOREIGN LICENSE GRANTED**

11/30/2020

**** SMALL ENTITY ****

STATE OR COUNTRY

UNITED STATES

ADDRESS

Withers Bergman LLP

101 Federal Street, Suite 1900

Boston, MA 02110

UNITED STATES

FILING FEE RECEIVED

$3,000

Electronic Acknowledgement Receipt	
EFS ID:	41170965
Application Number:	16952764
International Application Number:	
Confirmation Number:	7812
Title of Invention:	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS
First Named Inventor/Applicant Name:	Gregory Porreca
Customer Number:	21710
Filer:	Thomas C. Meyers/Sandra Richards
Filer Authorized By:	Thomas C. Meyers
Attorney Docket Number:	MLB-002/06US 34468/40
Receipt Date:	19-NOV-2020
Filing Date:	
Time Stamp:	16:02:33
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	DA
Payment was successfully received in RAM	$3000
RAM confirmation Number	E2020AIG02510877
Deposit Account	
Authorized User	
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	TrackOne Request	MLB-002-06US-Track-One-Cert.pdf	32692 5e42dd346dced0a6b6dd8f40ab38aff3d8b6cc4a	no	1
Warnings:					
Information:					
2		MLB-002-06US-Application.pdf	13469785 eea29a2a9eb6404bca8415921b4900ff1113d9c6	yes	99

	Multipart Description/PDF files in .zip description		
	Document Description	**Start**	**End**
	Abstract	99	99
	Claims	89	98
	Specification	1	88

Warnings:					
Information:					
3	Drawings-only black and white line drawings	MLB-002-05US-Drawings.pdf	567883 ef44392ca5d1044ad72fba472a98dfbba47b92cc	no	16
Warnings:					
Information:					
4	Sequence Listing	MLB-002-04-SEQ-LIST.pdf	3148128 5813db53322097cf1cb3135f55bf79dcfe2ee7e6	no	42
Warnings:					
Information:					
5	Oath or Declaration filed	MLB-002-06US-Decs.pdf	196381 bb5de06acc564f693725d0e5ced349dd6208cf64	no	2

Warnings:					
Information:					
6	Application Data Sheet	MLB-002-06US-ADS.pdf	1256074 5d190380b61505e5f85e082bc0e542ab2a14f614	no	9
Warnings:					
Information:					
7		MLB-002-06US-Prelim-amend.pdf	26583 498f8a0b4ae1381beba8ea64775bad57b86d744d	yes	6
	Multipart Description/PDF files in .zip description				
	Document Description		**Start**	**End**	
	Applicant Arguments/Remarks Made in an Amendment		5	6	
	Claims		3	4	
	Specification		2	2	
	Preliminary Amendment		1	1	
Warnings:					
Information:					
8	Fee Worksheet (SB06)	fee-info.pdf	39981 c54680988d531fee83dcf1685be0876a3c5ad7b1	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	18737507		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Doc Code: TRACK1.REQ
Document Description: TrackOne Request

PTO/AIA/424 (04-14)

CERTIFICATION AND REQUEST FOR PRIORITIZED EXAMINATION UNDER 37 CFR 1.102(e) (Page 1 of 1)

First Named Inventor:	Gregory Porreca	Nonprovisional Application Number (if known):	Not Yet Assigned
Title of Invention:	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS		

APPLICANT HEREBY CERTIFIES THE FOLLOWING AND REQUESTS PRIORITIZED EXAMINATION FOR THE ABOVE-IDENTIFIED APPLICATION.

1. The processing fee set forth in 37 CFR 1.17(i)(1) and the prioritized examination fee set forth in 37 CFR 1.17(c) have been filed with the request. The publication fee requirement is met because that fee, set forth in 37 CFR 1.18(d), is currently $0. The basic filing fee, search fee, and examination fee are filed with the request or have been already been paid. I understand that any required excess claims fees or application size fee must be paid for the application.
2. I understand that the application may not contain, or be amended to contain, more than four independent claims, more than thirty total claims, or any multiple dependent claims, and that any request for an extension of time will cause an outstanding Track I request to be dismissed.
3. The applicable box is checked below:

I. [x] **Original Application (Track One) - Prioritized Examination under § 1.102(e)(1)**

i. (a) The application is an original nonprovisional utility application filed under 35 U.S.C. 111(a). This certification and request is being filed with the utility application via EFS-Web.

---OR---

(b) The application is an original nonprovisional plant application filed under 35 U.S.C. 111(a). This certification and request is being filed with the plant application in paper.

ii. An executed inventor's oath or declaration under 37 CFR 1.63 or 37 CFR 1.64 for each inventor, **or** the application data sheet meeting the conditions specified in 37 CFR 1.53(f)(3)(i) is filed with the application.

II. [ ] **Request for Continued Examination - Prioritized Examination under § 1.102(e)(2)**

i. A request for continued examination has been filed with, or prior to, this form.
ii. If the application is a utility application, this certification and request is being filed via EFS-Web.
iii. The application is an original nonprovisional utility application filed under 35 U.S.C. 111(a), or is a national stage entry under 35 U.S.C. 371.
iv. This certification and request is being filed prior to the mailing of a first Office action responsive to the request for continued examination.
v. No prior request for continued examination has been granted prioritized examination status under 37 CFR 1.102(e)(2).

Signature	/Thomas C. Meyers/	Date	November 19, 2020
Name (Print/Typed)	Thomas C. Meyers	Practitioner Registration Number	36,989

***Note**: This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4(d) for signature requirements and certifications. Submit multiple forms if more than one signature is required.**

[ ] *Total of 1 forms are submitted.

63911455 v1

Petitioner Illumina, Inc., EX1004, p. 119

METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS

RELATED APPLICATIONS

This application claims the benefit under 35 U.S.C. §119(e) of U.S. provisional application serial number 61/174,470, filed April 30, 2009, U.S. provisional application serial number 61/178,923, filed May 15, 2009, U.S. provisional application serial number 61/179,358, filed May 18, 2009, and U.S. provisional application serial number 61/182,089, filed May 28, 2009, the entire contents of each of which are incorporated herein by reference.

FIELD OF INVENTION

The invention relates to methods and compositions for determining genotypes in patient samples.

BACKGROUND OF THE INVENTION

Information about the genotype of a subject is becoming more important and relevant for a range of healthcare decisions as the genetic basis for many diseases, disorders, and physiological characteristics is further elucidated. Medical advice is increasingly personalized, with individual decisions and recommendations being based on specific genetic information. Information about the type and number of alleles at one or more genetic loci impacts disease risk, prognosis, therapeutic options, and genetic counseling amongst other healthcare considerations.

For cost-effective and reliable medical and reproductive counseling on a large scale, it is important to be able correctly and unambiguously identify the allelic status for many different genetic loci in many subjects.

Numerous technologies have been developed for detecting and analyzing nucleic acid sequences from biological samples. These technologies can be used to genotype subjects and determine the allelic status of any locus of interest. However, they are not sufficiently robust and cost-effective to be scaled up for reliable high throughput analysis of many genetic loci in large numbers of patients. The frequency of incorrect or ambiguous calls is too high for current technology to manage large numbers of patient

samples without involving expensive and time-consuming steps to resolve uncertainties and provide confidence in the information output.

SUMMARY OF THE INVENTION

Aspects of the invention relate to preparative and analytical methods and compositions for evaluating genotypes, and in particular, for determining the allelic identity (or identities in a diploid organism) of one or more genetic loci in a subject.

Aspects of the invention are based, in part, on the identification of different sources of ambiguity and error in genetic analyses, and, in part, on the identification of one or more approaches to avoid, reduce, recognize, and/or resolve these errors and ambiguities at different stages in a genetic analysis.

According to aspects of the invention, certain types of genetic information can be under-represented or over-represented in a genetic analysis due to a combination of stochastic variation and systematic bias in any of the preparative stages (e.g., capture, amplification, etc.), determining stages (e.g., allele-specific detection, sequencing, etc.), data interpretation stages (e.g., determining whether the assay information is sufficient to identify a subject as homozygous or heterozygous), and/or other stages.

According to aspects of the invention, error or ambiguity may be apparent in a genetic analysis, but not readily resolved without running additional samples or more expensive assays (*e.g.*, array-based assays may report no-calls due to noisy/low signal). According to further aspects of the invention, error or ambiguity may not be accounted for in a genetic analysis and incorrect base calls may be made even when the evidence for them is limited and/or not statistically significant (*e.g.*, next-generation sequencing technologies may report base calls even if the evidence for them is not statistically significant). According to further aspects of the invention error or ambiguity may be problematic for a multi-step genetic analysis because it is apparent but not readily resolved in one or more steps of the analysis and not apparent or accounted for in other steps of the analysis.

In some embodiments, sources of error and ambiguity in one or more steps can be addressed by capturing and/or interrogating each target locus of interest with one or more sets of overlapping probes that are designed to overcome any systematic bias or stochastic effects that may impact the complexity and/or fidelity of the genetic information that is generated.

Petitioner Illumina, Inc., EX1004, p. 121

In some embodiments, sources of error and ambiguity in one or more steps can be addressed by capturing and/or interrogating each target locus of interest with at least one set of probes, wherein different probes are labeled with different identifiers that can be used to track the assay reactions and determine whether certain types of genetic information are under-represented or over-represented in the information that is generated.

In some embodiments, errors and ambiguities associated with the analysis of regions containing large numbers of sequence repeats are addressed by systematically analyzing frequencies of certain nucleic acids at particular stages in an assay (e.g., at a capture, sequencing, or detection stage). It should be appreciated that such techniques may be particularly useful in the context of a standardized protocol that is designed to allow many different loci to be evaluated in parallel without requiring different assay procedures for each locus. In some embodiments, the use of a single detection modality (*e.g.*, sequencing) to assay multiple types of genetic lesions (*e.g.*, point mutations, insertions/deletions, length polymorphisms) is advantageous in the clinical setting. In some embodiments of the invention, methods are provided that facilitate the use of multiple sample preparation steps in parallel, coupled with multiple analytical processes following sequence detection. Thus, in some embodiments of the invention, an improved workflow is provided that reduces error and uncertainty when simultaneously assaying different types of genetic lesions across multiple loci in multiple patients.

In some embodiments, aspects of the invention provide methods for overcoming preparative and/or analytical bias by combining two or more techniques, each having a different bias (e.g., a known bias towards under-representation or over-representation of one or more types of sequences), and using the resulting data to determine a genetic call for a subject with greater confidence.

It should be appreciated that in some embodiments, aspects of the invention relate to multiplex diagnostic methods. In some embodiments, multiplex diagnostic methods comprise capturing a plurality of genetic loci in parallel (e.g., one or more genetic loci from Table 1). In some embodiments, the genetic loci possess one or more polymorphisms (e.g., one or more polymorphisms from Table 2) the genotypes of which correspond to disease causing alleles. Accordingly, in some embodiments, the disclosure provides methods for assessing multiple heritable disorders in parallel. In some

Petitioner Illumina, Inc., EX1004, p. 122

embodiments, methods are provided for diagnosing multiple heritable disorders in parallel at a pre-implantation, prenatal, perinatal, or postnatal stage. In some embodiments, the disclosure provides methods for analyzing multiple genetic loci (e.g., a plurality of target nucleic acids selected from Table 1) from a patient sample, such as a blood, pre-implantation embryo, chorionic villus or amniotic fluid sample, or other sample (e.g., other biological fluid or tissue sample such as a biopsy sample) as aspects of the invention are not limited in this respect.

Other samples may include tumor tissue or circulating tumor cells. In some embodiments, a patient sample (*e.g.*, a tumor tissue or cell sample) is mosaic for one or more mutations of interest, and thus, may require higher sensitivity than is needed for a germline mutation analysis. In some embodiments, a sample comprises cells from a non-host organism (*e.g.*, bacterial or viral infections in a human subject) or a sample for environmental monitoring (*e.g.*, bacterial, viral, fungal composition of a soil, water, or air sample).

Accordingly, in some embodiments, aspects of the methods disclosed herein relate to genotyping a polymorphism of a target nucleic acid. In some embodiments, the genotyping may comprise determining that one or more alleles of the target nucleic acid are heterozygous or homozygous. In further embodiments, the genotyping may comprise determining the sequence of a polymorphism and comparing that sequence to a control sequence that is indicative of a disease risk. In some embodiments, the polymorphism is selected from a locus in Table 1 or Table 2. However, it should be appreciated that any locus associated with a disease or condition of interest may be used.

In some embodiments, a diagnosis, prognosis, or disease risk assessment is provided to a subject based on a genotype determined for that subject at one or more genetic loci (e.g., based on the analysis of a biological sample obtained from that subject). In some embodiments, an assessment is provided to a couple, based on their respective genotypes at one or more genetic loci, of the risk of their having one or more children having a genotype associated with a disease or condition (e.g., a homozygous or heterozygous genotype associated with a disease or condition). In some embodiments, a subject or a couple may seek genetic or reproductive counseling in connection with a genotype determined according to embodiments of the invention. In some embodiments, genetic information from a tumor or circulating tumor cells is used to determine prognosis and guide selection of appropriate drugs/treatments.

Petitioner Illumina, Inc., EX1004, p. 123

It should be appreciated that any of the methods or compositions described herein may be used in combination with any of the medical evaluations associated with one or more genetic loci as described herein.

In some embodiments, aspects of the invention provide effective methods for overcoming challenges associated with systematic errors (bias) and/or stochastic effects in multiplex genomic capture and/or analysis (including sequencing analysis). In some embodiments, aspects of the invention are useful to avoid, reduce and/or account for variability in one or more sampling and/or analytical steps. For example, in some embodiments, variability in target nucleic acid representation and unequal sampling of heterozygous alleles in pools of captured target nucleic acids can be overcome.

Accordingly, in some embodiments, the disclosure provides methods that reduce variability in the detection of target nucleic acids in multiplex capture methods. In other embodiments, methods improve allelic representation in a capture pool and, thus, improve variant detection outcomes. In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve the use of different sets of multiple probes (e.g., molecular inversion probes MIPs) that capture overlapping regions of a target nucleic acid to achieve a more uniform representation of the target nucleic acids in a capture pool compared with methods of the prior art. In other embodiments, methods reduce bias, or the risk of bias, associated with large scale parallel capture of genetic loci, e.g., for diagnostic purposes. In other embodiments, methods are provided for increasing reproducibility (e.g., by reducing the effect of polymorphisms on target nucleic acid capture) in the detection of a plurality of genetic loci in parallel. In further embodiments, methods are provided for reducing the effect of probe synthesis and/or probe amplification variability on the analysis of a plurality of genetic loci in parallel.

According to some aspects, methods of analyzing a plurality of genetic loci are provided. In some embodiments, the methods comprise contacting each of a plurality of target nucleic acids with a probe set, wherein each probe set comprises a plurality of different probes, each probe having a central region flanked by a 5' region and a 3' region that are complementary to nucleic acids flanking the same strand of one of a plurality of subregions of the target nucleic acid, wherein the subregions of the target nucleic acid are different, and wherein each subregion overlaps with at least one other subregion, isolating a plurality of nucleic acids each having a nucleic acid sequence of a different

Petitioner Illumina, Inc., EX1004, p. 124

subregion for each of the plurality of target nucleic acids, and analyzing the isolated nucleic acids.

In other embodiments, methods comprise contacting each of a plurality of target nucleic acids with a probe set, wherein each probe set comprises a plurality of different probes, each probe having a central region flanked by a 5' region and a 3' region that are complementary to nucleic acids flanking the same strand of one of a plurality of subregions of the target nucleic acid, wherein the subregions of the target nucleic acid are different, and wherein a portion of the 5' region and a portion of the 3' region of a probe have, respectively, the sequence of the 5' region and the sequence of the 3' region of a different probe, isolating a plurality of nucleic acids each having a nucleic acid sequence of a different subregion for each of the plurality of target nucleic acids, and analyzing the isolated nucleic acids.

Aspects of the disclosure are based, in part, on the discovery of methods for overcoming problems associated with systematic and random errors (bias) in genome capture, amplification and sequencing methods, namely high variability in the capture and amplification of nucleic acids and disproportionate representation of heterozygous alleles in sequencing libraries. Accordingly, in some embodiments, the disclosure provides methods that reduce errors associated with the variability in the capture and amplification of nucleic acids. In other embodiments, the methods improve allelic representation in sequencing libraries and, thus, improve variant detection outcomes. In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve the use of differentiator tag sequences to uniquely tag individual nucleic acid molecules. In some embodiments, the differentiator tag sequence permit the detection of bias based on the occurrence of combinations of differentiator tag and target sequences observed in a sequencing reaction. In other embodiments, the methods reduce errors caused by bias, or the risk of bias, associated with the capture, amplification and sequencing of genetic loci, e.g., for diagnostic purposes.

Aspects of the invention relate to providing sequence tags (referred to as differentiator tags) that are useful to determine whether target nucleic acid sequences identified in an assay are from independently isolated target nucleic acids or from multiple copies of the same target nucleic acid molecule (e.g., due to bias in a preparative step, for example, amplification). This information can be used to help analyze a

Petitioner Illumina, Inc., EX1004, p. 125

threshold number of independently isolated target nucleic acids from a biological sample in order to obtain sequence information that is reliable and can be used to make a genotype conclusion (e.g., call) with a desired degree of confidence. This information also can be used to detect bias in one or more nucleic acid preparative steps.

In some embodiments, the methods disclosed herein are useful for any application where reduction of bias, e.g., associated with genomic isolation, amplification, sequencing, is important. For example, detection of cancer mutations in a heterogeneous tissue sample, detection of mutations in maternally-circulating fetal DNA, and detection of mutations in cells isolated during a preimplantation genetic diagnostic procedure.

Accordingly, in some aspects, methods of genotyping a subject are provided. In some embodiments, the methods comprise determining the sequence of at least a threshold number of independently isolated nucleic acids, wherein the sequence of each isolated nucleic acid comprises a target nucleic acid sequence and a differentiator tag sequence, wherein the threshold number is a number of unique combinations of target nucleic acid and differentiator tag sequences, wherein the isolated nucleic acids are identified as independently isolated if they comprise unique combinations of target nucleic acid and differentiator tag sequences, and wherein the target nucleic acid sequence is the sequence of a genomic locus of a subject.

In some embodiments, the isolated nucleic acids are products of a circularization selection-based preparative method, e.g., molecular inversion probe capture products. In other embodiments, the isolated nucleic acids are products of an amplification-based preparative methods. In other embodiments, the isolated nucleic acids are products of hybridization-based preparative methods.

Circularization selection-based preparative methods selectively convert regions of interest (target nucleic acids) into a covalently-closed circular molecule which is then isolated typically by removal (usually enzymatic, e.g. with exonuclease) of any non-circularized linear nucleic acid. Oligonucleotide probes (e.g., molecular inversion probes) are designed which have ends that flank the region of interest (target nucleic acid) and, optionally, primer sites, e.g., sequencing primer sites. The probes are allowed to hybridize to the genomic target, and enzymes are used to first (optionally) fill in any gap between probe ends and second ligate the probe closed. Following circularization, any remaining (non-target) linear nucleic acid is typically removed, resulting in isolation

Petitioner Illumina, Inc., EX1004, p. 126

(capture) of target nucleic acid. Circularization selection-based preparative methods include molecular inversion probe capture reactions and 'selector' capture reactions. In some embodiments, molecular inversion probe capture of a target nucleic acid is indicative of the presence of a polymorphism in the target nucleic acid.

In amplification-based (e.g., PCR-based or LCR-based, etc.) preparative methods, genomic loci (target nucleic acids) are isolated directly by means of a polymerase chain reaction or ligase chain reaction (or other amplification method) that selectively amplifies each locus using one or more oligonucleotide primers. It is to be understood that primers will be sufficiently complementary to the target sequence to hybridize with and prime amplification of the target nucleic acid. Any one of a variety of art known methods may be utilized for primer design and synthesis. One or more of the primers may be perfectly complementary to the target sequence. Degenerate primers may also be used. Primers may also include additional nucleic acids that are not complementary to target sequences but that facilitate downstream applications, including for example restriction sites and differentiator tag sequences. Amplification-based methods include amplification of a single target nucleic acid and multiplex amplification (amplification of multiple target nucleic acids in parallel).

Hybridization-based preparative methods involve selectively immobilizing target nucleic acids for further manipulation. It is to be understood that one or more oligonucleotides (immobilization oligonucleotides), which comprise differentiator tag sequences, and which may be from 15 to 170 nucleotides in length, are used which hybridize along the length of a target region of a genetic locus to immobilize it. In some embodiments, immobilization oligonucleotides, are either immobilized before hybridization is performed (e.g., Roche/Nimblegen 'sequence capture'), or are prepared such that they include a moiety (e.g. biotin) which can be used to selectively immobilize the target nucleic acid after hybridization by binding to e.g., streptavidin-coated microbeads (e.g. Agilent 'SureSelect').

It should be appreciated that any of the circularization, amplification, and/or hybridization based methods described herein may be used in connection with one or more of the tiling/staggering, tagging, size-detection, and/or sensitivity enhancing algorithms described herein.

In some embodiments, the methods disclosed herein comprise determining the sequence of molecular inversion probe capture products, each comprising a molecular

Petitioner Illumina, Inc., EX1004, p. 127

inversion probe and a target nucleic acid, wherein the sequence of the molecular inversion probe comprises a differentiator tag sequence and, optionally, a primer sequence, and wherein the target nucleic acid is a captured genomic locus of a subject, and genotyping the subject at the captured genomic locus based on the sequence of at least a threshold number of unique combinations of target nucleic acid and differentiator tag sequences of molecular inversion probe capture products.

In some embodiments, the methods disclosed herein comprise obtaining molecular inversion probe capture products, each comprising a molecular inversion probe and a target nucleic acid, wherein the sequence of the molecular inversion probe comprises a differentiator tag sequence and, optionally, a primer sequence, wherein the target nucleic acid is a captured genomic locus of the subject, amplifying the molecular inversion probe capture products, and genotyping the subject by determining, for each target nucleic acid, the sequence of at least a threshold number of unique combinations of target nucleic acid and differentiator tag sequence of molecular inversion probe capture products. In certain embodiments, obtaining comprises capturing target nucleic acids from a genomic sample of the subject with molecular inversion probes, each comprising a unique differentiator tag sequence. In specific embodiments, capturing is performed under conditions wherein the likelihood of obtaining two or more molecular inversion probe capture products with identical combinations of target and differentiator tag sequences is equal to or less than a predetermined value, optionally wherein the predetermined value is about 0.05.

In one embodiment, the threshold number for a specific target nucleic acid sequence is selected based on a desired statistical confidence for the genotype. In some embodiments, the methods further comprising determining a statistical confidence for the genotype based on the number of unique combinations of target nucleic acid and differentiator tag sequences.

According to some aspects, methods of analyzing a plurality of genetic loci are provided. In some embodiments, the methods comprise obtaining a plurality of molecular inversion probe capture products each comprising a molecular inversion probe and a target nucleic acid, wherein the sequence of the molecular inversion probe comprises a differentiator tag sequence and, optionally, a primer sequence (e.g., a sequence that is complementary to the sequence of a nucleic acid that is used as a primer for sequencing or other extension reaction), amplifying the plurality of molecular

Petitioner Illumina, Inc., EX1004, p. 128

inversion probe capture products, determining numbers of occurrence of combinations of target nucleic acid and differentiator tag sequence of molecular inversion probe capture products in the amplified plurality, and if the number of occurrence of a specific combination of target nucleic acid sequence and differentiator tag sequence exceeds a predetermined value, detecting bias in the amplification of the molecular inversion probe comprising the specific combination. In some embodiments, the methods further comprise genotyping target sequences in the plurality, wherein the genotyping comprises correcting for bias, if detected.

In some embodiments, the target nucleic acid is a gene (or portion thereof) selected from Table 1. In some embodiments, the genotyping comprises determining the sequence of a target nucleic acid (e.g., a polymorphic sequence) at one or more (both) alleles of a genome (a diploid genome) of a subject. In certain embodiments, the genotyping comprises determining the sequence of a target nucleic acid at both alleles of a diploid genome of a subject, wherein in the target nucleic acid comprises, or consists of, a sequence of Table 1, Table 2, or other locus of interest.

In some embodiments, aspects of the invention provide methods and compositions for identifying nucleic acid insertions or deletions in genomic regions of interest without determining the nucleotide sequences of these regions. Aspects of the invention are particularly useful for detecting nucleic acid insertions or deletions in genomic regions containing nucleic acid sequence repeats (e.g., di- or tri-nucleotide repeats). However, the invention is not limited to analyzing nucleic acid repeats and may be used to detect insertions or deletions in any target nucleic acid of interest. Aspects of the invention are particularly useful for analyzing multiple loci in a multiplex assay.

In some embodiments, aspects of the invention relate to determining whether an amount of target nucleic acid that is captured in a genomic capture assay is higher or lower than expected. In some embodiments, a statistically significant deviation from an expected amount (e.g., higher or lower) is indicative of the presence of a nucleic acid insertion or deletion in the genomic region of interest. In some embodiments, the amount is a number of nucleic acid molecules that are captured. In some embodiments, the amount is a number of independently captured nucleic acid molecules in a sample. It should be appreciated that the captured nucleic acids may be literally captured from a sample, or their sequences may be captured without actually capturing the original nucleic acids in the sample. For example, nucleic acid sequences may be captured in an

Petitioner Illumina, Inc., EX1004, p. 129

assay that involves a template-based extension of nucleic acids having the region of interest, in the sample.

Aspects of the invention are based on the recognition that the efficiency of certain capture techniques is affected by the length of the nucleic acid being captured. Accordingly, an increase or decrease in the length of a target nucleic acid (e.g., due to an insertion or deletion of a repeated sequence) can alter the capture efficiency of that nucleic acid. In some embodiments, a difference in the capture efficiency (e.g., a statistically significant difference in the capture efficiency) of a target nucleic acid is indicative of an insertion or deletion in the target nucleic acid. It should be appreciated that the capture efficiency for a target nucleic acid may be evaluated based on an amount of captured nucleic acid (e.g., number of captured nucleic acid molecules) relative to a control amount (e.g., based on an amount of control nucleic acid that is captured). However, the invention is not limited in this respect and other techniques for evaluating capture efficiency also may be used.

According to aspects of the invention, evaluating the capture efficiency as opposed to determining the sequence of the entire repeat region reduces errors associated with sequencing through repeat regions. Repeat sequences often give rise to stutters or skips in sequencing reactions that make it very difficult to accurately determine the number of repeats in a target region without running multiple sequencing reactions under different conditions and carefully analyzing the results. Such procedures are cumbersome and not readily scalable in a manner that is consistent with high throughput analyses of target nucleic acids. In some embodiments, repeat regions may be longer than the length of the individual sequence read, making length determination on the basis of a single read impossible. For example, when using next-generation sequencing the repeat regions may be longer than the length of the individual sequence read, making length determination on the basis of a single read impossible. Accordingly, aspects of the invention are useful to increase the sensitivity of detecting insertions or deletions in target regions, particularly target regions containing repeated sequences.

In some embodiments, aspects of the invention relate to capturing genomic nucleic acid sequences using a molecular inversion probe (e.g., MIP or Padlock probe) technique, and determining whether the amount (e.g., number) of captured sequences is higher or lower than expected. In some embodiments, the amount (e.g., number) of captured sequences is compared to an amount (e.g., number) of sequences captured in a

Petitioner Illumina, Inc., EX1004, p. 130

control assay. The control assay may involve analyzing a control sample that contains a nucleic acid from the same genetic locus having a known sequence length (e.g., a known number of nucleic acid repeats). However, a control may involve analyzing a second (e.g., different) genetic locus that is not expected to contain any insertions or deletions. The second genetic locus may be analyzed in the same sample as the locus being interrogated or in a different sample where its length has been previously determined. The second genetic locus may be a locus that is not characterized by the presence of nucleic acid repeats (and thus not expected to contain insertions or deletions of the repeat sequence).

In some embodiments, a target nucleic acid region that is being evaluated may be determined by the identity of the targeting arms of a probe that is designed to capture the target region (or sequence thereof). For example, the targeting arms of a MIP probe may be designed to be complementary (e.g., sufficiently complementary for selective hybridization and//or polymerase extension and/or ligation) to genomic regions flanking a target region suspected of containing an insertion or deletion. It should be appreciated that two targeting arms may be designed to be complementary (e.g., sufficiently complementary for selective hybridization and/or polymerase extension and/or ligation) to the two flanking regions that are immediately adjacent (e.g., immediately 5' and 3', respectively) to a region of a sequence repeat on one strand of a genomic nucleic acid. However, one or both targeting arms may be designed to hybridize several bases (e.g., 1-5, 5-10, 10-25, 25-50, or more) upstream or downstream from the repeat region in such a way that the captured sequence includes a region of unique genomic sequence that on one or both sides of the repeat region. This unique region can then be used to identify the captured target (e.g., based on sequence or hybridization information).

In some embodiments, two or more (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10 or more) different loci may be interrogated in parallel in a single assay (e.g., in a multiplex assay). In some embodiments, the ratio of captured nucleic acids for each locus may be used to determine whether a nucleic acid insertion or deletion is present in one locus relative to the other. For example, the ratio may be compared to a control ratio that is representative of the two loci when neither one has an insertion or deletion relative to control sequences (e.g., sequences that are normal or known to be associated with healthy phenotypes for those loci). However, the amount of captured nucleic acids may be compared to any suitable control as discussed herein.

Petitioner Illumina, Inc., EX1004, p. 131

The locus of a captured sequence may be identified by determining a portion of unique sequence 5' and/or 3' to the repeat region in the target nucleic acid suspected of containing a deletion or insertion. This does not require sequencing the captured repeat region itself. However, some or all of the repeat region also could be sequenced as aspects of the invention are not limited in this respect.

Aspects of the invention may be combined with one or more sequence-based assays (e.g., SNP detection assays), for example in a multiplex format, to determine the genotype of one or more regions of a subject.

In some embodiments, methods of detecting a polymorphism in a nucleic acid in a biological sample are provided. In some embodiments, the methods comprise evaluating the efficiency of capture at one or more loci and determining whether one or both alleles at that locus contain an insertion or deletion relative to a control locus (e.g., a locus indicative of a length of repeat sequence that is associated with a healthy phenotype).

Accordingly, aspects of the invention relate to methods for determining whether a target nucleic acid has an abnormal length by evaluating the capture efficiency of a target nucleic acid in a biological sample from a subject, wherein a capture efficiency that is different from a reference capture efficiency is indicative of the presence, in the biological sample, of a target nucleic acid having an abnormal length. It should be appreciated that the term "abnormal" is a relative term based on a comparison to a "normal" length. In some embodiments, a normal length is a length that is associated with a normal (e.g., healthy or non-carrier phenotype). Accordingly, an abnormal length is a length that is either shorter or longer than the normal length. In some embodiments, the presence of an abnormal length is indicative of an increased risk that the locus is associated with a disease or a disease carrier phenotype. In some embodiments, the abnormal length is indicative that the subject is either has a disease or condition or is a carrier of a disease or condition (e.g., associated with the locus). However, it should be appreciated that the description of embodiments relating to detecting the presence of an abnormal length also support detecting the presence of a length that is different from an expected or control length.

In some embodiments, aspects of the invention relate to estimating the length of a target nucleic acid (e.g., of a sub-target region within a target nucleic acid). In some embodiments, aspects of the invention relate to methods for estimating the length of a

Petitioner Illumina, Inc., EX1004, p. 132

target nucleic acid by contacting the target nucleic acid with a plurality of detection probes under conditions that permit hybridization of the detection probes to the target nucleic acid, wherein each detection probe is a polynucleotide that comprises a first arm that hybridizes to a first region of the target nucleic acid and a second arm that hybridizes to a second region of the target nucleic acid, wherein the first and second regions are on a common strand of the target nucleic acid, and wherein the nucleotide sequence of the target between the 5' end of the first region and the 3' end of the second region is the nucleotide sequence of a sub-target nucleic acid; and capturing a plurality of sub-target nucleic acids that are hybridized with the plurality of detection probes; and measuring the frequency of occurrence of a sub-target nucleic acid in the plurality of sub-target nucleic acids, wherein the frequency of occurrence of the sub-target nucleic acid in the plurality of sub-target nucleic acids is indicative of the length of the sub-target nucleic acid. It should be appreciated that methods for estimating a nucleic acid length may involve comparing a capture efficiency for a target nucleic acid region to two or more reference efficiencies for known nucleic acid lengths in order to determine whether the target nucleic acid region is smaller, intermediate, or larger in size than the known control lengths. In some embodiments, a series of nucleic acids of known different lengths may be used to provide a calibration curve for evaluating the length of a target nucleic acid region of interest.

In some embodiments, the capture efficiency of a target region suspected of having a deletion or insertion is determined by comparing the capture efficiency to a reference indicative of a normal capture efficiency. In some embodiments, the capture efficiency is lower than the reference capture efficiency. In some embodiments, the subject is identified as having an insertion in the target region. In some embodiments, the capture efficiency is higher than the reference capture efficiency. In some embodiments, the subject is identified as having a deletion in the target region. In some embodiments, the subject is identified as being heterozygous for the insertion. In some embodiments, the subject is identified as being heterozygous for the deletion.

In some embodiments of any of the methods described herein (e.g., tiling/staggering, tagging, size-detection, and/or sensitivity enhancement) aspects of the invention relate to capturing a sub-target nucleic acid (or a sequence of a sub-target nucleic acid). In some embodiments, a molecular inversion probe technique is used. In some embodiments, a molecular inversion probe is a single linear strand of nucleic acid

Petitioner Illumina, Inc., EX1004, p. 133

that comprises a first targeting arm at its 5' end and a second targeting arm at its 3' end, wherein the first targeting arm is capable of specifically hybridizing to a first region flanking one end of the sub-target nucleic acid, and wherein the second targeting arm is capable of specifically hybridizing to a second region flanking the other end of the sub-target nucleic acid on the same strand of the target nucleic acid. In some embodiments, the first and second targeting arms are between about 10 and about 100 nucleotides long. In some embodiments, the first and second targeting arms are about 10-20, 20-30, 30-40, or 40-50 nucleotides long. In some embodiments, the first and second targeting arms are about 20 nucleotides long. In some embodiments, the first and second targeting arms have the same length. In some embodiments, the first and second targeting arms have different lengths. In some embodiments, each pair of first and second targeting arms in a set of probes has the same length. Accordingly, if one of the targeting arms is longer, the other one is correspondingly shorter. This allows for a quality control step in some embodiments to confirm that all captured probe/target sequence products have the same length after a multiplexed plurality of capture reactions. In some embodiments, a set of probes may be designed to have the same length if the intervening region is varied to accommodate any differences in the length of either one or both of the first and second targeting arms.

In some embodiments, the hybridization Tms of the first and second targeting arms are similar. In some embodiments, the hybridization Tms of the first and second targeting arms are within 2-5 °C of each other. In some embodiments, the hybridization Tms of the first and second targeting arms are identical. In some embodiments, the hybridization Tms of the first and second targeting arms are close to empirically-determined optima but not necessarily identical.

In some embodiments, the first and second targeting arms of a molecular inversion probe have different Tms. For example, the Tm of the first targeting arm (at the 5' end of the molecular inversion probe) may be higher than the Tm of the second targeting arm (at the 3' end of the molecular inversion probe). According to aspects of the invention, and without wishing to be bound by theory, a relatively high Tm for the first targeting arm may help avoid or prevent the first targeting arm from being displaced after hybridization by the extension product of the 3' end of the second targeting arm. It should be appreciated that a reference to the Tm of a targeting arm as used herein relates to the Tm of hybridization of the targeting arm to a nucleic acid having the

Petitioner Illumina, Inc., EX1004, p. 134

complementary sequence (e.g., the region of the target nucleic acid that has a sequence that is complementary to the sequence of the targeting arm). It also should be appreciated that the Tms of the targeting arms described herein may be calculated using any appropriate method. For example, in some embodiments an experimental method (e.g., a gel shift assay, a hybridization assay, a melting curve analysis, for example in a PCR machine with a SYBR dye by stepping through a temperature ramp while monitoring signal level from an intercalating dye, for example, bound to a double-stranded DNA, etc.) may be used to determine one or more Tms empirically. In some embodiments, an optimal Tm may be determined by evaluating the number of products formed (e.g., for each of a plurality of MIP probes), and determining the optimal Tm as the center point in a histogram of Tm for all targeting arms. In some embodiments, a predictive algorithm may be used to determine a Tm theoretically. In some embodiments, a relatively simple predictive algorithm may be used based on the number of G/C and A/T base pairs when the sequence is hybridized to its target and/or the length of the hybridized product (e.g., for example, 64.9 + 41 * ([G+C] - 16.4) / (A + T + G + C), see for example, Wallace,R.B., Shaffer,J., Murphy,R.F., Bonner,J., Hirose,T., and Itakura,K. (1979) Nucleic Acids Res 6:3543-3557). In some embodiments, a more complex algorithm may be used to account for the effects of base stacking entropy and enthalpy, ion concentration, and primer concentration (see, for example, SantaLucia J (1998), Proc Natl Acad Sci USA, 95:1460-5). In some embodiments an algorithm may use modified parameters (e.g., nearest-neighbor parameters for basepair entropy/enthalpy values). It should be appreciated that any suitable algorithm may be used as aspects of the invention are not limited in this respect. However, it also should be appreciated that different methodologies may results in different calculated or predicted Tms for the same sequences. Accordingly, in some embodiments, the same empirical and/or theoretical method is used to determine the Tms of different sequences for a set of probes to avoid a negative impact of any systematic difference in the Tm determination or prediction when designing a set of probes with predetermined similarities or differences for different Tms.

In some embodiments, the Tm of the first targeting arm may be about 1 °C, about 2 °C, about 3 °C, about 4 °C, about 5 °C, or more than about 5 °C higher than the Tm of the second targeting arm. In some embodiments, each probe in a plurality of probes (e.g., each probe in a set of 5-10, each probe in a set of at least 10, each probe in a set of

Petitioner Illumina, Inc., EX1004, p. 135

10-50, each probe in a set of 50-100, each probe in a set of 100-500, each probe in a set of 500-1,000, each probe in a set of 1,000-1,500, each probe in a set of 1,500-2,000, each probe in a set of 2,000-3,000, 3,000-5,000, 5,000-10,000 or each probe in a set of at least 5,000 different probes) has a unique first targeting arm (e.g., they all have different sequences) and a unique second targeting arm (e.g., they all have different sequences). In some embodiments, for at least 10% of the probes (e.g., at least 25%, 25%-50%, 50%-75%, 75%-90%, 90%-95% or over 95%, or all of the probes) the first targeting arm has a Tm for its complementary sequence that is higher (e.g., about 1 °C, about 2 °C, about 3 °C, about 4 °C, about 5 °C, or more than about 5 °C higher) than the Tm of the second targeting arm for its complementary sequence. In some embodiments, each of the first targeting arms have similar or identical Tms for their respective complementary sequences and each of the second targeting arms have similar or identical Tms for their respective complementary sequences (and the first targeting arms have higher Tms than the second targeting arms). For example, in some embodiments, the Tm of the first arm(s) may be about 58 °C and the Tm of the second arm(s) may be about 56 °C. In some embodiments, the Tm of the first arm(s) may be about 68 °C, and the Tm of the second arm(s) may be about 65 °C. It should be appreciated that in some embodiments the similarity (e.g., within a range of 1 °C, 2 °C, 3 °C, 4 °C, 5 °C) or identity of the Tms for the different targeting arms should be based either on empirical data for each arm or based on the same predictive algorithm for each arm (e.g., Wallace,R.B., Shaffer,J., Murphy,R.F., Bonner,J., Hirose,T., and Itakura,K. (1979) Nucleic Acids Res 6:3543-3557, SantaLucia J (1998), Proc Natl Acad Sci USA, 95:1460-5, or other algorithm).

In some embodiments, the Tm of the first targeting arm of a molecular inversion probe (at the 5' end of the molecular inversion probe) is selected to be sufficiently stable to prevent displacement of the first targeting arm from its complementary sequence on a target nucleic acid. In some embodiments, the Tm of the first targeting arm is 50-55 °C, at least 55 °C, 55-60 °C, at least 60 °C, 60-65 °C, at least 65 °C, at least 70 °C, at least 75 °C, or at least 80 °C. As discussed above, it should be appreciated that the for a particular targeting arm may be determined empirically or theoretically. Different theoretical models may be used to determine a Tm and it should be appreciated that the predicted Tm for a particular sequence may be different depending on the algorithm used for the prediction. In some embodiments, each probe in a plurality of probes (e.g., each probe in a set of 5-10, each probe in a set of at least 10, each probe in a set of 10-50, each

Petitioner Illumina, Inc., EX1004, p. 136

probe in a set of 50-100, each probe in a set of 100-500, or each probe in a set of at least 500 different probes) has a different first targeting arm (e.g., different sequences) but each different first targeting arm has a similar or identical Tm for its complementary sequence on a target nucleic acid. It should be appreciated that in some embodiments the similarity (e.g., within a range of 1 C, 2 C, 3 C, 4 C, 5 C) or identity of the Tms for the different targeting arms should be based either on empirical data for each arm or based on the same predictive algorithm for each arm (e.g., Wallace,R.B., Shaffer,J., Murphy,R.F., Bonner,J., Hirose,T., and Itakura,K. (1979) Nucleic Acids Res 6:3543-3557, SantaLucia J (1998), Proc Natl Acad Sci USA, 95:1460-5, or other algorithm).

In some embodiments, the sub-target nucleic acid contains a nucleic acid repeat. In some embodiments, the nucleic acid repeat is a dinucleotide or trinucleotide repeat. In some embodiments, the sub-target nucleic acid contains 10-100 copies of the nucleic acid repeat in the absence of an abnormal increase or decrease in nucleic acid repeats. In some embodiments, the sub-target nucleic acid is a region of the Fragile-X locus that contains a nucleic acid repeat. In some embodiments, one or both targeting arms hybridize to a region on the target nucleic acid that is immediately adjacent to a region of nucleic acid repeats. In some embodiments, one or both targeting arms hybridize to a region on the target nucleic acid that is separated from a region of nucleic acid repeats by a region that does not contain any nucleic acid repeats. In some embodiments, the molecular inversion probe further comprises a primer-binding region that can be used to sequence the captured sub-target nucleic acid and optionally the first and/or second targeting arm.

In some embodiments, aspects of the invention relate to evaluating the length of a plurality of different target nucleic acids in a biological sample. In some embodiments, the plurality of target nucleic acids are analyzed using a plurality of different molecular inversion probes. In some embodiments, each different molecular inversion probe comprises a different pair of first and second targeting arms at each of the 3' and 5' ends. In some embodiments, each different molecular inversion probe comprises the same primer-binding sequence.

In some embodiments, aspects of the invention relate to analyzing nucleic acid from a biological sample obtained from a subject. In some embodiments, the biological sample is a blood sample. In some embodiments, the biological sample is a tissue sample, specific cell population, tumor sample, circulating tumor cells, or environmental

Petitioner Illumina, Inc., EX1004, p. 137

sample. In some embodiments, the biological sample is a single cell. In some embodiments, nucleic acids are analyzed in biological samples obtained from a plurality of different subjects. In some embodiments, nucleic acids from a biological sample are analyzed in multiplex reactions. It should be appreciated that a biological sample contains a plurality of copies of a genome derived from a plurality of cells in the sample. Accordingly, a sample may contain a plurality of independent copies of a target nucleic acid region of interest, the capture efficiency of which can be used to evaluate its size as described herein.

In some embodiments, aspects of the invention relate to evaluating a nucleic acid capture efficiency by determining an amount of target nucleic acid that is captured (e.g., an amount of sub-target nucleic acid sequences that are captured). In some embodiments, the amount of target nucleic acid that is captured is determined by determining a number of independently captured target nucleic acid molecules (e.g., the amount of independently captured molecules that have the sequence of the sub-target region). In some embodiments, the amount of target nucleic acid that is captured is compared to a reference amount of captured nucleic acid. In some embodiments, the reference amount is determined by determining a number of independently captured molecules of a reference nucleic acid. In some embodiments, the reference nucleic acid is a nucleic acid of a different locus in the biological sample that is not suspected of containing a deletion or insertion. In some embodiments, the reference nucleic acid is a nucleic acid of known size and amount that is added to the capture reaction. As described herein, a number of independently captured nucleic acid sequences can be determined by contacting a nucleic acid sample with a preparation of a probe (e.g., a MIP probe as described herein). It should be appreciated that the preparation may comprise a plurality of copies of the same probe and accordingly a plurality of independent copies of the target region may be captured by different probe molecules. The number of probe molecules that actually capture a sequence can be evaluated by determining an amount or number of captured molecules using any suitable technique. This number is a reflection of both the number of target molecules in the sample and the efficiency of capture of those target molecules, which in turn is related to the size of the target molecules as described herein. Accordingly, the capture efficiency can be evaluated by controlling for the abundance of the target nucleic acid, for example by comparing the number or amount of captured target molecules to an appropriate control (e.g., a known size and

Petitioner Illumina, Inc., EX1004, p. 138

amount of control nucleic acid, or a different locus that should be present in the same amount in the biological sample and is not expected to contain any insertions or deletions). It should be appreciated that other factors may affect the capture efficiency of a particular target nucleic acid region (e.g., the sequence of the region, the GC content, the presence of secondary structures, etc.). However, these factors also can be accounted for by using appropriate controls (e.g., known sequences having similar properties, the same sequences, other genomic sequences expected to be present in the biological sample at the same frequency, etc., or any combination thereof).

In some embodiments, aspects of the invention relate to identifying a subject as having an insertion or deletion in one or more alleles of a genetic locus if the capture efficiency for that genetic locus is statistically significantly different than a reference capture efficiency.

It should be appreciated that hybridization conditions used for any of the capture techniques described herein (e.g., MIP capture techniques) can be based on known hybridization buffers and conditions.

In some embodiments, the methods disclosed herein are useful for any application where the detection of deletions or insertions is important.

In some embodiments, aspects of the invention relate to basing a nucleic acid sequence analysis on results from two or more different nucleic acid preparatory techniques that have different systematic biases in the types of nucleic acids that they sample. According to the invention, different techniques have different sequence biases that are systematic and not simply due to stochastic effects during nucleic acid capture or amplification. Accordingly, the degree of oversampling required to overcome variations in nucleic acid preparation needs to be sufficient to overcome the biases (e.g., an oversampling of 2-5 fold, 5-10 fold, 5-15 fold, 15-20 fold, 20-30 fold, 30-50 fold, or intermediate to higher fold).

According to some embodiments, different techniques have different characteristic or systematic biases. For example, one technique may bias a sample analysis towards one particular allele at a genetic locus of interest, whereas a different technique would bias the sample analysis towards a different allele at the same locus. Accordingly, the same sample may be identified as being different depending on the type of technique that is used to prepare nucleic acid for sequence analysis. This effectively

Petitioner Illumina, Inc., EX1004, p. 139

represents a sensitivity limitation, because each technique has different relative sensitivities for polymorphic sequences of interest.

According to aspects of the invention, the sensitivity of a nucleic acid analysis can be increased by combining the sequences from different nucleic acid preparative steps and using the combined sequence information for a diagnostic assay (e.g., for a making a call as to whether a subject is homozygous or heterozygous at a genetic locus of interest).

In some embodiments, the invention provides a method of increasing the sensitivity of a nucleic acid detection assay by obtaining a first preparation of a target nucleic acid using a first preparative method on a biological sample, obtaining a second preparation of a target nucleic acid using a second preparative method on the biological sample, assaying the sequences obtained in both first and second nucleic acid preparations, and using the sequence information from both first and second nucleic acid preparations to determine the genotype of the target nucleic acid in the biological sample, wherein the first and second preparative methods have different systematic sequence biases. In some embodiments, the first and second nucleic acid preparations are combined prior to performing a sequence assay. In some embodiments, separate sequence assays are performed on the first and second nucleic acid preparations and the sequence information from both assays are combined to determine the genotype of the target nucleic acid in the biological sample. In some embodiments, the first preparative method is an amplification-based, a hybridization-based, or a circular probe-based preparative method. In some embodiments, the second method is an amplification-based, a hybridization-based, or a circular probe-based preparative method. In some embodiments, the first and second methods are of different types (e.g., only one of them is an amplification-based, a hybridization-based, or a circular probe-based preparative method, and the other one is one or the other two types of method). Accordingly, in some embodiments the second preparative method is an amplification-based, a hybridization-based, or a circular probe-based preparative method, provided that the second method is different from the first method. However, in some embodiments, both methods may be of the same type, provided they are different methods (e.g., both are amplification based or hybridization-based, but are different types of amplification or hybridization methods, e.g., with different relative biases).

Petitioner Illumina, Inc., EX1004, p. 140

In amplification-based (e.g., PCR-based or LCR-based, etc.) preparative methods, genomic loci (target nucleic acids) are isolated directly by means of a polymerase chain reaction or ligase chain reaction (or other amplification method) that selectively amplifies each locus using a pair of oligonucleotide primers. It is to be understood that primers will be sufficiently complementary to the target sequence to hybridize with and prime amplification of the target nucleic acid. Any one of a variety of art known methods may be utilized for primer design and synthesis. One or both of the primers may be perfectly complementary to the target sequence. Degenerate primers may also be used. Primers may also include additional nucleic acids that are not complementary to target sequences but that facilitate downstream applications, including for example restriction sites and identifier sequences (e.g., source sequences). PCR based methods may include amplification of a single target nucleic acid and multiplex amplification (amplification of multiple target nucleic acids in parallel).

Hybridization-based preparative may methods involve selectively immobilizing target nucleic acids for further manipulation. It is to be understood that one or more oligonucleotides (immobilization oligonucleotides), which in some embodiments may be from 10 to 200 nucleotides in length, are used which hybridize along the length of a target region of a genetic locus to immobilize it. In some embodiments, immobilization oligonucleotides are either immobilized before hybridization is performed (e.g., Roche/Nimblegen 'sequence capture'), or are prepared such that they include a moiety (e.g., biotin) which can be used to selectively immobilize the target nucleic acid after hybridization by binding to e.g., streptavidin-coated microbeads (e.g., Agilent 'SureSelect').

Circularization selection-based preparative methods selectively convert each region of interest into a covalently-closed circular molecule which is then isolated by removal (usually enzymatic, e.g., with exonuclease) of any non-circularized linear nucleic acid. Oligonucleotide probes are designed which have ends that flank the region of interest. The probes are allowed to hybridize to the genomic target, and enzymes are used to first (optionally) fill in any gap between probe ends and second ligate the probe closed. In some embodiments, following circularization, any remaining (non-target) linear nucleic acid can be removed, resulting in isolation (capture) of target nucleic acid. Circularization selection-based preparative methods include molecular inversion probe capture reactions and 'selector' capture reactions. However, other techniques may be

Petitioner Illumina, Inc., EX1004, p. 141

used as aspects of the invention are not limited in this respect. In some embodiments, molecular inversion probe capture of a target nucleic acid is indicative of the presence of a polymorphism in the target nucleic acid.

A variety of methods may be used to evaluate and compare bias profiles of each preparative technique. Next-generation sequencing may be used to quantitatively measure the abundance of each isolated target nucleic acid obtained from a certain preparative method. This abundance may be compared to a control abundance value (e.g., a known starting abundance of the target nucleic acid) and/or with an abundance determined through the use of an alternative preparative method. For example, a set of target nucleic acids may be isolated by one or more of the three preparative methods; the target nucleic acid may be observed x times using the amplification technique, y times using the hybridization enrichment technique, and z times using the circularization selection technique. A pairwise correlation coefficient may be computed between each abundance value (e.g., x and y, x and z, and y and z) to assess bias in nucleic acid isolation between pairs of preparative methods. Since the mechanisms of isolation are different in each approach, the abundances will usually be different and largely uncorrelated with each other.

In some embodiments, the invention provides a method of obtaining a nucleic acid preparation that is representative of a target nucleic acid in a biological sample by obtaining a first preparation of a target nucleic acid using a first preparative method on a biological sample, obtaining a second preparation of a target nucleic acid using a second preparative method on the biological sample, and combining the first and second nucleic acid preparations to obtain a combined preparation that is representative of the target nucleic acid in the biological sample.

In some embodiments of any of the methods described herein, a third preparation of the target nucleic acid is obtained using a third preparative method that is different from the first and second preparative methods, wherein the first, second, and third preparative methods all have different systematic sequence biases. In some embodiments of any of the methods described herein, the different preparative methods are used for a plurality of different loci in the biological sample to increase the sensitivity of a multiplex nucleic acid analysis. In some embodiments, the target nucleic acid has a sequence of a gene selected from Table 1.

Petitioner Illumina, Inc., EX1004, p. 142

However, it should be appreciated that a genotyping method of the invention may include several steps, each of which independently may involve one or more different preparative techniques described herein. In some embodiments, a nucleic acid preparation may be obtained using one or more (e.g., 2, 3, 4, 5, or more) different techniques described herein (e.g., amplification, hybridization capture, circular probe capture, etc., or any combination thereof) and the nucleic acid preparation may be analyzed using one or more different techniques (e.g., amplification, hybridization capture, circular probe capture, etc., or any combination thereof) that are selected independently of the techniques used for the initial preparation.

In some embodiments, aspects of the invention also provide compositions, kits, devices, and analytical methods for increasing the sensitivity of nucleic acid assays. Aspects of the invention are particularly useful for increasing the confidence level of genotyping analyses. However, aspects of the invention may be used in the context of any suitable nucleic acid analysis, for example, but not limited to, a nucleic acid analysis that is designed to determine whether more than one sequence variant is present in a sample.

In some embodiments, aspects of the invention relate to a plurality of nucleic acid probes (e.g., 10-50, 50-100, 100-250, 250-500, 500-1,000, 1,000-2,000, 2,000-5,000, 5,000-7,500, 7,500-10,000, or lower, higher, or intermediate number of different probes). In some embodiments, each probe or each of a subset of probes (e.g., 10-25%, 25-50%, 50-75%, 75-90%, or 90-99%) has a different first targeting arm. In some embodiments, each probe or each probe of a subset of probes (e.g., 10-25%, 25-50%, 50-75%, 75-90%, or 90-99%) has a different second targeting arm. In some embodiments, the first and second targeting arms are separated by the same intervening sequence. In some embodiments, the first and second targeting arms are complementary to target nucleic acid sequences that are separated by the same or a similar length (e.g., number of nucleic acids, for example, 0-25, 25-50, 50-100, 100-250, 250-500, 500-1,000, 1,000-2,500 or longer or intermediate number of nucleotides) on their respective target nucleic acids (e.g., genomic loci). In some embodiments, each probe or a subset of probes (e.g., 10-25%, 25-50%, 50-75%, 75-90%, or 90-99%) includes a first primer binding sequence. In some embodiments, the primer binding sequence is the same (e.g., it can be used to prime sequencing or other extension reaction). In some embodiments, each probe or a subset of probes (e.g., 10-25%, 25-50%, 50-75%, 75-90%, or 90-99%) includes a unique

Petitioner Illumina, Inc., EX1004, p. 143

identifier sequence tag (e.g., that is predetermined and can be used to distinguish each probe).

In some embodiments, the methods disclosed herein are useful for any application where sensitivity is important. For example, detection of cancer mutations in a heterogenous tissue sample, detection of mutations in maternally-circulating fetal DNA, and detection of mutations in cells isolated during a preimplantation genetic diagnostic procedure.

According to some aspects of the invention, methods of detecting a polymorphism in a nucleic acid in a biological sample are provided. In some embodiments, the methods comprise obtaining a nucleic acid preparation using a preparative method (e.g., any of the preparative methods disclosed herein) on a biological sample, and performing a molecular inversion probe capture reaction on the nucleic acid preparation, wherein a molecular inversion probe capture (e.g., using a mutation-detection MIP) of a target nucleic acid of the nucleic acid preparation is indicative of the presence of a mutation (polymorphism) in the target nucleic acid, optionally wherein the polymorphism is selected from Table 2.

According to some aspects of the invention, methods of genotyping a nucleic acid in a biological sample are provided. In some embodiments, the methods comprise obtaining a nucleic acid preparation using a preparative method on a biological sample, sequencing a target nucleic acid of the nucleic acid preparation, and performing a molecular inversion probe capture reaction on the biological sample, wherein a molecular inversion probe capture of the target nucleic acid in the biological sample is indicative of the presence of a polymorphism in the target nucleic acid, genotyping the target nucleic acid based on the results of the sequencing and the capture reaction.

In some embodiments of the methods disclosed herein, the target nucleic acid has a sequence of a gene selected from Table 1.

It should be appreciated that any one or more embodiments described herein may be used for evaluating multiple genetic markers in parallel. Accordingly, in some embodiments, aspects of the invention relate to determining the presence of one or more markers (e.g., one or more alleles) at multiple different genetic loci in parallel. Accordingly, the risk or presence of multiple heritable disorders may be evaluated in parallel. In some embodiments, the risk of having offspring with one or more heritable

Petitioner Illumina, Inc., EX1004, p. 144

disorders may be evaluated. In some embodiments, an evaluation may be performed on a biological sample of a parent or a child (e.g., at a pre-implantation, prenatal, perinatal, or postnatal stage). In some embodiments, the disclosure provides methods for analyzing multiple genetic loci (e.g., a plurality of target nucleic acids selected from Table 1 or 2) from a patient sample, such as a blood, pre-implantation embryo, chorionic villus or amniotic fluid sample. A patient or subject may be a human. However, aspects of the invention are not limited to humans and may be applied to other species (e.g., mammals, birds, reptiles, other vertebrates or invertebrates) as aspects of the invention are not limited in this respect. A subject or patient may be male or female. In some embodiments, in connection with reproductive genetic counseling, samples from a male and female member of a couple may be analyzed. In some embodiments, for example, in connection with an animal breeding program, samples from a plurality of male and female subjects may be analyzed to determine compatible or optimal breeding partners or strategies for particular traits or to avoid one or more diseases or conditions. Accordingly, reproductive risks may be determined and/or reproductive recommendations may be provided based on information derived from one or more embodiments of the invention.

However, it should be appreciated that aspects of the invention may be used in connection with any medical evaluation where the presence of one or more alleles at a genetic locus of interest is relevant to a medical determination (e.g., risk or detection of disease, disease prognosis, therapy selection, therapy monitoring, etc.). Further aspects of the invention may be used in connection with detection, in tumor tissue or circulating tumor cells, of mutations in cellular pathways that cause cancer or predict efficacy of treatment regimens, or with detection and identification of pathogenic organisms in the environment or a sample obtained from a subject, *e.g.*, a human subject.

These and other aspects of the invention are described in more detail in the following description and non-limiting examples and drawings.

BRIEF DESCRIPTION OF DRAWINGS

FIG. 1 illustrates a non-limiting embodiment of a tiled probe layout;

FIG. 2 illustrates a non-limiting embodiment of a staggered probe layout;

FIG. 3 illustrates a non-limiting embodiment of an alternating staggered probe layout;

Petitioner Illumina, Inc., EX1004, p. 145

FIGs. 4A, B, and C depict various non-limiting methods for combining differentiator tag sequence and target sequences (NNNN depicts a differentiator tag sequence);

FIG. 5 depicts a non-limiting method for genotyping based on target and differentiator tag sequences;

FIG. 6 depicts non-limiting results of a simulation of a MIP capture reaction;

FIG. 7 depicts a non-limiting graph of sequencing coverage;

FIG. 8 illustrates that shorter sequences are captured with higher efficiency that longer sequences using MIPs;

FIG. 9 illustrates a non-limiting scheme of padlock (MIP) capture of a region that includes both repetitive regions (thick wavy line) and the adjacent unique sequence (thick strait line);

FIG. 10 illustrates a non-limiting hypothetical relationship between target gap size and the relative number of reads of the repetitive region;

FIG. 11A depicts MIP capture of FMR1 repeat regions from a diploid genome;

FIG. 11B depicts preparative methods for biallelic resolution of FMR1 repeat region lengths in a diploid genome using MIP capture probes and unique differentiator tags;

FIG. 11C depicts an analysis of FMR1 repeat region lengths in a diploid genome;

FIG. 12 is a schematic of an embodiment of an algorithm of the invention;

FIG. 13 illustrates a non-limiting example of a graph of per-target abundance with MIP capture; and,

FIG. 14 shows a non-limiting a graph of correlation between two MIP capture reactions.

DETAILED DESCRIPTION

Aspects of the invention relate to preparative and analytical methods and compositions for evaluating genotypes, and in particular, for determining the allelic identity (or identities in a diploid organism) of one or more genetic loci in a subject. Aspects of the invention are based, in part, on the identification of different sources of ambiguity and error in genetic analyses, and, in part, on the identification of one or more approaches to avoid, reduce, recognize, and/or resolve these errors and ambiguities at different stages in a genetic analysis. Aspects of the invention relate to methods and

Petitioner Illumina, Inc., EX1004, p. 146

compositions for addressing bias and/or stochastic variation associated with one or more preparative and/or analytical steps of a nucleic acid evaluation technology. In some embodiments, preparative methods can be adapted to avoid or reduce the risk of bias skewing the results of a genetic analysis. In some embodiments, analytical methods can be adapted to recognize and correct for data variations that may give rise to misinterpretation (e.g., incorrect calls such as homozygous when the subject is actually heterozygous or heterozygous when the subject is actually homozygous). Methods of the invention may be used for any type of mutation, for example a single base change (e.g., insertion, deletion, transversion or transition, etc.), a multiple base insertion, deletion, duplication, inversion, and/or any other change or combination thereof.

In some embodiments, additional or alternative techniques may be used to address loci characterized by multiple repeats of a core sequence where the length of the repeat is longer than a typical sequencing read thereby making it difficult to determine whether a deletion or duplication of one or more core sequence units has occurred based solely on a sequence read.

In some embodiments, increased confidence in an assay result may be obtained by i) selecting two or more different preparative and/or analytical techniques that have different biases (e.g., known to have different biases), ii) evaluating a patient sample using the two or more different techniques, iii) comparing the results from the two or more different techniques, and/or iv) determining whether the results are consistent for the two or more different techniques. In some embodiments, if determining in step (iv) indicates that the results are consistent (e.g., the same) then increased confidence in the assay result is obtained. In other embodiments, if determining in step (iv) indicates that the results are inconsistent (e.g., that the results are ambiguous) then one or more additional preparative and/or analytical techniques, which have a different bias (e.g., known to have a different bias) compared with the two or more different preparative and/or analytical techniques selected in step (i), are used to evaluate the patient sample, and the results of the one or more additional preparative and/or analytical techniques are compared with the results from step (ii) to resolve the inconsistency.

In some embodiments, two or more independent samples may be obtained from a subject and independently analyzed. In some embodiments, two or more independent samples are obtained at approximately the same time point. In some embodiments, two or more independent samples are obtained at multiple different time points. In some

Petitioner Illumina, Inc., EX1004, p. 147

embodiments, the use of two or more independent sample facilitates the elimination, normalization, and/or quantification of stochastic measurement noise. It is to be appreciated that two or more independent samples may be obtained in connection with any of the methods disclosed herein, including, for example, methods for pathogen profiling in a human or other animal subjects, monitoring tumor progression/regression, analyzing circulating tumor cells, analyzing fetal cells in maternal circulation, and analyzing/monitoring/profiling of environmental pathogens.

In some embodiments, one or more of the techniques described herein may be combined in a single assay protocol for evaluating multiple patient samples in parallel.

It should be appreciated that aspects of the invention may be useful for high throughput, cost-effective, yet reliable, genotyping of multiple patient samples (e.g., in parallel, for example in multiplex reactions). In some embodiments, aspects of the invention are useful to reduce the error frequency in a multiplex analysis. Certain embodiments may be particularly useful where multiple reactions (e.g., multiple loci and/or multiple patient samples) are being processed. For example, 10-25, 25-50, 50-75, 75-100 or more loci may be evaluated for each subject out of any number of subject samples that may be processed in parallel (e.g., 1-25, 25-50, 50-100, 100-500, 500-1,000, 1,000-2,500, 2,500-5,000 or more or intermediate numbers of patient samples). It should be appreciated that different embodiments of the invention may involve conducting two or more target capture reactions and/or two or more patient sample analyses in parallel in a single multiplex reaction. For example, in some embodiments a plurality of capture reactions (e.g., using different capture probes for different target loci) may be performed in a single multiplex reaction on a single patient sample. In some embodiments, a plurality of captured nucleic acids from each one of a plurality of patient samples may be combined in a single multiplex analysis reaction. In some embodiments, samples from different subjects are tagged with subject-specific (e.g., patient-specific) tags (e.g., unique sequence tags) so that the information from each product can be assigned to an identified subject. In some embodiments, each of the different capture probes used for each patient sample have a common patient-specific tag. In some embodiments, the capture probes do not have patient-specific tags, but the captured products from each subject may be amplified using one or a pair of amplification primers that are labeled with a patient-specific tag. Other techniques for associating a patient-specific tag with the captured product from a single patient sample may be used as aspects of the

Petitioner Illumina, Inc., EX1004, p. 148

invention are not limited in this respect. It should be appreciated that patient-specific tags as used herein may refer to unique tags that are assigned to identified patients in a particular assay. The same tags may be used in a separate multiplex analysis with a different set of patient samples (e.g., from different patients) each of which is assigned one of the tags. In some embodiments, different sets of unique tags may be used in sequential (e.g., alternating) multiplex reactions in order to reduce the risk of contamination from one assay to the next and allow contamination to be detected on the basis of the presence of tags that are not expected to be present in a particular assay.

Embodiments of the invention may be used for any of a number of different settings: reproductive settings, disease screening, identifying subjects having cancer, identifying subjects having increased risk for a disease, stratifying a population of subjects according to one or more of a number of factors, for example responsiveness to a particular drug, lack or not of an adverse reaction (or risk therefore) to a particular drug, and/or providing information for medical records (e.g., homozygosity, heterozygosity at one or more loci). It should be appreciated that the invention is not limited to genomic analysis of patient samples. For example, aspects of the invention may be useful for high throughput genetic analysis of environment samples to detect pathogens.

In some embodiments, the methods disclosed herein are useful for diagnosis of one or more heritable disorders. In some embodiments, a heritable disorder that may be diagnosed with the methods disclosed herein is a genetic disorder that is prevalent in the Ashkenazi Jewish population. In some embodiments, the heritable disorders are selected from: 21-Hydroxylase-Defiocient Congenital Adrenal Hyperplasia; ABCC8-Related Hyperinsulinism; Alpha-Thalassemia, includes Constant Spring, & MR associated; Arylsulfatase A Deficiency-Metyachromatic Leukodystrophy; Biotinidase Deficiency-Holocarboxylase Synthetase Deficiency; Bloom's Syndrome; Canavan Disease; CFTR-Related Disorders-cystic fibrosis; Citrullinemia Type I; Combined MMA & Homocystinuria-dblC; Dystrophinopathies (DMD & BMD); Familial Dysautonomia; Fanconi Anemia-FANCC; Galactosemia-Classical: Galactokinase Defiency & Galactose Epimerase Deficiency; Gaucher Disease; GJB2-Related DFNB 1 Nonsyndromic Hearing Loss and Deafness; Glutaric acidemia Type 1; Hemoglobinopathies beta-chain disorders; Glycogen Storage Disease Type 1A; Maple Syrup Urine Disease; Types 1A, 1B, 2, 3; Medium Chain Acyl-Coenzyme A; Dehydrogenase Deficiency-MCADD;

Petitioner Illumina, Inc., EX1004, p. 149

Methylmalonic Acidemia; Mucolipidosis IV; Nemaline Myopathy; Nieman-Pick Type A-Acid Sphingomyelinase Deficiency; Non-Ketotic Hyperglycinemia-Glycine Encephalopathy; Ornithine Transcarbamylase Deficiency; PKU Phenylalanine Hydroxylase Deficiency; Propionic Acidemia; Short Chain Acyl-CoA Dehydrogenase Deficiency-SCADD; Smith-Lemli-Opitz Syndrome; Spinal Muscular Atrophy (SMN1)-SMA; Tay Sachs-HexA Deficiency; Usher Synbdrome-Type I (Type IB, Type IC, Type ID, Type IF, Type IG) ; X-Linked Mental Retardation ARX-Related Disorders; X-Linked Mental Retardation with Cerebellar Cypoplasia and sistinctive Facial Appearance; X-Linked Mental Retardation; includes 9, 21, 30, 46, 58, 63, 88, 89; X-linked mental retardation: FM1-Related Disorders-FRXA, Fragile X MR; X-linked SMR: Renpenning Syndrome 1; Zellweger Spectrum disorders - Peroxisomal Bifunctional Enzyme Deficiencies including Zellweger, NALD, and/or infantile Refsums. However, all of these, subsets of these, other genes, or combinations thereof may be used.

According to some aspects, the disclosure relates to multiplex diagnostic methods. In some embodiments, multiplex diagnostic methods comprise capturing a plurality of genetic loci in parallel (e.g., a genetic locus of Table 1). In some embodiments, genetic loci possess one or more polymorphisms (e.g., a polymorphism of Table 2) the genotypes of which correspond to disease causing alleles. Accordingly, in some embodiments, the disclosure provides methods for assessing multiple heritable disorders in parallel.

In some embodiments, methods are provided for diagnosing multiple heritable disorders in parallel at a pre-implantation, prenatal, perinatal, or postnatal stage. In some embodiments, the disclosure provides methods for analyzing multiple genetic loci (e.g., a plurality of target nucleic acids selected from Table 1) from a patient sample, such as a blood, pre-implantation embryo, chorionic villus or amniotic fluid sample. A patient or subject may be a human. However, aspects of the invention are not limited to humans and may be applied to other species (e.g., mammals, birds, reptiles, other vertebrates or invertebrates) as aspects of the invention are not limited in this respect. A subject or patient may be male or female. In some embodiments, in connection with reproductive genetic counseling, samples from a male and female member of a couple may be analyzed. In some embodiments, for example, in connection with an animal breeding program, samples from a plurality of male and female subjects may be analyzed to

Petitioner Illumina, Inc., EX1004, p. 150

determine compatible or optimal breeding partners or strategies for particular traits or to avoid one or more diseases or conditions.

However, it should be appreciated that any other diseases may be studied and/or risk factors for diseases or disorders including, but not limited to allergies, responsiveness to treatment, cancer tumor profiling for treatment and prognosis, monitoring and identification of patient infections, and monitoring of environmental pathogens.

1. Reducing representational bias in multiplex amplification reactions:

In some embodiments, aspects of the invention relate to methods that reduce bias and increase reproducibility in multiplex detection of genetic loci, e.g., for diagnostic purposes.

Molecular inversion probe technology is used to detect or amplify particular nucleic acid sequences in potentially complex mixtures. Use of molecular inversion probes has been demonstrated for detection of single nucleotide polymorphisms (Hardenbol et al. 2005 Genome Res 15:269-75) and for preparative amplification of large sets of exons (Porreca et al. 2007 Nat Methods 4:931-6, Krishnakumar et al. 2008 Proc Natl Acad Sci USA 105:9296-301). One of the main benefits of the method is in its capacity for a high degree of multiplexing, because generally thousands of targets may be captured in a single reaction containing thousands of probes. However, challenges associated with, for example, amplification efficiency (See, e.g., Turner EH, et al., Nat Methods. 2009 Apr 6:1-2.) have limited the practical utility of the method in research and diagnostic settings.

Aspects of the disclosure are based, in part, on the discovery of effective methods for overcoming challenges associated with systematic errors (bias) in multiplex genomic capture and sequencing methods, namely high variability in target nucleic acid representation and unequal sampling of heterozygous alleles in pools of captured target nucleic acids (e.g., isolated from a biological sample). Accordingly, in some embodiments, the disclosure provides methods that reduce variability in the detection of target nucleic acids in multiplex capture methods. In other embodiments, methods improve allelic representation in a capture pool and, thus, improve variant detection outcomes. In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve the use of different sets of

Petitioner Illumina, Inc., EX1004, p. 151

multiple probes (e.g., molecular inversion probes MIPs) that capture overlapping regions of a target nucleic acid to achieve a more uniform representation of the target nucleic acids in a capture pool compared with methods of the prior art. In other embodiments, methods reduce bias, or the risk of bias, associated with large scale parallel capture of genetic loci, e.g., for diagnostic purposes. In other embodiments, methods are provided for increasing reproducibility (e.g., by reducing the effect of polymorphisms on target nucleic acid capture) in the detection of a plurality of genetic loci in parallel. In further embodiments, methods are provided for reducing the effect of probe synthesis and/or probe amplification variability on the analysis of a plurality of genetic loci in parallel.

In some aspects, the disclosure provides probe sets that comprise a plurality of different probes. As used herein, a 'probe' is a nucleic acid having a central region flanked by a 5' region and a 3' region that are complementary to nucleic acids flanking the same strand of a target nucleic acid or subregion thereof. An exemplary probe is a molecular inversion probe (MIP). A 'target nucleic acid' may be a genetic locus. Exemplary genetic loci are disclosed herein in Table 1 (RefSeqGene Column).

While probes have been typically designed to meet certain constraints (e.g. melting temperature, G/C content, etc.) known to partially affect capture/amplification efficiency (Ball et al (2009) Nat Biotech 27:361-8 AND Deng et al (2009) Nat Biotech 27:353-60), a set of constraints which is sufficient to ensure either largely uniform or highly reproducible capture/amplification efficiency has not previously been achieved. As disclosed herein, uniformity and reproducibility can be increased by designing multiple probes per target, such that each base in the target is captured by more than one probe. In some embodiments, the disclosure provides multiple MIPs per target to be captured, where each MIP in a set designed for a given target nucleic acid has a central region and a 5' region and 3' region ('targeting arms') which hybridize to (at least partially) different nucleic acids in the target nucleic acid (immediately flanking a subregion of the target nucleic acid). Thus, differences in efficiency between different targeting arms and fill-in sequences may be averaged across multiple MIPs for a single target, which results in more uniform and reproducible capture efficiency.

In some embodiments, the methods involve designing a single probe for each target (a target can be as small as a single base or as large as a kilobase or more of contiguous sequence).

Petitioner Illumina, Inc., EX1004, p. 152

It may be preferable, in some cases, to design probes to capture molecules (e.g., target nucleic acids or subregions thereof) having lengths in the range of 1 - 200 bp (as used herein, a bp refers to a base pair on a double-stranded nucleic acid – however, where lengths are indicated in bps, it should be appreciated that single-stranded nucleic acids having the same number of bases, as opposed to base pairs, in length also are contemplated by the invention). However, probe design is not so limited. For example, probes can be designed to capture targets having lengths in the range of up to 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 1000, or more bps, in some cases.

It is to be appreciated that the length of a capture molecule (e.g., a target nucleic acid or subregion thereof) is selected based upon multiple considerations. For example, where analysis of a target involves sequencing, e.g., with a next-generation sequencer, the target length should typically match the sequencing read-length so that shotgun library construction is not necessary. However, it should be appreciated that captured nucleic acids may be sequenced using any suitable sequencing technique as aspects of the invention are not limited in this respect.

It is also to be appreciated that some target nucleic acids are too large to be captured with one probe. Consequently, it may be necessary to capture multiple subregions of a target nucleic acid in order to analyze the full target.

In some embodiments, a subregion of a target nucleic acid is at least 1 bp. In other embodiments, a subregion of a target nucleic acid is at least 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 bp or more. In other embodiments, a subregion of a target nucleic acid has a length that is up to 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or more percent of a target nucleic acid length.

The skilled artisan will also appreciate that consideration is made, in the design of MIPs, for the relationship between probe length and target length. In some embodiments, MIPs are designed such that they are several hundred basepairs (e.g., up to 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 bp or more) longer than corresponding target (e.g., subregion of a target nucleic acid, target nucleic acid).

In some embodiments, lengths of subregions of a target nucleic acid may differ. For example, if a target nucleic acid contains regions for which probe hybridization is not possible or inefficient, it may be necessary to use probes that capture subregions of

Petitioner Illumina, Inc., EX1004, p. 153

one or more different lengths in order to avoid hybridization with problematic nucleic acids and capture nucleic acids that encompass a complete target nucleic acid.

Aspects of the invention involve using multiple probes, e.g., MIPs, to amplify each target nucleic acid. In some embodiments, the set of probes for a given target can be designed to 'tile' across the target, capturing the target as a series of shorter sub-targets. In some embodiments, where a set of probes for a given target is designed to 'tile' across the target, some probes in the set capture flanking non-target sequence). Alternately, the set can be designed to 'stagger' the exact positions of the hybridization regions flanking the target, capturing the full target (and in some cases capturing flanking non-target sequence) with multiple probes having different targeting arms, obviating the need for tiling. The particular approach chosen will depend on the nature of the target set. For example, if small regions are to be captured, a staggered-end approach might be appropriate, whereas if longer regions are desired, tiling might be chosen. In all cases, the amount of bias-tolerance for probes targeting pathological loci can be adjusted ('dialed in') by changing the number of different MIPs used to capture a given molecule.

In some embodiments, the 'coverage factor', or number of probes used to capture a basepair in a molecule, is an important parameter to specify. Different numbers of probes per target are indicated depending on whether one is using the tiling approach (see, e.g., FIG. 1) or one of the staggered approaches (see, e.g., FIGs. 2 or 3).

FIG. 1 illustrates a non-limiting embodiment of a tiled probe layout showing ten captured sub-targets tiled across a single target. Each position in the target is covered by three sub-targets such that MIP performance per base pair is averaged across three probes.

FIG. 2 illustrates a non-limiting embodiment of a staggered probe layout showing the targets captured by a set of three MIPs. Each MIP captures the full target, shown in black, plus (in some cases) additional extra-target sequence, shown in gray, such that the targeting arms of each MIP fall on different sequence. Each position in the target is covered by three sub-targets such that MIP performance per basepair is averaged across three probes. Targeting arms land immediately adjacent to the black or gray regions shown. It should be appreciated that in some embodiments, the targeting arms (not shown) can be designed so that they do not overlap with each other.

FIG. 3 illustrates a non-limiting embodiment of an alternating staggered probe layout showing the targets captured by a set of three MIPs. Each MIP captures the full

Petitioner Illumina, Inc., EX1004, p. 154

target, shown in black, plus (in some cases) additional extra-target sequence, shown in gray, such that the targeting arms of each MIP fall on different sequence. Each position in the target is covered by three sub-targets such that MIP performance per basepair is averaged across three probes. Targeting arms land immediately adjacent to the black or gray regions shown.

It should be appreciated that for any of the layouts, the targeting arms on adjacent tiled or staggered probes may be designed to either overlap, not overlap, or overlap for only a subset of the probes.

In certain embodiments for any of the layouts, a coverage factor of about 3 to about 10 is used. However, the methods are not so limited and coverage factors of up to 2, 3, 4, 5, 6, 7, 8, 9, 10, 20 or more may be used. It is to be appreciated that the coverage factor selected may depend the probe layout being employed. For example, in the tiling approach, for a desired coverage factor, the number of probes per target is typically a function of target length, sub-target length, and spacing between adjacent sub-target start locations (step size). For example, for a desired coverage factor of 3, a 200 bp target with a start-site separation of 20 bp and sub-target length of 60 bp may be encompassed with 12 MIPs (FIG. 1). Thus, a specific coverage factor may be achieved by varying the number of probes per target nucleic acid and the length of the molecules captured. In the staggered approach, a fixed-length target nucleic acid is captured as several subregions or as 'super-targets', which are molecules comprising the target nucleic acid and additional flanking nucleic acids, which may be of varying lengths. For example, a target of 50 bp can be captured at a coverage factor of 3 with 3 probes in either a 'staggered' (FIG. 2) or 'alternating staggered' configuration (FIG. 3).

The coverage factor will be driven by the extent to which detection bias is tolerable. In some cases, where the bias tolerance is small, it may be desirable to target more subregions of target nucleic acid with, perhaps, higher coverage factors. In some embodiments, the coverage factor is up to 2, 3, 4, 5, 6, 7, 8, 9, 10 or more.

In some embodiments, when a tiled probe layout is used, when the target length is greater than 1 bp and when a step size (distance between the 5'-end of a target and the 5' end of its adjacent target) is less than the length of a target or subregion thereof, it is possible to compute probe number for a particular target based on target length (T), sub-target length (S), and coverage factor (C), such that probe number = T/(S/C) + (C-1).

Petitioner Illumina, Inc., EX1004, p. 155

In some aspects, the disclosure provides methods to increase the uniformity of amplification efficiency when multiple molecules are amplified in parallel; methods to increase the reproducibility of amplification efficiency; methods to reduce the contribution of targeting probe variability to amplification efficiency; methods to reduce the effect on a given target nucleic acid of polymorphisms in probe hybridization regions; and/or methods to simplify downstream workflows when multiplex amplification by MIPs is used as a preparative step for analysis by nucleic acid sequencing.

Polymorphisms in the target nucleic acid under the regions flanking a target can interfere with hybridization, polymerase fill-in, and/or ligation. Furthermore, this may occur for only one allele, resulting in allelic drop-out, which ultimately decreases downstream sequencing accuracy. In some embodiments, using a set of MIPs having multiple hybridization sites for the capture of any given target, the probability of loss from polymorphism is substantially decreased because not all targeting arms in the set of MIPs will cover the location of the mutation.

Probes for MIP capture reactions may be synthesized on programmable microarrays because of the large number of sequences required. Because of the low synthesis yields of these methods, a subsequent amplification step is required to produce sufficient probe for the MIP amplification reaction. The combination of multiplex oligonucleotide synthesis and pooled amplification results in uneven synthesis error rates and representational biases. By synthesizing multiple probes for each target, variation from these sources may be averaged out because not all probes for a given target will have the same error rates and biases.

Multiplex amplification strategies disclosed herein may be used analytically, as in detection of SNPs, or preparatively, often for next-generation sequencing or other sequencing techniques. In the preparative setting, the output of an amplification reaction is generally the input to a shotgun library protocol, which then becomes the input to the sequencing platform. The shotgun library is necessary in part because next-generation sequencing yields reads significantly shorter than amplicons such as exons. In addition to the bias-reduction afforded by the multi-tiled approach described here, tiling also obviates the need for shotgun library preparation. Since the length of the capture molecule can be specified when the probes, e.g., MIPs, are designed, it can be chosen to

Petitioner Illumina, Inc., EX1004, p. 156

match the readlength of the sequencer. In this way, reads can 'walk' across an exon by virtue of the start position of each capture molecule in the probe set for that exon.

2. Reducing analytical errors associated with bias in nucleic acid preparations:

In some embodiments, aspects of the invention relate to preparative steps in DNA sequencing-related technologies that reduce bias and increase the reliability and accuracy of downstream quantitative applications.

There are currently many genomics assays that utilize next-generation (polony-based) sequencing to generate data, including genome resequencing, RNA-seq for gene expression, bisulphite sequencing for methylation, and Immune-seq, among others. In order to make quantitative measurements (including genotype calling), these methods utilize the counts of sequencing reads of a given genomic locus as a proxy for the representation of that sequence in the original sample of nucleic acids. The majority of these techniques require a preparative step to construct a high-complexity library of DNA molecules that is representative of a sample of interest. This may include chemical or biochemical treatment of the DNA (e.g., bisulphite treatment), capture of a specific subset of the genome (e.g., padlock probe capture, solution hybridization), and a variety of amplification techniques (e.g., polymerase chain reaction, whole genome amplification, rolling circle amplification).

Systematic and random errors are common problems associated with genome amplification and sequencing library construction techniques. For example, genomic sequencing library may contain an over- or under-representation of particular sequences from a source genome as a result of errors (bias) in the library construction process. Such bias can be particularly problematic when it results in target sequences from a genome being absent or undetectable in the sequencing libraries. For example, an under-representation of particular allelic sequences (e.g., heterozygotic alleles) from a genome in a sequencing library can result in an apparent homozygous representation in a sequencing library. As most downstream sequencing library quantification techniques depend on stochastic counting processes, these problems have typically been addressed by sampling enough (over-sampling) to obtain a minimum number of observations necessary to make statistically significant decisions. However, the strategy of oversampling is generally limited to elimination of low-count Poisson noise, and the

Petitioner Illumina, Inc., EX1004, p. 157

approach wastes resources and increases the expense required to perform such experiments. Moreover, oversampling can result in a reduced statistical confidence in certain conclusions (e.g., diagnostic calls) based on the data. Accordingly, new approaches are needed for overcoming bias in sequencing library preparatory methods.

Aspects of the disclosure are based, in part, on the discovery of methods for overcoming problems associated with systematic and random errors (bias) in genome capture, amplification and sequencing methods, namely high variability in the capture and amplification of nucleic acids and disproportionate representation of heterozygous alleles in sequencing libraries. Accordingly, in some embodiments, the disclosure provides methods that reduce variability in the capture and amplification of nucleic acids. In other embodiments, the methods improve allelic representation in sequencing libraries and, thus, improve variant detection outcomes. In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve the use of differentiator tag sequences to uniquely tag individual nucleic acid molecules. In some embodiments, the differentiator tag sequence permits the detection of bias based on the frequency with which pairs of differentiator tag and target sequences are observed in a sequencing reaction. In other embodiments, the methods reduce errors caused by bias, or the risk of bias, associated with the capture, amplification and sequencing of genetic loci, e.g., for diagnostic purposes.

Aspects of the invention relate to associating unique sequence tags (referred to as differentiator tag sequences) with individual target molecules that are independently captured and/or analyzed (e.g., prior to amplification or other process that may introduce bias). These tags are useful to distinguish independent target molecules from each other thereby allowing an analysis to be based on a known number of individual target molecules. For example, if each of a plurality of target molecule sequences obtained in an assay is associated with a different differentiator tag, then the target sequences can be considered to be independent of each other and a genotype likelihood can be determined based on this information. In contrast, if each of the plurality of target molecule sequences obtained in the assay is associated with the same differentiator tag, then they probably all originated from the same target molecule due to over-representation (e.g., due to biased amplification) of this target molecule in the assay. This provides less information than the situation where each nucleic acid was associated with a different

Petitioner Illumina, Inc., EX1004, p. 158

differentiator tag. In some embodiments, a threshold number of independently isolated molecules (e.g., unique combinations of differentiator tag and target sequences) is analyzed to determine the genotype of a subject.

In some embodiments, the invention relates to compositions comprising pools (libraries) of preparative nucleic acids that each comprise "differentiator tag sequences" for detecting and reducing the effects of bias, and for genotyping target nucleic acid sequences. As used herein, a "differentiator tag sequence" is a sequence of a nucleic acid (a preparative nucleic acid), which in the context of a plurality of different isolated nucleic acids, identifies a unique, independently isolated nucleic acid. Typically, differentiator tag sequences are used to identify the origin of a target nucleic acid at one or more stages of a nucleic acid preparative method. For example, in the context of a multiplex nucleic acid capture reaction, differentiator tag sequences provide a basis for differentiating between multiple independent, target nucleic acid capture events. Also, in the context of a multiplex nucleic acid amplification reaction, differentiator tag sequences provide a basis for differentiating between multiple independent, primary amplicons of a target nucleic acid, for example. Thus, combinations of target nucleic acid and differentiator tag sequence (target:differentiator tag sequences) of an isolated nucleic acid of a preparative method provide a basis for identifying unique, independently isolated target nucleic acids. FIG. 4A-C depict various non-limiting examples of methods for combining differentiator tag sequence and target sequences.

It will be apparent to the skilled artisan that differentiator tags may be synthesized using any one of a number of different methods known in the art. For example, differentiator tags may be synthesized by random nucleotide addition. Differentiator tag sequences are typically of a predefined length, which is selected to control the likelihood of producing unique target:differentiator tag sequences in a preparative reaction (e.g., amplification-based reaction, a circularization selection-based reaction, e.g., a MIP reaction). Differentiator tag sequences may be, up to 5, up to 6, up to 7 up to 8, up to 9, up to 10, up to 11, up to 12, up to 13, up to 14, up to 15, up to 16, up to 17, up to 18, up to 19, up to 20, up to 21, up to 22, up to 23, up to 24, up to 25, or more nucleotides in length. For purposes of genotyping, isolated nucleic acids are identified as independently isolated if they comprise unique combinations of target nucleic acid and differentiator tag sequences, and observance of threshold numbers of

Petitioner Illumina, Inc., EX1004, p. 159

unique combinations of target nucleic acid and differentiator tag sequences provide a certain statistical confidence in the genotype.

During a library preparation process, each nucleic acid molecule may be tagged with a unique differentiator tag sequence in a configuration that permits the differentiator tag sequence to be sequenced along with the target nucleic acid sequence of interest (the nucleic acid sequence for which the library is being prepared, e.g., a polymorphic sequence). The incorporation of the nucleic acid comprising a differentiator tag sequence at a particular step allows the detection and correction of biases in subsequent steps of the protocol.

A large library of unique differentiator tag sequences may be created by using degenerate, random-sequence polynucleotides of defined length. The differentiator tag sequences of the polynucleotides may be read at the final stage of the sequencing. The observations of the differentiator tag sequences may be used to detect and correct biases in the final sequencing read-out of the library. For example, the total possible number of differentiator tag sequences, which may be produced, e.g., randomly, is 4^N, where N is the length of the differentiator tag sequence. Thus, it is to be understood that the length of the differentiator tag sequence may be adjusted such that the size of the population of MIPs having unique differentiator tag sequences is sufficient to produce a library of MIP capture products in which identical independent combinations of target nucleic acid and differentiator tag sequence are rare. As used herein combinations of target nucleic acid and differentiator tag sequences, may also be referred to as "target:differentiator tag sequences".

In the final readout of a sequencing process, each read may have an additional unique differentiator tag sequence. In some embodiments, when differentiator tag sequences are distributed randomly in a library, all the unique differentiator tag sequences will be observed about an equal number of times. Accordingly, the number of occurrences of a differentiator tag sequence may follow a Poisson distribution.

In some embodiments, overrepresentation of target:differentiator tag sequences in a pool of preparative nucleic acids (e.g., amplified MIP capture products) is indicative of bias in the preparative process (e.g., bias in the amplification process). For example, target:differentiator tag sequence combinations that are statistically overrepresented are indicative of bias in the protocol at one or more steps between the incorporation of the

Petitioner Illumina, Inc., EX1004, p. 160

differentiator tag sequences into MIPs and the actual sequencing of the MIP capture products.

The number of reads of a given target:differentiator tag sequence may be indicative (may serve as a proxy) of the amount of that target sequence present in the originating sample. In some embodiments, the numbers of occurrence of sequences in the originating sample is the quantity of interest. For example, using the methods disclosed herein, the occurrence of differentiator tag sequences in a pool of MIPs may be predetermined (e.g., may be the same for all differentiator tag sequences). Accordingly, changes in the occurrence of differentiator tag sequences after amplification and sequencing may be indicative of bias in the protocol. Bias may be corrected to provide an accurate representation of the composition of the original MIP pool, e.g., for diagnostic purposes.

According to some aspects, a library of preparative nucleic acid molecules (e.g., MIPs, each nucleic acid in the library having a unique differentiator tag sequence, may be constructed such that the number of nucleic acid molecules in the library is significantly larger than the number prospective target nucleic acid molecules to be captured using the library. This ensures that products of the preparative methods include only unique target:differentiator tag sequence; e.g., in a MIP reaction the capture step would undersample the total population of unique differentiator tag sequences in the MIP library. For example, an experiment utilizing 1 μg of genomic DNA will contain about ~150,000 copies of a diploid genome. For a MIP library, each MIP in the library comprising a randomly produced 12-mer differentiator tag sequence (~1.6 million possible unique differentiator tag sequences), there would be more than 100 unique differentiator tag sequences per genomic copy. For a MIP library, each MIP in the library comprising a randomly produced 15-mer differentiator tag sequence (~1 billion possible unique differentiator tag sequences), there would be more than 7000 unique differentiator tag sequences per genomic copy. Therefore, the probability of the same differentiator tag sequence being incorporated multiple times is incredibly small. Thus, it is to be appreciated that the length of the differentiator tag sequence is to be selected based on the amount of target sequence in a MIP capture reaction and the desired probability for having multiple, independent occurrences of target:differentiator tag sequence combinations.

Petitioner Illumina, Inc., EX1004, p. 161

FIG. 5 depicts a non-limiting method for genotyping based on target and differentiator tag sequences. Sequencing reads of target and differentiator tags sequences are collapsed to make diploid genotype calls. FIG. 6 depicts non-limiting results of a simulation of a MIP capture reaction in which MIP probes, each having a differentiator tag sequence of 15 nucleotides, are combined with 10000 target sequence copies (e.g., genome equivalents). In this simulated reaction, the probability of capturing one or more copies of a target sequence having the same differentiator tag sequence is 0.05. The Y axis reflects the number of observations. The X axis reflects the number of independent occurrences of target:differentiator tag combinations. FIG. 7 depicts a non-limiting graph of sequencing coverage, which can help ensure that alleles are sampled to sufficient depth (e.g., either 10x or 20x minimum sampling per allele, assuming 1000 targets). In this non-limiting example, the X axis is total per-target coverage required, and the Y axis is the probability that a given total coverage will result in at least 10x or 20x coverage for each allele.

The skilled artisan will appreciate that as part of a MIP library preparation process, adapters may be ligated onto the ends of the molecules of interest. Adapters often contain PCR primer sites (for amplification or emulsion PCR) and/or sequencing primer sites. In addition, barcodes may be included, for example, to uniquely identify individual samples (e.g., patient samples) that may be mixed together. (See, e.g., USPTO Publication Number US 2007/0020640 A1 (McCloskey et al.)

The actual incorporation of the random differentiator tag sequences can be performed through various methods known in the art. For example, nucleic acids comprising differentiator tag sequences may be incorporated by ligation. This is a flexible method, because molecules having differentiator tag sequence can be ligated to any blunt-ended nucleic acids. The sequencing primers must be incorporated subsequently such that they sequence both the differentiator tag sequence and the target sequence. Alternatively, the sequencing adaptors can be synthesized with the random differentiator tag sequences at their 3' end (as degenerate bases), so that only one ligation must be performed. Another method is to incorporate the differentiator tag sequence into a PCR primer, such that the primer structure is arranged with the common adaptor sequence followed by the random differentiator tag sequence followed by the PCR priming sequence (in 5' to 3' order). A differentiator tag sequence and adaptor sequence (which may contain the sequencing primer site) are incorporated as tags. Another

Petitioner Illumina, Inc., EX1004, p. 162

method to incorporate the differentiator tag sequences is to synthesize them into a padlock probe prior to performing a gene capture reaction. The differentiator tag sequence is incorporated 3' to the targeting arm but 5' to the amplification primer that will be used downstream in the protocol. Another method to incorporate the differentiator tag sequences is as a tag on a gene-specific or poly-dT reverse-transcription primer. This allows the differentiator tag sequence to be incorporated directly at the cDNA level.

In some embodiments, at the incorporation step, the distribution of differentiator tag sequences can be assumed to be uniform. In this case, bias in any part of the protocol would change the uniformity of this distribution, which can be observed after sequencing. This allows the differentiator tag sequence to be used in any preparative process where the ultimate output is sequencing of many molecules in parallel.

Differentiator tag sequences may be incorporated into probes (e.g., MIPs) of a plurality when they are synthesized on-chip in parallel, such that degeneracy of the incorporated nucleotides is sufficient to ensure near-uniform distribution in the plurality of probes. It is to be appreciated that amplification of a pool of unique differentiator tag sequences may itself introduce bias in the initial pool. However, in most practical cases, the scale of synthesis (e.g., by column synthesis, chip based synthesis, etc.) is large enough that amplification of an initial pool of differentiator tag sequences is not necessary. By avoiding amplification or selection steps on the pool of unique differentiator tag sequences, potential bias may be minimized.

One example of the use of the differentiator tag sequences is in genome re-sequencing. Considering that the raw accuracy of most next-generation sequencing instruments is relatively low, it is crucial to oversample the genomic loci of interest. Furthermore, since there are two alleles at every locus, it is important to sample enough to ensure that both alleles have been observed a sufficient number of times to determine with a sufficient degree of statistical confidence whether the sample is homozygous or heterozygous. Indeed, the sequencing is performed to sample the composition of molecules in the originating sample. However, after multiple reads have been collected for a given locus, it is possible that due to bias (e.g., caused by PCR amplification steps), a large fraction of the reads are derived from a single originating molecule. This would skew the population of target sequences observed, and would affect the outcome of the genotype call. For example, it is possible that a locus that is heterozygous is called as

Petitioner Illumina, Inc., EX1004, p. 163

homozygous, because there are only a few observations of the second allele out of many observations of that locus. However, if information is available on differentiator tag sequences, this situation could be averted, because the over-represented allele would be seen to also have an over-represented differentiator tag sequence (i.e., the sequences with the overrepresented differentiator tag sequence all originated from the same single molecule). Therefore, the sequences and corresponding distribution of differentiator tag sequences can be used as an additional input to the genotype-calling algorithm to significantly improve the accuracy and confidence of the genotype calls.

In some aspects, the disclosure provides methods for analyzing a plurality of target sequences which are genetic loci or portions of genetic loci (e.g., a genetic locus of Table 1). The genetic loci may be analyzed by sequencing to obtain a genotype at one or more polymorphisms (e.g., SNPs). Exemplary polymorphisms are disclosed in Table 2. The skilled artisan will appreciate that other polymorphisms are known in the art and may be identified, for example, by querying the Entrez Single Nucleotide Polymorphism database, for example, by searching with a GeneID from Table 1.

Table 1: Target Nucleic Acids

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
CYP21A2	1589	cytochrome P450, family 21, subfamily A, polypeptide 2	CAH1; CPS1; CA21H; CYP21; CYP21B; P450c21B; MGC150536; MGC150537; CYP21A2	201910	NG_008337.1	6p21.3
ABCC8	6833	ATP-binding cassette, sub-family C (CFTR/MRP), member 8	HI; SUR; HHF1; MRP8; PHHI; SUR1; ABC36; HRINS; TNDM2; ABCC8	600509	NG_008867.1	11p15.1
ATRX	546	alpha thalassemia/mental retardation syndrome X-linked (RAD54 homolog, S. cerevisiae)	SHS; XH2; XNP; ATR2; SFM1; RAD54; MRXHF1; RAD54L; ZNF-HX; MGC2094; ATRX	300032	NG_008838.1	Xq13.1-q21.1
ARSA	410	arylsulfatase A	MLD; ARSA	607574	NG_009260.1	22q13.31-qter; 22q13.33
PSAP	5660	Prosaposin	GLBA; SAP1; FLJ00245; MGC110993; PSAP	176801	NG_008835.1	10q21-q22
BTD	686	Biotinidase	BTD	609019	NG_008019.1	3p25

Petitioner Illumina, Inc., EX1004, p. 164

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
HLCS	3141	holocarboxylase synthetase (biotin-(proprionyl-Coenzyme A-carboxylase (ATP-hydrolysing)) ligase)	HCS; HLCS	609018	NC_000021.7	21q22.1; 21q22.13
BLM	641	Bloom syndrome, RecQ helicase-like	BS; RECQ2; RECQL2; RECQL3; MGC126616; MGC131618; MGC131620; BLM	604610	NG_007272.1	15q26.1
ASPA	443	aspartoacylase (Canavan disease)	ASP; ACY2; ASPA	608034	NG_008399.1	17pter-p13
CFTR	1080	cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7)	CF; MRP7; ABC35; ABCC7; CFTR/MRP; TNR-CFTR; dJ760C5.1; CFTR	602421	NC_000007.12	7q31.2
ASS1	445	argininosuccinate synthetase 1	ASS; CTLN1; ASS1	603470	NG_011542.1	9q34.1
MMACHC	25974	methylmalonic aciduria (cobalamin deficiency) cblC type, with homocystinuria	cblC; FLJ25671; DKFZp564I122; RP11-291L19.3; MMACHC	609831	NC_000001.9	1p34.1
IKBKAP	8518	inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase complex-associated protein	FD; DYS; ELP1; IKAP; IKI3; TOT1; FLJ12497; DKFZp781H1425; IKBKAP	603722	NG_008788.1	9q31
FANCC	2176	Fanconi anemia, complementation group C	FA3; FAC; FACC; FLJ14675; FANCC	227645	NG_011707.1	9q22.3
GALK1	2584	galactokinase 1	GK1; GALK; GALK1	604313	NG_008079.1	17q24
GALT	2592	galactose-1-phosphate uridylyltransferase	GALT	606999	NC_000009.10	9p13
GALE	2582	UDP-galactose-4-epimerase	SDR1E1; FLJ95174; FLJ97302; GALE	606953	NG_007068.1	1p36-p35
GBA	2629	glucosidase, beta; acid (includes glucosylceramidase)	GCB; GBA1; GLUC; GBA	606463	NG_009783.1	1q21
GJB2	2706	gap junction protein, beta 2, 26kDa	HID; KID; PPK; CX26; DFNA3; DFNB1; NSRD1; DFNA3A; DFNB1A; GJB2	121011	NG_008358.1	13q11-q12
GCDH	2639	glutaryl-Coenzyme A dehydrogenase	GCD; ACAD5; GCDH	608801	NG_009292.1	19p13.2
G6PC	2538	glucose-6-phosphatase, catalytic subunit	G6PT; GSD1; GSD1a; MGC163350; G6PC	232200	NG_011808.1	17q21
HBB	3043	hemoglobin, beta	CD113t-C; beta-globin; HBB	141900	NG_000007.3	11p15.5

Petitioner Illumina, Inc., EX1004, p. 165

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
BCKDHA	593	branched chain keto acid dehydrogenase E1, alpha polypeptide	MSU; MSUD1; OVD1A; BCKDE1A; FLJ45695; BCKDHA	608348	NC_000019.8	19q13.1-q13.2
BCKDHB	594	branched chain keto acid dehydrogenase E1, beta polypeptide	E1B; FLJ17880; dJ279A18.1; BCKDHB	248611	NG_009775.1	6q13-q15
DBT	1629	dihydrolipoamide branched chain transacylase E2	E2; E2B; BCATE2; MGC9061; DBT	248610	NG_011852.1	1p31
DLD	1738	dihydrolipoamide dehydrogenase	E3; LAD; DLDH; GCSL; PHE3; DLD	238331	NG_008045.1	7q31-q32
ACADM	34	acyl-Coenzyme A dehydrogenase, C-4 to C-12 straight chain	MCAD; ACAD1; MCADH; FLJ18227; FLJ93013; FLJ99884; ACADM	607008	NG_007045.1	1p31
MMAA	166785	methylmalonic aciduria (cobalamin deficiency) cblA type	cblA; MGC120010; MGC120011; MGC120012; MGC120013; MMAA	607481	NG_007536.1	4q31.22
MMAB	326625	methylmalonic aciduria (cobalamin deficiency) cblB type	ATR; cblB; MGC20496; MMAB	607568	NG_007096.1	12q24
MUT	4594	methylmalonyl Coenzyme A mutase	MCM; MUT	609058	NG_007100.1	6p12.3
MCOLN1	57192	mucolipin 1	ML4; MLIV; MST080; TRPML1; MSTP080; TRP-ML1; TRPM-L1; MCOLN1	605248	NC_000019.8	19p13.3-p13.2
ACTA1	58	actin, alpha 1, skeletal muscle	ACTA; ASMA; CFTD; MPFD; NEM1; NEM2; NEM3; CFTD1; CFTDM; ACTA1	102610	NG_006672.1	1q42.13
TPM3	7170	tropomyosin 3	TM3; TRK; NEM1; TM-5; TM30; TM30nm; TPMsk3; hscp30; MGC3261; FLJ41118; MGC14582; MGC72094; OK/SW-cl.5; TPM3	191030	NG_008621.1	1q21.2
TNNT1	7138	troponin T type 1 (skeletal, slow)	ANM; TNT; STNT; TNTS; FLJ98147; MGC104241; TNNT1	191041	NG_011829.1	19q13.4
NEB	4703	nebulin	NEM2; NEB177D; FLJ11505; FLJ36536; FLJ39568; FLJ39584; DKFZp686C1456; NEB	161650	NG_009382.1	2q22
SMPD1	6609	sphingomyelin phosphodiesterase 1, acid lysosomal	ASM; NPD; SMPD1	607608	NG_011780.1	11p15.4-p15.1
GLDC	2731	glycine dehydrogenase (decarboxylating)	GCE; NKH; GCSP; HYGN1; MGC138198; MGC138200; GLDC	238300	NC_000009.10	9p22

Petitioner Illumina, Inc., EX1004, p. 166

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
GCSH	2653	glycine cleavage system protein H (aminomethyl carrier)	GCE; NKH; GCSH	238330	NC_000016.8	16q23.2
AMT	275	aminomethyltransferase	GCE; NKH; GCST; AMT	238310	NC_000003.10	3p21.2-p21.1
OTC	5009	ornithine carbamoyltransferase	OCTD; MGC129967; MGC129968; MGC138856; OTC	300461	NG_008471.1	Xp21.1
PAH	5053	phenylalanine hydroxylase	PH; PKU; PKU1; PAH	612349	NG_008690.1	12q22-q24.2
DHPR	5860	quinoid dihydropteridine reductase	DHPR; PKU2; SDR33C1; FLJ42391; QDPR	612676	NG_008763.1	4p15.31
PTS	5805	6-pyruvoyltetrahydropterin synthase	PTPS; FLJ97081; PTS	261640	NG_008743.1	11q22.3-q23.3
PCCA	5095	propionyl Coenzyme A carboxylase, alpha polypeptide	PCCA	232000	NG_008768.1	13q32
PCCB	5096	propionyl Coenzyme A carboxylase, beta polypeptide	DKFZp451E113; PCCB	232050	NG_008939.1	3q21-q22
ACADS	35	acyl-Coenzyme A dehydrogenase, C-2 to C-3 short chain	SCAD; ACAD3; ACADS	606885	NG_007991.1	12q22-qter
DHCR7	1717	7-dehydrocholesterol reductase	SLOS; DHCR7	602858	NC_000011.8	11q13.2-q13.5
SMN1	6606	survival of motor neuron 1, telomeric	SMA; SMN; SMA1; SMA2; SMA3; SMA4; SMA@; SMNT; BCD541; T-BCD541; SMN1	600354	NG_008691.1	5q13
HEXA	3073	hexosaminidase A (alpha polypeptide)	TSD; MGC99608; HEXA	606869	NG_009017.1	15q23-q24
MYO7A	4647	myosin VIIA	DFNB2; MYU7A; NSRD2; USH1B; DFNA11; MYOVIIA; MYO7A	276903	NG_009086.1	11q13.5
USH1C	10083	Usher syndrome 1C (autosomal recessive, severe)	PDZ73; AIE-75; DFNB18; PDZ-45; PDZ-73; NY-CO-37; NY-CO-38; ush1cpst; PDZ-73/NY-CO-38; USH1C	605242	NC_000011.8	11p15.1-p14
CDH23	64072	cadherin-like 23	USH1D; DFNB12; FLJ00233; FLJ36499; KIAA1774; KIAA1812; MGC102761; DKFZp434P2350; CDH23	605516	NG_008835.1	10q21-q22
PCDH15	65217	protocadherin 15	USH1F; DFNB23; DKFZp667A1711; PCDH15	605514	NG_009191.1	10q21.1

Petitioner Illumina, Inc., EX1004, p. 167

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
SANS	124590	Usher syndrome 1G (autosomal recessive)	SANS; ANKS4A; FLJ33924; USH1G	607696	NG_007882.1	17q25.1
ARX	170302	aristaless related homeobox	ISSX; PRTS; MRX29; MRX32; MRX33; MRX36; MRX38; MRX43; MRX54; MRX76; MRX87; MRXS1; ARX	300382	NG_008281.1	Xp21
OPHN1	4983	oligophrenin 1	OPN1; MRX60; OPHN1	300127	NG_008960.1	Xq12
JARID1C	8242	lysine (K)-specific demethylase 5C	MRXJ; SMCX; MRXSJ; XE169; JARID1C; DXS1272E; KDM5C	314690	NG_008085.1	Xp11.22-p11.21
FTSJ1	24140	FtsJ homolog 1 (E. coli)	JM23; MRX9; SPB1; TRM7; CDLIV; MRX44; FTSJ1	300499	NG_008879.1	Xp11.23
SLC6A8	6535	solute carrier family 6 (neurotransmitter transporter, creatine), member 8	CRT; CT1; CRTR; MGC87396; SLC6A8	300036	NC_000023.9	Xq28
DLG3	1741	discs, large homolog 3 (Drosophila)	MRX; MRX90; NEDLG; NE-Dlg; SAP102; SAP-102; KIAA1232; DLG3	300189	NC_000023.9	Xq13.1
TM4SF2	7102	tetraspanin 7	A15; MXS1; CD231; MRX58; CCG-B7; TM4SF2; TALLA-1; TM4SF2b; DXS1692E; TSPAN7	300096	NG_009160.1	Xp11.4
ZNF41	7592	zinc finger protein 41	MRX89; MGC8941; ZNF41	314995	NG_008238.1	Xp11.23
FACL4	2182	acyl-CoA synthetase long-chain family member 4	ACS4; FACL4; LACS4; MRX63; MRX68; ACSL4	300157	NG_008053.1	Xq22.3-q23
PQBP1	10084	polyglutamine binding protein 1	SHS; MRX55; MRXS3; MRXS8; NPW38; RENS1; PQBP1	300463	NC_000023.9	Xp11.23
PEX1	5189	peroxisomal biogenesis factor 1	ZWS1; PEX1	602136	NG_008341.1	7q21.2
PXMP3	5828	peroxisomal membrane protein 3, 35kDa	PAF1; PEX2; PMP3; PAF-1; PMP35; RNF72; PXMP3	170993	NG_008371.1	8q21.1
PEX6	5190	peroxisomal biogenesis factor 6	PAF2; PAF-2; PXAAA1; PEX6	601498	NG_008370.1	6p21.1
PEX10	5192	peroxisomal biogenesis factor 10	NALD; RNF69; MGC1998; PEX10	602859	NG_008342.1	1p36.32
PEX12	5193	peroxisomal biogenesis factor 12	PAF-3; PEX12	601758	NG_008447.1	17q12
PEX5	5830	peroxisomal biogenesis factor 5	PXR1; PTS1R; PTS1-BP; FLJ50634; FLJ50721; FLJ51948; PEX5	600414	NG_008448.1	12p13.31
PEX26	55670	peroxisomal biogenesis factor 26	FLJ20695; PEX26M1T; Pex26pM1T; PEX26	608666	NG_008339.1	22q11.21

Petitioner Illumina, Inc., EX1004, p. 168

The mutations listed in Table 2 are documented polymorphisms in several disease-associated genes (CFTR is mutated in cystic fibrosis, GBA is mutated in Gaucher disease, ASPA is mutated in Canavan disease, HEXA is mutated in Tay Sachs disease). The polymorphisms are of several types: insertion/deletion polymorphisms which will cause frameshifts (and thus generally interrupt protein function) unless the insertion/deletion length is a multiple of 3bp, and substitutions which can alter the amino acid sequence of the protein and in some cases cause complete inactivation by introduction of a stop codon.

Table 2: Non-limiting examples of polymorphisms

Gene name	GeneID	SNP ID	Mutation	SEQ ID NO:
CFTR	1080	rs63500661	TCACATCACCAAGTTAAAAAAAAAAA[A/G]GGGGCGGGGGGGCAGAATGAAAATT	1
CFTR	1080	rs63107760	AAACAAGGATGAATTAAGTTTTTTTT[-/T]AAAAAAGAAACATTTGGTAAGGGGA	2
CFTR	1080	rs62469443	ATCACCAAGTTAAAAAAAAAAAAGGG[A/G]CGGGGGGGCAGAATGAAAATTGCAT	3
CFTR	1080	rs62469442	CTATTGAACCAGAACCAAACAGGAAT[A/G]CCATAGCATTTTGTAAACTAAACTG	4
CFTR	1080	rs62469441	CAGGAGTTCAAGACCAGCCTACTAAA[A/C]CACACACACACACACACACACACAC	5
CFTR	1080	rs62469439	GATTAAATAATAGTGTTTATGTACCC[C/G]GCTTATAGGAGAAGAGGGTGTGTGT	6
CFTR	1080	rs62469438	ATTGTTATCTTTTCATATAAGGTAAC[A/T]GAGGCCCAGAGAGATTAAATAACAT	7
CFTR	1080	rs62469437	TAATTTTAATTAAGTAAATTTAATTG[A/G]TAGATAAATAAGTAGATAAAAAATA	8
CFTR	1080	rs62469436	GTATAAAAAAAAAAAAAAAAAAAGTT[A/T]GAATGTTTTCTTGCATTCAGAGCCT	9
CFTR	1080	rs62469435	ATACTAAAAATTTAAAGTTCTCTTGC[A/G]ATATATTTTCTTAATATCTTACATC	10
CFTR	1080	rs62469434	TGCTGGGATTACAGGCGTGAGCCACC[A/G]CGCCTGGCCTGATGGGACATATTTT	11
CFTR	1080	rs62469433	CTACAATATAAGTATAGTATTGCAAA[A/C]CCATCAGGAAGGGTGTTAACTATTT	12
CFTR	1080	rs61763210	GTTGTCTCCAAACTTTTTTTCAGGTG[-/AGA]AGGTGGCCAACCGAGCTTCGGAAAG	13
CFTR	1080	rs61720488	TTTTTTCATAAAAGATTATATAAAGG[A/C]TATTGCTTTTGAATCACAAACACTA	14
CFTR	1080	rs61481156	ATCTAGTGAGCAGTCAGGAAAGAGAA[C/T]TTCCAGATCCTGGAAATCAGGGTTA	15
CFTR	1080	rs61443875	TAGAGTATAAAAAAAAAAAAAAAAA[-/A]GTTTGAATGTTTTCTTGCATTCAGA	16

Petitioner Illumina, Inc., EX1004, p. 169

CFTR	1080	rs61312222	TGCAAATGCCAACTATCAAAGATATT[C/G]GA GTATACTGTCAATAAACTTCATA	17
CFTR	1080	rs61159372	TCCTCAACAGTTAGAAACAATATTTT[C/G]AG TGATTTCCCATGCCAACTTTACT	18
CFTR	1080	rs61094145	TTTTTGGTATTGTTGTTAAATAAGTG[A/G]GA ATTCAATACAGTATAATGTCTGT	19
CFTR	1080	rs61086387	CTTGAAATCGGATATATATATATATA[- /TGTATATATATATATATATATATATATAT ACATATATATATATA]GTATTATCCCTGTTTTC ACAGTTTT	20
CFTR	1080	rs60996744	AGAGGGGCTGTGAAGGACACCAAGGA[A/G]G AGACTAAGAGCCAGGAGGGAAAAC	21
CFTR	1080	rs60960860	TAGAGTTTATTAGCTTTTACTACTCT[A/G]CTT AGTTACTTTGTGTTACAGAATA	22
CFTR	1080	rs60923902	ACTAGTGATGATGAGCTTCTTTTCAT[- /AT]GTTTGTTGGCTGCATAAATGTCTTC	23
CFTR	1080	rs60912824	GCAGAGAAAAGAGGGGCTGTGAAGGA[C/G]A CCAAGGAGGAGACTAAGAGCCAGG	24
CFTR	1080	rs60887846	TTCAGAGGTCTACCACTGGTGCATAC[G/T]CT AATCACAGTGTCGAAAATTTTAC	25
CFTR	1080	rs60793174	AAGAAAGAGCAAAAGAGGGCAAACTT[C/T]T CATACATTTTTGATGTCGAAACCA	26
CFTR	1080	rs60788575	CCTAAAGTTTAAAAAGAAAAAAAAAA[- /A]GGAAGAAGGAATTAAAAATCCAAAG	27
CFTR	1080	rs60760741	GTGTGTGTGTGTATATATATATATAT[A/T]TA TATATTTTTTTTTTCCTGAGCCA	28
CFTR	1080	rs60456599	AAACTGTTGATGTTTTCATTTATTTA[C/G]ATC ATTGGAAAACTTTAGATTCTAG	29
CFTR	1080	rs60363249	TTTATCCATTCTTAACCAGAACAGAC[A/G]TT TTTTCAGAGCTGGTCCAGGAAAA	30
CFTR	1080	rs60355115	TTGAAATCGGATATATATATATATAT[A/G]TA TATATATATATATATATATATAT	31
CFTR	1080	rs60308689	TAGTTTTTTATTTCCTCATATTATTT[- /T]CAGTGGCTTTTTCTTCCACATCTTT	32
CFTR	1080	rs60271242	ACATAGTTCTCAGTGGTACAACTACA[A/G]GT GATTTCTCTTTTCTTATTTCTGG	33
CFTR	1080	rs60010318	AGAGCAATGGCATCCCTTGTCTTGTG[C/T]TA TACAGGATGCAGCAATTTATAGG	34
CFTR	1080	rs59961323	TTCTGTCTACATAAGATGTCATACTA[A/G]AT TATCTTTTCCAGCATGCATTCAG	35
CFTR	1080	rs59961270	CAGGGTGGCATGTTAGGCAGTGCTTA[A/G]A ATAAATGAGTTGGTTATACAAGTA	36
CFTR	1080	rs59837506	AGGACACACACACACACACACACACA[- /CA]TGCACACACATTTAAATAGATGCAT	37
CFTR	1080	rs59572090	TAAAAAATTGGTATAATGAAATTGCA[C/T]TT GTAGTCTTTGGACATTTAAATCC	38
CFTR	1080	rs59548252	TTTCAATACTTAAGAGGTACGCAGAG[A/G]A AAGAGGGGCTGTGAAGGACACCAA	39
CFTR	1080	rs59519859	CAGCAATGAATATTTTGAGGCTGAGG[C/T]GC TGAGGGGTAAAATTGCAGCCTGG	40
CFTR	1080	rs59509837	TTATGGTTTATATTTTTGTGTCTTCT[- /CTTT]AACACATCTTTTCTAGCAGAATTCA	41

Petitioner Illumina, Inc., EX1004, p. 170

CFTR	1080	rs59417037	GTATTTTAGTTTTTTTTTTTTGTTTG[-/T]TTTGTTTTGTTTTGTTTTGTTTTTG	42
CFTR	1080	rs59159458	TGGGTGACTCCATTTTTACTTTTAGT[C/T]TGG TCTGTTGAGGCCTCGTGAGAGA	43
CFTR	1080	rs59048119	TATTTTCATGTATTTTAGTTTTTTTTT[-/TTTT]GTTTGTTTTGTTTTGTTTTGTTTTG	44
CFTR	1080	rs58970500	GTGTGTGTGTATATATATATATATAT[A/T]TA TATTTTTTTTTTCCTGAGCCAAA	45
CFTR	1080	rs58942292	AACCTATTAGCATGTCTGGCAGAAAA[-/A]TAGATACTTAATAAATTTCTTAAAT	46
CFTR	1080	rs58917054	GAGGCTTAGACAGTTTAAGTAACTCA[A/G]G CATGGTTACACAACTAGCTAGGGC	47
CFTR	1080	rs58837484	GTGTGAGTATTATGAGACCATATGTT[A/G]GG AGATTTTATTTGGTATTGAGGAT	48
CFTR	1080	rs58829491	GAAACCCCACCCCTTCTATAGTTTTC[C/T]CTT TAATATTTACAATGGAACCATT	49
CFTR	1080	rs58805195	CATATATATATAGTGTGTGTGTGTGT[A/G]TA TATATATATATATATATATTTTT	50
GBA	2629	rs60866785	CGAGCGAGAGAGAGAGAGAGAGAGAG[-/AG]GAGCCGGCGCGAGAACTACGCATGC	51
GBA	2629	rs60239603	GGCAGGTAATATCTAGTACCTTACTT[A/T]TA TTTCCTGAGCACATTCTACATTT	52
GBA	2629	rs56310840	GGCCAGGAATGGGAGTGCTTAGGTGC[A/G]G AGGTGGCACTGTTCCCGCAGCTGC	53
GBA	2629	rs41264927	TGAAAACTCCATCCCCTCAGGGTCAT[C/T]AG ATGAAGAGAAGACCACAGGGGTT	54
GBA	2629	rs41264925	TGTAGGTAAGGGTCACATGTGGGAGA[C/G]G CAGCTGTGGGTAGGTCAGCCCTGT	55
GBA	2629	rs36024691	CCAAGAAGGCGCCATTACACTCCAGC[-/C]TGGGCGACAGGGCGAGACTCCCTCA	56
GBA	2629	rs36024092	TGCCACACCCAGCTAATTTGTGTGTG[-/G]TATGTGTGTGTATGTATGTGTGTGT	57
GBA	2629	rs35682967	GTTCCTCCAGTAATTTTTTTTTTTTT[-/T]GGTTTTGAGACAGAGTCTTGCCCTG	58
GBA	2629	rs35033592	ATCATGCCCAGATAATTTTTTTTTTT[-/T]GTATTTTAGTAGACACAGGGTTTCA	59
GBA	2629	rs34732744	CGAGCGAGAGAGAGAGAGAGAGAGAG[-/AG]GAGCCGGCGCGAGAACTACGCATGC	60
GBA	2629	rs34620635	CCTGTGAGGGGCACATTCCTTAGTAG[-/C]TAAGGAGTTGGGGGTGTGAAGATCC	61
GBA	2629	rs34302637	ACAGGCTACTGGCTGGGCCCAGGCAA[-/A]GGGGGCCTTGGCAGGAAAAGTTCCT	62
GBA	2629	rs33949225	GCGAGAGAGAGAGAGAGAGAGAGAGG[-/AG]AGCCGGCGCGAGAACTACGCATGCG	63
GBA	2629	rs28678003	AAGAAGAAAAATAAAAAGAAAGTGGG[C/T]C AGACCGAGAGAACAGGAAGCCTGA	64
GBA	2629	rs28559737	AAGGACAAAGGCAAAGAGACAAAGGC[G/T]C AACACTGGGGGTCCCCAGAGAGTG	65
GBA	2629	rs28373017	TACCTAGTCACTTCCTGCCTCCATGG[C/T]GC AAAAGGGGATGGGTGTGCCTCTT	66
GBA	2629	rs12752133	CTCTTCCGAGGTTCCACCCTGAACAC[C/T]TT CCTGCTCCCTCGTGGTGTAGAGT	67

Petitioner Illumina, Inc., EX1004, p. 171

GBA	2629	rs12747811	TTCTGACTGGCAACCAGCCCCACTCT[C/T]TG GGAGCCCTCAGGAATGAACTTGC	68
GBA	2629	rs12743554	gctcagcctcccaggctggagtgcag[A/T]ggcgcgatctcggctcac cgcaacc	69
GBA	2629	rs12041778	CATGAACCACATCAAATGAGATTTAG[C/T]GG GAGTGGCACACACAGTCATGACC	70
GBA	2629	rs12034326	AAGCAGCCCTGGGGAGTCGGGGCGGG[A/G]C CTGGATTGGAAAAGAGACGGTCAC	71
GBA	2629	rs11558184	CTCCAAGTTCTGGGAGCAGAGTGTGC[A/G]G CTAGGCTCCTGGGATCGAGGGATG	72
GBA	2629	rs11430678	GTTCCTCCAGTAAttttttttttt[-/G/T]gttttgagacagagtcttgccctgt	73
GBA	2629	rs11264345	CTAGTACCTTACTTCCCTCAAGTTCA[A/T]TC ATCTCACAGATATTTCCTGAGCA	74
GBA	2629	rs10908459	aattagccgtgcgtggtggcgggtgc[C/T]tgtaatcccacgtacttgg gaggct	75
GBA	2629	rs10796940	CCATGGCCAGCCGGGGAGGGGACGGG[A/C]A CACACAGACCCACACAGAGACTCA	76
GBA	2629	rs10668496	agcgagagagagagagagagagag[-/AG]gagCCGGCGCGAGAACTACGCATGC	77
GBA	2629	rs7416991	CGTAGCAGTTAGCAGATGATAGGCGG[C/G/T] GAAATCTTATTTCACAGGGCATTAA	78
GBA	2629	rs4024049	CTGGCCCTGGTGACAGTGGGGCTGTG[C/T]GT GGGGCCAGAGCCTTCTCAGAGGT	79
GBA	2629	rs4024048	CAGATACTGGCCCTGGTGACAGTGGG[A/G]C TGTGCGTGGGGCCAGAGCCTTCTC	80
GBA	2629	rs4024047	GACAGATACTGGCCCTGGTGACAGTG[G/T]G GCTGTGCGTGGGGCCAGAGCCTTC	81
GBA	2629	rs3841430	GGCTCctctctctctctctctctc[-/TC]gctcgctctctcgctctctcgctct	82
GBA	2629	rs3754485	GTTTCAGACCAGCCTGGCCAACATAG[C/T]GA AACCCCATCTCTACTAAAAATAA	83
GBA	2629	rs3205619	AGTGGGCGATTGGATGGAGCTGAGTA[C/T]G GGGCCCATCCAGGCTAATCACACC	84
GBA	2629	rs2990227	CCGGGCTCCGTGAATGTTTGTCACAT[C/G]TC TGAAGAACGTATGAATTACATAA	85
GBA	2629	rs2990226	GAATCCCAACCCCGACGCTCGTCGCC[C/G]G GCTCCGTGAATGTTTGTCACATGT	86
GBA	2629	rs2990225	GCGAATCCCAACCCCGACGCTCGTCG[C/T]CG GGCTCCGTGAATGTTTGTCACAT	87
GBA	2629	rs2990224	TGGGCAGAAGTCAGGGTCCAAAGAAA[G/T]G GCAAAGAAAAGTGTcagtggctca	88
ASPA	443	rs63751297	TAAGAAAGACGTTTTTGATTTTTTTC[A/G]GA CTTCTCTGGCTCCACTACCCTGC	89
ASPA	443	rs62071301	CTGATTCCTGGCCAGGAGCGGTGGCT[C/T]AC GCCTGTAATCCCAGCGCTTTGGG	90
ASPA	443	rs62071300	TAAAAATGCTGATTCCTGGCCAGGAG[C/T]GG TGGCTCACGCCTGTAATCCCAGC	91
ASPA	443	rs62071299	TTTAAAAATGCTGATTCCTGGCCAGG[A/C]GC GGTGGCTCACGCCTGTAATCCCA	92
ASPA	443	rs62071297	CAAGACCTGTCAAAGATCTGAGAAAT[A/T]TT ACCCGACTTACAAGCTAACCATT	93

Petitioner Illumina, Inc., EX1004, p. 172

ASPA	443	rs61697033	ACTGTAATAAGTGCTGTAAAAGAAAT[A/G]C ACAAAATAATATAGCAGAGGGTAT	94
ASPA	443	rs60743592	CTTGAGGTCAGGAGTTCAAGACCAGT[C/T]TG GGCAACATGGGGAAAACCTTGTC	95
ASPA	443	rs60666840	AGGTTGCAGTGAGCCGAGATCATGCC[A/G]TT GCACTCCAGCCGGGGCAACAAAA	96
ASPA	443	rs60147514	ACAAGTGTCTTGAAATTATCTGTGAT[C/T]TG CTATAGAGCAATACTTTTGTAAA	97
ASPA	443	rs59930743	GTGGGTATATGCAGCTCTATGCACTA[C/T]CT GCTCATTTATTTGGTAAATCTAA	98
ASPA	443	rs59690349	TGTGTGTGTGTGCGTGTGTGTGTGTG[-/TGTGTGTG]ATCATAAGAGTGGCTGCAGCAA ACT	99
ASPA	443	rs59676360	AGTCTGGAGTGCAATGGTGCAATCTC[A/G]GC TCACTGCAGCCTCCACCTCCGGG	100
ASPA	443	rs59335404	CTCCTAATGGATATTTCCTAAATTTT[G/T]CTG AACAGAATTTAACTTGAGCTGG	101
ASPA	443	rs58879097	ATTTAAAAATGGATTTCTAGAAAAAC[A/G]AT CACATACTTGAATATTTTAGCAA	102
ASPA	443	rs58686774	CTATAAATGGGTAGCATGAGGGATTC[A/G]A GGAGGTGGCTGAAAGAAGCACGTA	103
ASPA	443	rs57511162	AAGAAACCAAGCATAGTAGAGTGTTA[A/G]A AAACCAAAGCAACTAAACAACTGT	104
ASPA	443	rs55859596	CGGGGCTCAGAACTTGTAACAGAAAA[A/T]T AAAATATACTCCACTCAAGGGAAT	105
ASPA	443	rs55742972	TACTACACTTCACGGATACTGTACTT[-/GTACTT]TTTTTCCAAATTGAAGGTTTTTGGC	106
ASPA	443	rs55640436	TTGTTTTTGTTTTTGTTTTTGTTTTT[-/GTTTTTGTTTTT]TGAGATGGAGTCTCGCTCT GTCGCC	107
ASPA	443	rs36225687	TTTGCCTTACTACACTTCACGGATAC[-/TGTACT]TGTACTTTTTTTCCAAATTGAAGGT	108
ASPA	443	rs36051310	GAGGTGGCTGAAAGAAGCACGTATCC[-/C]TGATGGCATGGTTGCGGGTTATATG	109
ASPA	443	rs36034906	GAGAAAAGCAGTTCCTGGAACACCCC[-/C]ACCCCTTAACCCCTTATCTCTGCTT	110
ASPA	443	rs36033666	TTACATATGTATACATGTGCCATGTT[-/T]GGTGTGCCGCACCCATTAACTCGTC	111
ASPA	443	rs35730123	CTTTTTCCAGATTTTTTTTTTTTTTT[-/T]GAGACAGAGTTTCACTCTTGTTGCC	112
ASPA	443	rs35629100	TTTGGAAATCTTAAGCTTTTATTTGG[-/G]TGTCACAGAGAAACAGGATCTGTAT	113
ASPA	443	rs35614631	TACTTTAAGTTTTAGGGTACATGTGC[-/A]CCATGTGCAGGTTTGTTACATATGT	114
ASPA	443	rs35225782	ATTCATGACCAGCCACATAAATGCAC[-/A]GTATTACTTCGCAAGCATGCCAATG	115
ASPA	443	rs35178659	GTGCACTAGAATTAGCTAAAGTGGGG[-/G]AAAAAAAGATGCATTTGATGGTCTA	116
ASPA	443	rs35095578	AACCTCCACCTCCCAGGTTCAAGAGA[-/A]TTCTCCTGCCTCAGCCTCCCAAGTA	117
ASPA	443	rs35002210	CCTCCCTGTGATCCGAAGTAGCAGAC[A/G]TA CTTAACTTCCATGGTGGATTGTT	118

Petitioner Illumina, Inc., EX1004, p. 173

ASPA	443	rs34744839	AAAACATTATTATATCTAGAAAAAAA[-/A]TGTATCTTAACCATTGTGGGAAGTG	119
ASPA	443	rs34680506	TTGAAGGTAAAATCATAGGGAGTTGG[-/G]AGCTGTCCTCTTGCGCTGAATCAGT	120
ASPA	443	rs34365618	ACTTGTGGCCTTTTTGGAGAGGTTAG[-/CA]ACTCTGAAAACTCTGTCCCTGGACC	121
ASPA	443	rs34275920	GAAGGAGAAAAAGAGAGGAAATAAGT[-/T]AAAATAATAAACACAATTAATAAAG	122
ASPA	443	rs34109510	TGTATACATGTGCCATGTTGGTGTGC[C/T]GC ACCCATTAACTCGTCATTTAGCA	123
ASPA	443	rs34054576	TCACCTGTCACCTCCTATAGAACTTT[-/C]CCCTGACCCTCCTCTATAGCATTAA	124
ASPA	443	rs34015272	ATAAATGATCATCATTCACAGTAGGG[-/G]TTTTGTTTTGTTTTTTTTCTGGAAA	125
ASPA	443	rs34002091	ACAGACATATCTACAAACACACTTTT[-/T]CACATATTTGTGTAAGTCATTTATG	126
ASPA	443	rs28940574	AAAGACAACTAAACTAACGCTCAATG[A/C]A AAAAGTATTCGCTGCTGTTTACAT	127
ASPA	443	rs28940279	TACCGTGTACCCCGTGTTTGTGAATG[A/C]GG CCGCATATTACGAAAAGAAAGAA	128
ASPA	443	rs17850703	CAGGGCTGGAGGTAAAACCATTTATT[A/G]CT AACCCCAGAGCAGTGAAGAAGTG	129
ASPA	443	rs17222495	TTCTTCATTGCCTATTGAAGAGAGAG[C/T]GG AATGCTTTGGTTGCCAGATATGG	130
ASPA	443	rs17175228	CACAAGATCTCATTACTCAGGAGCTG[C/T]CC AAGTGTCTAATGTACTTAGTTAA	131
ASPA	443	rs16953074	TTCTGTGTAACATTTCATTTAAGCAA[A/G]GG ATTCGGCAAATCAAAAATTGTCA	132
ASPA	443	rs16953070	TAAAACGTATTGAAGGTATTATTGAC[G/T]CT GTTGAAGCAAAGAGAACAAAACA	133
HEXA	3073	rs62022858	ATCTGCTCTTCCAGTTGGATGACAAG[C/T]CT TGCTGTCTAACACCTGCTGCAGA	134
HEXA	3073	rs62022857	CCATTTTTTGTTGTATTTTTTTTTTC[C/T]TGAA TACTTTTTATCGCAGTTGGTT	135
HEXA	3073	rs62017872	CCCTGTCTCTAAAAGAAAAAAAAAAA[A/G]A AAAAAAAAAGAAAACAAAACCCAA	136
HEXA	3073	rs62017871	AGTGGCTCCAAAAAGGTCATGGAACC[C/T]CT TGAGGATGATGCAAATTGACTCT	137
HEXA	3073	rs61662730	TAAAGTTACTTTTCTTTTATTGACTT[C/T]CCC TTATTTTTTAACCTTATGCTTT	138
HEXA	3073	rs61329913	CAGAGTTAAAAAAAAAAAAAAAAAA[-/A]GGAAGTAGCAGCAACAGCTTGGAAA	139
HEXA	3073	rs60920713	GTTGCCCAGGGTTGAGTGCAGAGGCA[C/T]AT CTTGGCTCACAGCAACCTCTGCC	140
HEXA	3073	rs60783213	AAGGCTTTTTTTTTTTTTTTTTTTTT[-/TTTT]GAGACAGAGTCTTGCTGTGTCACCC	141
HEXA	3073	rs60644867	GCCTACATTCTGCAAAGAGGAGGGAA[C/G]A TTCACAGCTCCATACTTGAACCCT	142
HEXA	3073	rs60288568	CCAAAGGAGAATAGCTCTAGGGGAGG[C/G]A GGTGGATGAGTATGCATGGGGGAG	143
HEXA	3073	rs59888548	GACTCCATCTCAAAAAAAAAAAAAAA[-/A]TGCAGTCTAATGGCAGAATTAGACT	144

Petitioner Illumina, Inc., EX1004, p. 174

HEXA	3073	rs59733856	TTATTTATTTATTTATTTATTTTTGA[A/G]ACA GGGTCTCTGTTGTCCAGGCTGG	145
HEXA	3073	rs59427837	TTTTGAGGCAGGGTCTCACTCTGTTG[C/T]CC AGGGTTGAGTGCAGAGGCACATC	146
HEXA	3073	rs59171976	CGCCTTGCGAAGGCCCCACAGCTTGC[C/T]TG TGACAAACGTTCATAGGCAAATG	147
HEXA	3073	rs58706602	GGAGGTCTGTACAAAGCACCACCTAC[C/T]TC ATGGGTCAGTTTCCACAGCAGAA	148
HEXA	3073	rs58696963	GAATCTTATAATTCACTGTGTACCTC[- /CCTC]TGTTTCATATTTTCGCAATTGAACT	149
HEXA	3073	rs58610850	AACATAGTATCTAATATAGCTTTACA[C/T]CC AAAGCCAAAATATGAATACACTG	150
HEXA	3073	rs58016062	TTGTTTTGTTTTGTTTGGGGGGGGG[- /G]TTGTTTTTCTGAGAGGGAGTCTTGC	151
HEXA	3073	rs57733983	CATACCAAAGGGCAGCTGGAGGGATA[C/T]A GACGGAAGTCATGTGGAGAGTGAA	152
HEXA	3073	rs57476645	CAGGTGTGAGCCACCACGACCACCAA[A/T]T TAGCTCTTTTTACTCCTTCCCTTC	153
HEXA	3073	rs56870003	AGTGGTAGCTGATTTTGCTTCTGGAT[A/C]CT TTGCCACCTTCCCACTCTTTAAT	154
HEXA	3073	rs56338339	AAAGACCTGTTTCTTAAAAAAAAAAA[- /AGAAAAAAAAAAA]GAAAGAAAAGAAAAG AAAAAAACAG	155
HEXA	3073	rs55995352	TAAAAAATCTTTCAATGAGGAGATGT[C/T]CC CAGAGCAAGACAGCTGTAGGATG	156
HEXA	3073	rs55860138	AAAAGAAAAAAAAAAAAAAAAAAAA[- /A]GAAAACAAAACCCAAACCCATAAAG	157
HEXA	3073	rs55743646	CCTGTCTCTAAAAGAAAAAAAAAAAA[A/G]A AAAAAAAAGAAAACAAAACCCAAA	158
HEXA	3073	rs55665666	GTTATCATAGAAAAATATCACACTCT[- /GT]CTGTATCCCCACTTCCAGAAACTGT	159
HEXA	3073	rs36106892	CAGGAGCTCATAGAATTACATACAAT[- /C]TTTTTTTTTTTTTTTTTGAGACAGCG	160
HEXA	3073	rs36091525	TTGAGAATCTTATAATTCACTGTGTA[- /CCTC]CCTCTGTTTCATATTTTCGCAATTG	161
HEXA	3073	rs35949555	CCACTACCACAGTGCCTAGAGAACAA[C/T]A TGTGTTTAATAATATTTAAATAAT	162
HEXA	3073	rs35827424	CCCTGTCTCTAAAAGAAAAAAAAAAA[- /A]AAAAAAAAAAGAAAACAAAACCCAA	163
HEXA	3073	rs35729578	CCATTATATCATTCATTTCCCACTCA[- /T]TTTCTTCATTCCAACCAAGATATAT	164
HEXA	3073	rs35649102	TCCGTCTCAAAAAAAAAAAAAAAAAG[- /A]GAAAGGAATTATTCTCATGTATACA	165
HEXA	3073	rs35118677	CTGGGGCAGTTAAAAAGAAAAACAAA[- /C]CCCTGGTCCCTGCCCTTGAGGAGAT	166
HEXA	3073	rs35005352	CTCCAGGGTCCCATTCCAGGACCACA[- /C]GCCTGCTACCTCTGCAGCTCACTCA	167
HEXA	3073	rs34736306	GGATTGACATATACCAGTTAGACGGA[- /T]TTTTTTTTTCCATAAACCAGGCTCA	168
HEXA	3073	rs34607939	ACAAATAATTACTACATATCTACAAC[A/G]TT CCAGATACAGAAGAAATGGCCAA	169

Petitioner Illumina, Inc., EX1004, p. 175

HEXA	3073	rs34496117	TAAACACACTTGAAACATCATATAAA[-/ATG]ATATTACTACAAGACTTAACCGTAA	170
HEXA	3073	rs34300017	ACACAGGTAATCCATGTTTATTATAG[-/A]AAAATGCCACATTACTCTTTATTGA	171
HEXA	3073	rs34206496	AGTTATCATAGAAAAATATCACACTC[-/TG]TCTGTATCCCCACTTCCAGAAACTG	172
HEXA	3073	rs34110830	AATGAACTTACAGGAAGGTAATATAT[-/G]GGAAATAAACATCTTATTGAATTTA	173
HEXA	3073	rs34093438	GGACCCCTGAAAGGCACAAGACACCC[-/T]TTCAGGTTCACACTTCCTGAAAGCT	174
HEXA	3073	rs34085965	CCACCAATCACCAGAGCCTTCTGCTC[A/G]GGGGTACCTGAGGGAAAACAAGCAA	175
HEXA	3073	rs34004907	AAAGACTGAAAAAACATTCATAACTA[-/T]TTTTCTTGTTATCCTCGGAAATGTC	176
HEXA	3073	rs28942072	TATCTTCATCTTGGAGGAGATGAGGT[C/T]GATTTCACCTGCTGGAAGTCCAACC	177
HEXA	3073	rs28942071	TTGCCTATGAACGTTTGTCACACTTC[C/T]GCTGTGAGTTGCTGAGGCGAGGTGT	178
HEXA	3073	rs28941771	GCTTGCTGTTGGATACATCTCGCCAT[C/T]ACCTGCCACTCTCTAGCATCCTGGA	179
HEXA	3073	rs28941770	CCGGGGCTTGCTGTTGGATACATCTC[G/T]CCATTACCTGCCACTCTCTAGCATC	180

3. Nucleic acid target length evaluation:

In some embodiments, aspects of the invention relate to methods for detecting nucleic acid deletions or insertions in regions containing nucleic acid sequence repeats.

Genomic regions that contain nucleic acid sequence repeats are often the site of genetic instability due to the amplification or contraction of the number of sequence repeats (e.g., the insertion or deletion of one or more units of the repeated sequence). Instability in the length of genomic regions that contain high numbers of repeat sequences has been associated with a number of hereditary and non hereditary diseases and conditions.

For example, "Fragile X syndrome, or Martin-Bell syndrome, is a genetic syndrome which results in a spectrum of characteristic physical, intellectual, emotional and behavioral features which range from severe to mild in manifestation. The syndrome is associated with the expansion of a single trinucleotide gene sequence (CGG) on the X chromosome, and results in a failure to express the FMR-1 protein which is required for normal neural development. There are four generally accepted forms of Fragile X syndrome which relate to the length of the repeated CGG sequence; Normal (29-31 CGG

Petitioner Illumina, Inc., EX1004, p. 176

repeats), Premutation (55-200 CGG repeats), Full Mutation (more than 200 CGG repeats), and Intermediate or Gray Zone Alleles (40 - 60 repeats)."

Other examples include cancer, which has been associated with microsatellite instability (MSI) involving an increase or decrease in the genomic copy number of nucleic acid repeats at one or more microsatellite loci (e.g., BAT-25 and/or BAT-26). The are currently many sequencing-based assays for determining the number of nucleic acid sequence repeats at a particular locus and identifying the presence of nucleic acid insertions or deletions. However, such techniques are not useful in a high throughput multiplex analysis where the entire length of a region may not be sequenced.

In contrast, in some embodiments, aspects of the invention relate to detecting the presence of an insertion or deletion at a genomic locus without requiring the locus to be sequenced (or without requiring the entire locus to be sequenced). Aspects of the invention are particularly useful for detecting an insertion or deletion in a nucleic acid region that contains high levels of sequence repeats. The presence of sequence repeats at a genetic locus is often associated with relatively high levels of polymorphism in a population due to insertions or deletions of one or more of the sequence repeats at the locus. The polymorphisms can be associated with diseases or predisposition to diseases (e.g., certain polymorphic alleles are recessive alleles associated with a disease or condition). However, the presence of sequence repeats often complicates the analysis of a genetic locus and increases the risk of errors when using sequencing techniques to determine the precise sequence and number of repeats at that locus.

In some embodiments, aspects of the invention relate to determining the size of a genetic locus by evaluating the capture frequency of a portion of that locus suspected of containing an insertion or deletion (e.g., due to the presence of sequence repeats) using a nucleic acid capture technique (e.g., a nucleic acid sequence capture technique based on molecular inversion probe technology). According to aspects of the invention, a statistically significant difference in capture efficiency for a genetic locus of interest in different biological samples (e.g., from different subjects) is indicative of different relative lengths in those samples. It should be appreciated that the length differences may be at one or both alleles of the genetic locus. Accordingly, aspects of the invention may be used to identify polymorphisms regardless of whether biological samples being interrogated at heterozygous or homozygous for the polymorphisms. According to

Petitioner Illumina, Inc., EX1004, p. 177

aspects of the invention, subjects that contain one or more loci with an insertion or deletion can be identified by analyzing capture efficiencies for nucleic acids obtained from one or more biological samples using appropriate controls (e.g., capture efficiencies for known nucleic acid sizes, capture efficiencies for other regions that are not suspected of containing an insertion or deletion in the biological sample(s), or predetermined reference capture efficiencies, or any combination thereof. However, it should be appreciated that aspects of the invention are not limited by the nature or presence of the control. In some embodiments, if a statistically significant variation in capture efficiency is detected, a subject may be identified as being at risk for a disease or condition associated with insertions or deletions at that genetic locus. In some embodiments, the subject may be analyzed in greater detail in order to determine the precise nature of the insertion or deletion and whether the subject is heterozygous or homozygous for one or more insertions or deletions. For example, gel electrophoresis of an amplification (e.g., PCR) product of the locus, or Southern blotting, or any combination thereof can be used as an orthogonal approach to verify the length of the locus. In some embodiments, a more exhaustive and detailed sequence analysis of the locus can be performed to identify the number and types of insertions and deletions. However, other techniques may be used to further analyze a locus identified as having an abnormal length according to aspects of the invention.

Accordingly, aspects of the invention relate to detecting abnormal nucleic acid lengths in genomic regions of interest. In some embodiments, the invention aims to estimate the size of genomic regions that are hard to be accessed, such as repetitive elements. However, it should be appreciated that methods of the invention do not require that the precise length be estimated. In some embodiments, it is sufficient to determine that one or more alleles with abnormal lengths are present at a locus of interest (e.g., based on the detection of abnormal capture efficiencies).

In a non-limiting example, fragile X can be used to illustrate aspects of the invention where the size of trinucleotide repeats (genotype) is linked to a symptom (phenotype). However, it should be appreciated that fragile X is a non-limiting example and similar analyses may be performed for other genetic loci (e.g., independently or simultaneously in multiplex analyses).

Use of molecular inversion probes (MIPs) has been demonstrated for detection of single nucleotide polymorphisms (Hardenbol et al. 2005 Genome Res 15:269-75) and for

Petitioner Illumina, Inc., EX1004, p. 178

preparative amplification of large sets of exons (Porreca et al. 2007 Nat Methods 4:931-6, Krishnakumar et al. 2008 Proc Natl Acad Sci USA 105:9296-301). In both cases, oligonucleotide probes are designed which have ends ('targeting arms') that hybridize up-stream and down-stream of the locus that is to be amplified.

In some embodiments, aspects of the invention are based on the recognition that the effect of length on probe capturing efficiency can be used in the context of an assay (e.g., a high throughput and/or multiplex assay) to allow the length of sequences to be determined without requiring sequencing of the entire region being evaluated. This is particularly useful for repeat regions that are prone to changes in size. As illustrated in FIG. 8, which is reproduced from Deng et al., Nature Biotech. 27:353-60, (see Supplemental Fig. 1G of Deng et al.,) illustrates that shorter sequences are captured with higher efficiency that longer sequences using MIPs. The statistical package R and its *effects* module were used for this analysis. A linear model was used, and each individual factor was assumed to be independent. The dashed lines represent a 95% confidence interval. Shorter target sequences were captured with higher efficiency than long target sequences ($p < 2x10^{-16}$). However, the use of this differential capture efficiency for systematic sequence length analysis was not previously recognized.

In some embodiments, following probe hybridization, polymerase fill-in and ligation reactions are performed to convert the hybridized probe to a covalently-closed, circular molecule containing the desired target. PCR or rolling circle amplification plus exonuclease digestion of non-circularized material is performed to isolate and amplify the circular targets from the starting nucleic acid pool. Since one of the main benefits of the method is the potential for a high degree of multiplexing, generally thousands of targets are captured in a single reaction containing thousands of probes.

According to aspects of the invention, repetitive regions are surrounded by non-repetitive unique sequences, which can be used to amplify the repeat-containing regions using, for example, PCR or padlock (MIP)-based method.

In addition to the repetitive regions, a probe (e.g., a MIP or padlock probe) can be designed to include at least a sequence that is sufficient to be uniquely identified in the genome (or target pool). After the probe is circularized and amplified, the amplicon can be end-sequenced so that the unique sequence can be identified and served as the "representative" of the repetitive region as illustrated in FIG. 9. FIG. 9 illustrates a non-limiting scheme of padlock (MIP) capture of a region that includes both repetitive

Petitioner Illumina, Inc., EX1004, p. 179

regions (thick wavy line) and the adjacent unique sequence (thick strait line). The regions of the probe are indicated with the targeting arms shown as regions "1" and "3." An intervening region that may be, or include, a sequencing primer binding site is shown as "2." After the padlock is circularized and amplified, it can be end-sequenced to obtain the sequence of the unique sequence, which represents the repetitive region of interest. Although capturing efficiency is overall negatively correlated with target length, different probe sequences may have unique features. Therefore, multiple probes could be designed and tested so that an optimal one is chosen to be sensitive enough to differentiate repetitive sizes of roughly 0-150bp, 150-600bp, and beyond, which represent normal, premutation and full mutation of fragile X syndrome, respectively. However, it should be appreciated that other probe sizes and sequences can be designed, and optionally optimized, to distinguish a range of repeat region size differences (e.g., length differences of about 3-30 bases, about 30-60 bases, about 60-90 bases, about 90-120 bases, about 120-150 bases, about 150-300 bases, about 300-600 bases, about 600-900 bases, or any intermediate or longer length difference). It should be appreciated that a length difference may be an increase in size or a decrease in size.

In some embodiments, an initial determination of an unexpected capture frequency is indicative of the presence of size difference. In some embodiments, an increase in capture frequency is indicative of a deletion. In some embodiments, a decrease in capture frequency is indicative of an insertion. However, it should be appreciated that depending on specific sequence parameters and the relative sizes of the capture probes, the target region, and the deletions or insertions, a change in capture frequency can be associated with either an increase or decrease in target region length. In some embodiments, the precise nature of the change can be determined using one or more additional techniques as described herein.

Accordingly, in some aspects a MIP probe includes a linear nucleic acid strand that contains two hybridization sequences or targeting arms, one at each end of the linear probe, wherein each of the hybridization sequences is complementary to a separate sequence on a the same strand of a target nucleic acid, and wherein these sequences on the target nucleic acid flank the two ends of the target nucleic acid sequence of interest. It should be appreciated that upon hybridization, the two ends of the probe are inverted with respect to each other in the sense that both 5' and 3' ends of the probe hybridize to

Petitioner Illumina, Inc., EX1004, p. 180

the same strand to separate regions flanking the target region (as illustrated in FIG. 9 for example).

In some embodiments, the hybridization sequences are between about 10-100 nucleotides long, for example between about 10-30, about 30-60, about 60-90, or about 20, about 30, about 40, or about 50 nucleotides long. However, other lengths may be used depending on the application. In some embodiments, the hybridization Tms of both targeting arms of a probe are designed or selected to be similar. In some embodiments, the hybridization Tms of the targeting arms of a plurality of probes designed to capture different target regions are selected or designed to be similar so that they can be used together in a multiplex reaction. Accordingly, a typical size of a MIP probe prior to fill-in is about 60-80 nucleotides long. However, other sizes can be used depending on the sizes of the targeting arms and any other sequences (e.g., primer binding or tag sequences) that are present in the MIP probe. In some embodiments, MIP probes are designed to avoid sequence-dependent secondary structures. In some embodiments, MIP probes are designed such that the targeting arms do not overlap with known polymorphic regions. In some embodiments, targeting arms that can be used for capturing the repeat region of the Fragile X locus can have the following sequences or complementary to these sequences depending on the strand that is captured.

left: CTCCGTTTCGGTTTCACTTC (SEQ ID NO: 181)

right: ATCTTCTCTTCAGCCCTGCT (SEQ ID NO: 182)

The typical captured size using these targeting arms is about 100 nucleotides in length (e.g., about 30 repeats of a tri-nucleotide repeat).

In some embodiments, the number of reads obtained for the "representative" of the repetitive region is not informative to estimate the target length because it is dependent on the total number of reads obtained. To overcome this, it is useful to include one or more probes that target other "control" regions where no or minimal polymorphism exists among populations. Because of the systematic consistency of capturing efficiency (see, e.g., FIG. 9), the ratio of reads obtained for the repetitive "representative" to reads obtained for the control region(s) will be tuned using DNA with defined numbers of repeats. Ultimately, the ratio can serve as a measure of the repeat length as illustrated in FIG. 10. FIG. 10 illustrates a non-limiting hypothetical relationship between target gap size and the relative number of reads of the repetitive

Petitioner Illumina, Inc., EX1004, p. 181

region, which is measured by the ratio of the repeat "representative" reads vs. the "control" region reads. The unit of y-axis is arbitrary.

In some embodiments, to better tell targets with similar size range apart, the whole repetitive region can be sequenced by making a shotgun library (e.g., by making a shotgun library from a captured sequence, for example a sequence captured using a MIP probe). The longer the repeat is, the more short reads of repeats will be obtained. Therefore, the target length will contribute twice to the relative number of "repetitive" reads, which will gain better resolution of differentiating targets. In some embodiments, the expectation is that the number of reads from any given repeat will be a direct function of the number of repeats present. However, in some embodiments, a Poisson sampling-induced spread may need to be considered and in some embodiments may be sufficiently large to limit the resolution.

When a precise measurement of the length of both alleles from a diploid sample is desired, further manipulations may be required. This is because the capture efficiency measured will actually be the average efficiency of the two alleles. To effectively achieve separate measurements for each allele, barcodes (e.g., sequence tags) can be used that allow the efficiency of individual capture events (from individual genomic loci) to be followed. FIG. 11A-C shows the approach. For a given locus, MIPs are synthesized to contain one of a large number differentiator tags in their backbone such that the probability of any two MIPs in a reaction having the same differentiator tag sequence is low. MIP capture is performed on the sample; the reaction will be biased for shorter target lengths, and therefore the reaction product will be comprised of more 'short' circles than 'long' circles. Each circle should bear a unique differentiator tag sequence. Then, linear RCA (IRCA) is performed on the circles. In the IRCA reaction, circles are converted into long, linear concatemers of themselves. The IRCA reaction for a given circle stops when the concatemer has reached a 'fixed' length (based on the processivity/error rate of the polymerase). Concatemers derived from smaller circles will therefore contain more copies of the differentiator tag, and concatemers derived from larger circles will contain fewer copies of the differentiator tag. The number of each differentiator tag sequence is counted, for example, by next-generation sequencing. When number of occurrences is plotted against differentiator tag ID, the data will naturally cluster into two groups reflecting the lengths of the two alleles in the diploid sample. The allele lengths can therefore be read directly off this graph, after absolute

Petitioner Illumina, Inc., EX1004, p. 182

length calibration using known standards. In some embodiments, a sequencing technique (e.g., a next-generation sequencing technique) is used to sequence part of one or more captured targets (e.g., or amplicons thereof) and the sequences are used to count the number of different barcodes that are present. Accordingly, in some embodiments, aspects of the invention relate to a highly-multiplexed qPCR reaction.

Other non-limiting examples of loci at which insertions or deletions or repeat sequences may be associated with a disease or condition are provided in Tables 3 and 4. It should be appreciated that the presence of an abnormal length at any one or more of these loci may be evaluated according to aspects of the invention. In some embodiments, two or more of these loci or other loci may be evaluated in a single multiplex reaction using different probes designed to hybridize under the same reaction conditions to different target nucleic acid in a biological sample.

Table 3.

Polyglutamine (PolyQ) Diseases

Type	Gene	Normal/wildtype	Pathogenic
DRPLA (Dentatorubropallidoluysian atrophy)	ATN1 or DRPLA	6 - 35	49 - 88
HD (Huntington's disease)	HTT (Huntingtin)	10 - 35	35+
SBMA (Spinobulbar muscular atrophy or Kennedy disease)	Androgen receptor on the X chromosome.	9 - 36	38 - 62
SCA1 (Spinocerebellar ataxia Type 1)	ATXN1	6 - 35	49 - 88
SCA2 (Spinocerebellar ataxia Type 2)	ATXN2	14 - 32	33 - 77
SCA3 (Spinocerebellar ataxia Type 3 or Machado-Joseph disease)	ATXN3	12 - 40	55 - 86
SCA6 (Spinocerebellar ataxia Type 6)	CACNA1A	4 - 18	21 - 30
SCA7 (Spinocerebellar ataxia Type 7)	ATXN7	7 - 17	38 - 120
SCA17 (Spinocerebellar ataxia Type 17)	TBP	25 - 42	47 - 63

Table 4.

Petitioner Illumina, Inc., EX1004, p. 183

Non-Polyglutamine Diseases

Type	Gene	Codon	Normal/wildtype	Pathogenic
FRAXA (Fragile X syndrome)	*FMR1*, on the X-chromosome	CGG	6 - 53	230+
FXTAS (Fragile X-associated tremor/ ataxia syndrome)	*FMR1*, on the X-chromosome	CGG	6 - 53	55-200
FRAXE (Fragile XE mental retardation)	AFF2 or FMR2, on the X-chromosome	GCC	6 - 35	200+
FRDA (Friedreich's ataxia)	FXN or X25, (frataxin)	GAA	7 - 34	100+
DM (Myotonic dystrophy)	DMPK	CTG	5 - 37	50+
SCA8 (Spinocerebellar ataxia Type 8)	OSCA or SCA8	CTG	16 - 37	110 – 250
SCA12 (Spinocerebellar ataxia Type 12)	PPP2R2B or SCA12	CAG On 5' end	7 - 28	66 - 78

The following examples illustrate aspects and embodiments of the invention and are not intended to be limiting or restrictive. Many variations of the invention will become apparent to those skilled in the art upon review of this specification. The full scope of the invention should be determined by reference to the claims, along with their full scope of equivalents, and the specification, along with such variations.

4. Increasing detection sensitivity:

In some embodiments, aspects of the invention relate to methods for increasing the sensitivity of nucleic acid detection assays.

There are currently many genomic assays that utilize next-generation (e.g., polony-based) sequencing to generate data, including genome resequencing, RNA-seq for gene expression, bisulphite sequencing for methylation, and Immune-seq, among others. In order to make quantitative measurements (including genotype calling), these

Petitioner Illumina, Inc., EX1004, p. 184

methods utilize the counts of sequencing reads of a given genomic locus as a proxy for the representation of that sequence in the original sample of nucleic acids. The majority of these techniques require a preparative step to construct a high-complexity library of DNA molecules that is representative of a sample of interest. Current assays use one of several alternative nucleic acid preparative techniques (e.g., amplification, for example PCR-based amplification; sequence-specific capture, for example, using immobilized capture probes; or target capture into a circularized probe followed by a sequence analysis step. In order to reduce errors associated with the unpredictability (stochastic nature) of nucleic acid isolation and sequence analysis techniques, current methods involve oversampling a target nucleic acid preparation in order to increase the likelihood that all sequences that are present in the original nucleic acid sample will be represented in the final sequence data. For example, a genomic sequencing library may contain an over- or under-representation of particular sequences from a source nucleic acid sample (e.g., genome preparation) as a result of stochastic variations in the library construction process. Such variations can be particularly problematic when they result in target sequences from a genome being absent or undetectable in a sequencing library. For example, an under-representation of particular allelic sequences (e.g., heterozygotic alleles) from a genome in a sequencing library can result in an apparent homozygous representation in a sequencing library.

In contrast, aspects of the invention relate to basing a nucleic acid sequence analysis on results from two or more different nucleic acid preparatory techniques that have different systematic biases in the types of nucleic acids that they sample rather than simply oversampling the target nucleic acid. According to some embodiments, different techniques have different sequence biases that are systematic and not simply due to stochastic effects during nucleic acid capture or amplification. Accordingly, in some embodiments, the degree of oversampling required to overcome variations in nucleic acid preparation needs to be sufficient to overcome the biases. In some embodiments, the invention provides methods that reduce the need for oversampling by combining nucleic acid and/or sequence results obtained from two or more different nucleic acid preparative techniques that have different biases.

According to the invention, different techniques have different characteristic or systematic biases. For example, one technique may bias a sample analysis towards one

Petitioner Illumina, Inc., EX1004, p. 185

particular allele at a genetic locus of interest, whereas a different technique would bias the sample analysis towards a different allele at the same locus. Accordingly, the same sample may be identified as being different depending on the type of technique that is used to prepare nucleic acid for sequence analysis. This effectively represents a sensitivity issue, because each technique has a different relative sensitivities for polymorphic sequences of interest.

According to aspects of the invention, the sensitivity of a nucleic acid analysis can be increased by combining the sequences from different nucleic acid preparative steps and using the combined sequence information for a diagnostic assay (e.g., for a making a call as to whether a subject is homozygous or heterozygous at a genetic locus of interest).

Currently, the ability of DNA sequencing to detect mutations is limited by the ability of the upstream sample isolation (e.g., by amplification, immobilization enrichment, circularization capture, etc.) methods to reliably isolate the locus of interest. If one wishes to make heterozygote base-calls for a diploid genome (e.g. a human sample presented for molecular diagnostic sequencing), it is important in some embodiments that the isolation method produces near- or perfectly- uniform amounts of the two alleles to be sequenced (at least sufficiently uniform to be "called" unambiguously as a heterozygote or a homozygote for a locus of interest).

Sample preparative methods may fall into three classes: 1) single- or several-target amplification (e.g., uniplex PCR, 'multiplex' PCR), 2) multi-target hybridization enrichment (e.g., Agilent SureSelect 'hybrid capture' [Gnirke et al 2009, *Nature methods* 27:182-9], Roche/Nimblegen 'sequence capture' [Hodges et al 2007, *Nature genetics* 39:1522-7], and 3) multi-target circularization selection (e.g. molecular inversion probes or padlock probes, [Porreca et al 2007, *Nature methods* 4:931-6, Turner et al 2009, *Nature methods* 6:315-6], 'selectors' [Dahl et al 2005, *Nucleic acids research* 33:e71]). Each of these methods can result in a pool of isolated product that does not adequately represent the input abundance distribution. For example, the two alleles at a heterozygous position can become skewed far from their input 50:50 ratio to something that results in a missed basecall during downstream sequencing. For example, if the ratio was skewed from 50:50 to 10:90, and the sample was sequenced to 10X average coverage, there is a high probability that one of the two alleles would not be observed once in the ten sequencing reads. This would reduce the sensitivity of the sequencing

Petitioner Illumina, Inc., EX1004, p. 186

method by converting a heterozygous position to homozygous (where potentially the 'mutant' allele was the one not observed). In some embodiments, a skewed ratio is a particular issue that decreases the sensitivity of detecting mutations present in a heterogeneous tumor tissue. For example, if only 10% of the cells analyzed in a heterogeneous sample harbored a heterozygous mutation, the mutation would be expected to be present in 5% of sequence reads, not 50%. In this scenario, the need for robust, sensitive detection may be even more acute.

The methods disclosed herein are based, in part, on the discovery that certain classes of isolation methods have different modes of bias. The disclosure provide methods for increasing the sensitivity of the downstream sequencing by using a combination of multiple isolation methods (e.g., one or more from at least two of the classes disclosed herein) for a sample. This is particularly important in molecular diagnostics where high sensitivity is required to minimize the chances of 'missing' a disease-associated mutation. For example, given a nominal false-negative error rate of $1x10^{-3}$ for sequencing following circularization selection, and a false-negative error rate of $1x10^{-3}$ for sequencing following hybridization enrichment, one can achieve a final false-negative rate of $1x10^{-6}$ by performing both techniques on the sample (assuming failures in each method are fully independent). For a recessive disease with carrier frequency of 0.1, caused by a single fully-penetrant mutant allele, the number of missed carrier diagnoses would decrease from 1000 per million patients tested to 1 per million patients tested. Furthermore, if the testing was used in the context of prenatal carrier screening, the number of affected children born as a result of missing the carrier call in one parent would decrease from 25 per million to 25 per billion born.

Additionally, the disclosure provides combinations of preparative methods to effectively increase sequencing coverage in regions containing disease-associated alleles. Since heterozygote error rate is largely tied to both deviations from 50:50 allele representation, and in the case of next-generation DNA sequencing deviations from average abundance (such that less abundant isolated targets are more likely to be undersampled at one or both alleles), selectively increasing coverage in these regions will also selectively increase sensitivity. Furthermore, MIPs that detect presence or absence of specific known disease-associated mutations can be used to increase sensitivity selectively. In some embodiments, these MIPs would have a targeting arm whose 3' –most region is complementary to the expected mutation, and has a fill-in

Petitioner Illumina, Inc., EX1004, p. 187

length of 0 or more bp. Thus, the MIP will form only if the mutation is present, and its presence will be detected by sequencing.

Additionally, algorithms disclosed herein may be used to determine base identity with varying levels of stringency depending on whether the given position has any known disease-associated alleles. Stringency can be reduced in such positions by decreasing the minimum number of observed mutant reads necessary to make a consensus base-call. This will effectively increase sensitivity for mutant allele detection at the cost of decreased specificity.

An embodiment of the invention combines MIPs plus hybridization enrichment, plus optionally extra MIPs targeted to specific known, common disease-associated loci, e.g., to detect the presence of a polymorphism in a target nucleic acid. A non-limiting example is illustrated in FIG. 12 that illustrates a schematic using MIPs plus hybridization enrichment, plus optionally extra MIPs targeted to specific known, common disease-associated loci, e.g., to detect the presence of a polymorphism in a target nucleic acid.

FIGs. 13 and 14 illustrate different capture efficiencies for MIP-based captures. FIG. 13 shows a graph of per-target abundance with MIP capture. In this graph, bias largely drives the heterozygote error rate, since targets which are less abundant here are less likely to be covered in sufficient depth during sequencing to adequately sample both alleles. This is from Turner et al 2009, Nature methods 6:315-6. Hybridization enrichment results in a qualitatively similar abundance distribution, but the abundance of a given target is likely not correlated between the two methods. FIG. 14 shows a graph of correlation between two MIP capture reactions from Ball et al 2009, Nature biotechnology 27:361-8. Each point represents the target abundance in replicate 1 and replicate 2. Pearson correlation r=0.956. This indicates that MIP capture reproducibly biases targets to specific abundances. Hybridization enrichment is similarly correlated from one capture to the next.

According to aspects of the invention, such biases can be detected or overcome by systematically combining different capture and/or analytical techniques in an assay that interrogates a plurality of loci in a plurality of subject samples.

Accordingly, it should be appreciated that in any of the embodiments described herein (e.g., tiling/staggering, tagging, size-detection, sensitivity enhancing algorithms,

Petitioner Illumina, Inc., EX1004, p. 188

or any combination thereof), aspects of the invention involve preparing genomic nucleic acid and/or contacting them with one or more different probes (e.g., capture probes, hybridization probes, MIPs, others etc.). In some embodiments, the amount of genomic nucleic acid used per subject ranges from 1 ng to 10 micrograms (e.g., 500 ng to 5 micrograms). However, higher or lower amounts (e.g., less than 1 ng, more than 10 micrograms, 10-50 micrograms, 50-100 micrograms or more) may be used. In some embodiments, for each locus of interest, the amount of probe used per assay may be optimized for a particular application. In some embodiments, the ratio (molar ratio, for example measured as a concentration ratio) of probe to genome equivalent (e.g., haploid or diploid genome equivalent, for example for each allele or for both alleles of a nucleic acid target or locus of interest) ranges from 1/100, 1/10, 1/1, 10/1, 100/1, 1000/1. However, lower, higher, or intermediate ratios may be used.

In some embodiments, the amount of target nucleic acid and probe used for each reaction is normalized to avoid any observed differences being caused by differences in concentrations or ratios. In some embodiments, in order to normalize genomic DNA and probe, the genomic DNA concentration is read using a standard spectrophotometer or by fluorescence (e.g., using a fluorescent intercalating dye). The probe concentration may be determined experimentally or using information specified by the probe manufacturer.

Similarly, once a locus has been captured (e.g., on a MIP or other probe or in another form), it may be amplified and/or sequenced in a reaction involving one or more primers. The amount of primer added for each reaction can range from 0.1 pmol to 1 nmol, 0.15 pmol to 1.5 nmol (for example around 1.5 pmol). However, other amounts (e.g., lower, higher, or intermediate amounts) may be used.

In some embodiments, it should be appreciated that one or more intervening sequences (e.g., sequence between the first and second targeting arms on a MIP capture probe), identifier or tag sequences, or other probe sequences that are not designed to hybridize to a target sequence (e.g., a genomic target sequence) should be designed to avoid excessive complementarity (to avoid cross-hybridization) to target sequences or other sequences (e.g., other genomic sequences) that may be in a biological sample. For example, these sequences may be designed have a sufficient number of mismatches with any genomic sequence (e.g., at least 5, 10, 15, or more mismatches out of 30 bases) or as having a Tm (e.g., a mismatch Tm) that is lower (e.g., at least 5, 10, 15, 20, or more degrees C lower) than the hybridization reaction temperature.

Petitioner Illumina, Inc., EX1004, p. 189

It should be appreciated that a targeting arm as used herein may be designed to hybridize (e.g., be complementary) to either strand of a genetic locus of interest if the nucleic acid being analyzed is DNA (e.g., genomic DNA). However, in the context of MIP probes, whichever strand is selected for one targeting arm will be used for the other one. However, in the context of RNA analysis, it should be appreciated that a targeting arm should be designed to hybridize to the transcribed RNA. It also should be appreciated that MIP probes referred to herein as "capturing" a target sequence are actually capturing it by template-based synthesis rather than by capturing the actual target molecule (other than for example in the initial stage when the arms hybridize to it or in the sense that the target molecule can remain bound to the extended MIP product until it is denatured or otherwise removed).

It should be appreciated that in some embodiments a targeting arm may include a sequence that is complementary to one allele or mutation (e.g., a SNP or other polymorphism, a mutation, etc.) so that the probe will preferentially hybridize (and capture) target nucleic acids having that allele or mutation. However, in many embodiments, each targeting arm is designed to hybridize (e.g., be complementary) to a sequence that is not polymorphic in the subjects of a population that is being evaluated. This allows target sequences to be captured and/or sequenced for all alleles and then the differences between subjects (e.g., calls of heterozygous or homozygous for one or more loci) can be based on the sequence information and/or the frequency as described herein.

It should be appreciated that sequence tags (also referred to as barcodes) may be designed to be unique in that they do not appear at other positions within a probe or a family of probes and they also do not appear within the sequences being targeted. Thus they can be used to uniquely identify (e.g., by sequencing or hybridization properties) particular probes having other characteristics (e.g., for particular subjects and/or for particular loci).

It also should be appreciated that in some embodiments probes or regions of probes or other nucleic acids are described herein as comprising or including certain sequences or sequence characteristics (e.g., length, other properties, etc.). However, it should be appreciated that in some embodiments, any of the probes or regions of probes or other nucleic acids consist of those regions (e.g., arms, central regions, tags, primer sites, etc., or any combination thereof) of consist of those sequences or have sequences with characteristics that consist of one or more characteristics (e.g., length, or other

Petitioner Illumina, Inc., EX1004, p. 190

properties, etc.) as described herein in the context of any of the embodiments (e.g., for tiled or staggered probes, tagged probes, length detection, sensitivity enhancing algorithms or any combination thereof).

It should be appreciated that probes, primers, and other nucleic acids designed or used herein may be synthetic, natural, or a combination thereof. Accordingly, as used herein, the term "nucleic acid" refers to multiple linked nucleotides (i.e., molecules comprising a sugar (e.g., ribose or deoxyribose) linked to an exchangeable organic base, which is either a pyrimidine (e.g., cytosine (C), thymidine (T) or uracil (U)) or a purine (e.g., adenine (A) or guanine (G)). "Nucleic acid" and "nucleic acid molecule" may be used interchangeably and refer to oligoribonucleotides as well as oligodeoxyribonucleotides. The terms shall also include polynucleosides (i.e., a polynucleotide minus a phosphate) and any other organic base containing nucleic acid. The organic bases include adenine, uracil, guanine, thymine, cytosine and inosine. Unless otherwise stated, nucleic acids may be single or double stranded. The nucleic acid may be naturally or non-naturally occurring. Nucleic acids can be obtained from natural sources, or can be synthesized using a nucleic acid synthesizer (i.e., synthetic). Harvest and isolation of nucleic acids are routinely performed in the art and suitable methods can be found in standard molecular biology textbooks. (See, for example, Maniatis' Handbook of Molecular Biology.) The nucleic acid may be DNA or RNA, such as genomic DNA, mitochondrial DNA, mRNA, cDNA, rRNA, miRNA, or a combination thereof. Non-naturally occurring nucleic acids such as bacterial artificial chromosomes (BACs) and yeast artificial chromosomes (YACs) can also be used.

The invention also contemplates the use of nucleic acid derivatives. As will be described herein, the use of certain nucleic acid derivatives may increase the stability of the nucleic acids of the invention by preventing their digestion, particularly when they are exposed to biological samples that may contain nucleases. As used herein, a nucleic acid derivative is a non-naturally occurring nucleic acid or a unit thereof. Nucleic acid derivatives may contain non-naturally occurring elements such as non-naturally occurring nucleotides and non-naturally occurring backbone linkages.

Nucleic acid derivatives may contain backbone modifications such as but not limited to phosphorothioate linkages, phosphodiester modified nucleic acids, phosphorothiolate modifications, combinations of phosphodiester and phosphorothioate nucleic acid, methylphosphonate, alkylphosphonates, phosphate esters,

Petitioner Illumina, Inc., EX1004, p. 191

alkylphosphonothioates, phosphoramidates, carbamates, carbonates, phosphate triesters, acetamidates, carboxymethyl esters, methylphosphorothioate, phosphorodithioate, p-ethoxy, and combinations thereof. The backbone composition of the nucleic acids may be homogeneous or heterogeneous.

Nucleic acid derivatives may contain substitutions or modifications in the sugars and/or bases. For example, they include nucleic acids having backbone sugars which are covalently attached to low molecular weight organic groups other than a hydroxyl group at the 3' position and other than a phosphate group at the 5' position (e.g., an 2'-O-alkylated ribose group). Nucleic acid derivatives may include non-ribose sugars such as arabinose. Nucleic acid derivatives may contain substituted purines and pyrimidines such as C-5 propyne modified bases, 5-methylcytosine, 2-aminopurine, 2-amino-6-chloropurine, 2,6-diaminopurine, hypoxanthine, 2-thiouracil and pseudoisocytosine. In some embodiments, substitution(s) may include one or more substitutions/modifications in the sugars/bases, groups attached to the base, including biotin, fluorescent groups (fluorescein, cyanine, rhodamine, etc), chemically-reactive groups including carboxyl, NHS, thiol, etc., or any combination thereof.

A nucleic acid may be a peptide nucleic acid (PNA), locked nucleic acid (LNA), DNA, RNA, or co-nucleic acids of the same such as DNA-LNA co-nucleic acids. PNA are DNA analogs having their phosphate backbone replaced with 2-aminoethyl glycine residues linked to nucleotide bases through glycine amino nitrogen and methylenecarbonyl linkers. PNA can bind to both DNA and RNA targets by Watson-Crick base pairing, and in so doing form stronger hybrids than would be possible with DNA or RNA based oligonucleotides in some cases.

PNA are synthesized from monomers connected by a peptide bond (Nielsen, P.E. et al. Peptide Nucleic Acids, Protocols and Applications, Norfolk: Horizon Scientific Press, p. 1-19 (1999)). They can be built with standard solid phase peptide synthesis technology. PNA chemistry and synthesis allows for inclusion of amino acids and polypeptide sequences in the PNA design. For example, lysine residues can be used to introduce positive charges in the PNA backbone. All chemical approaches available for the modifications of amino acid side chains are directly applicable to PNA. Several types of PNA designs exist, and these include single strand PNA (ssPNA), bisPNA and pseudocomplementary PNA (pcPNA).

Petitioner Illumina, Inc., EX1004, p. 192

The structure of PNA/DNA complex depends on the particular PNA and its sequence. ssPNA binds to single stranded DNA (ssDNA) preferably in antiparallel orientation (i.e., with the N-terminus of the ssPNA aligned with the 3' terminus of the ssDNA) and with a Watson-Crick pairing. PNA also can bind to DNA with a Hoogsteen base pairing, and thereby forms triplexes with double stranded DNA (dsDNA) (Wittung, P. et al., *Biochemistry* 36:7973 (1997)).

A locked nucleic acid (LNA) is a modified RNA nucleotide. An LNA form hybrids with DNA, which are at least as stable as PNA/DNA hybrids (Braasch, D.A. et al., *Chem & Biol.* 8(1):1-7(2001)). Therefore, LNA can be used just as PNA molecules would be. LNA binding efficiency can be increased in some embodiments by adding positive charges to it. LNAs have been reported to have increased binding affinity inherently.

Commercial nucleic acid synthesizers and standard phosphoramidite chemistry are used to make LNAs. Therefore, production of mixed LNA/DNA sequences is as simple as that of mixed PNA/peptide sequences. The stabilization effect of LNA monomers is not an additive effect. The monomer influences conformation of sugar rings of neighboring deoxynucleotides shifting them to more stable configurations (Nielsen, P.E. et al. Peptide Nucleic Acids, Protocols and Applications, Norfolk: Horizon Scientific Press, p. 1-19 (1999)). Also, lesser number of LNA residues in the sequence dramatically improves accuracy of the synthesis. Most of biochemical approaches for nucleic acid conjugations are applicable to LNA/DNA constructs.

These and other aspects of the invention are illustrated by the following non-limiting examples.

EXAMPLES

The following examples illustrate non-limiting embodiments of the invention.

Example 1: Design a set of capture probes for a human target exon

All targets are captured as a set of partially-overlapping subtargets. For example, in the tiling approach, a 200 bp target exon might be captured as a set of 12 subtargets,

Petitioner Illumina, Inc., EX1004, p. 193

each 60bp in length (FIG. 1). Each subtarget is chosen such that it partially overlaps two or three other targets.

In some embodiments, all probes are composed of three regions: 1) a 20 bp 'targeting arm' comprised of sequence which hybridizes immediately upstream from the sub-target, 2) a 30 bp 'constant region' comprised of sequence used as a pair of amplification priming sites, and 3) a second 20 bp 'targeting arm' comprised of sequence which hybridizes immediately downstream from the sub-target. Targeting arm sequences will be different for each capture probe in a set, while constant region sequence will be the same for all probes in the set, allowing all captured targets to be amplified with a single set of primers. Targeting arm sequences should be designed such that any given pair of 20 bp sequences is unique in the target genome (to prevent spurious capture of undesired sites). Additionally, melting temperatures should be matched for all probes in the set such that hybridization efficiency is uniform for all probes at a constant temperature (e.g., 60 C). Targeting arm sequences should be computationally screened to ensure they do not form strong secondary structure that would impair their ability to basepair with the genomic target.

Hybridize capture probes to human genomic sample

Assemble hybridization reaction:

- 1.0 ul capture probe mix (~2.5 pmol)
- 2.0 ul 10X Ampligase buffer (Epicentre)
- 6.0 ul 500 ng/ul human genomic DNA (~16.7 fmol)
- 11 ul dH2O

In a thermal cycler, heat reaction to 95C for 5 min to denature genomic DNA, then cool to 60C. Allow to incubate at 60C for 40 hours.

Convert hybridized probes into covalently-closed circular products containing subtargets

Prepare fill-in / ligation reaction mixture:

- 0.25 ul 2 mM dNTP mix (Invitrogen)
- 2.5 ul 10X Ampligase buffer (Epicentre)
- 5.0 ul 5 U/ul Taq Stoffel fragment (Applied Biosystems)

Petitioner Illumina, Inc., EX1004, p. 194

- 12.5 ul 5 U/ul Ampligase (Epicentre)
- 4.75 ul dH2O

Add 1.0 ul of this mix to the hybridized probe reaction, and incubate at 60C for 10 hours.

Purify circularized probe / subtarget products from un-reacted probes and genomic DNA

Prepare exonuclease reaction mixture:

- 21 ul fill-in / ligation reaction product
- 2.0 ul 10X exonuclease I buffer (New England Biolabs)
- 2.0 ul 20 U/ul exonuclease I (New England Biolabs)
- 2.0 ul 100 U/ul exonuclease III (New England Biolabs)

Incubte at 37C for 60 min, then heat-inactivate by incubating at 80C for 15 min. Immediately cool to 4C for storage.

Amplify circular material by PCR using primers specific to the 'constant region' of the probes

Prepare PCR mixture:

- 5.0 ul 10X Accuprime reaction buffer (Invitrogen)
- 1.5 ul 10 uM CP-2-FA (5'-GCACGATCCGACGGTAGTGT-3') (SEQ ID NO: 183)
- 1.5 ul 10 uM CP-2-RA (5'-CCGTAATCGGGAAGCTGAAG-3') (SEQ ID NO: 184)
-
- 0.4 ul 25 mM dNTP mix (Invitrogen)
- 2.0 ul heat-inactivated exonuclease reaction mix
- 1.5 ul 10X SybrGreen (Invitrogen)
- 0.4 ul 2.5 U/ul Accuprime Pfx polymerase (Invitrogen)
- 37.7 ul dH2O

Thermal cycle in real-time thermal cycler according to the following protocol, but stop cycling before amplification yield plateaus (generally 8-12 cycles):

1. 95C for 5 min

Petitioner Illumina, Inc., EX1004, p. 195

2. 95C for 30 sec
3. 58C for 60 sec
4. 72C for 60 sec
5. goto 2, N more times

Prepare a shotgun next-generation sequencing library for analysis

- Purify desired amplicon population from non-specific amplification products by gel extraction.
- Concatemerize amplicons into high-molecular weight products suitable for shearing
- Mechanically shear, using either a nebulizer, BioRuptor, Hydroshear, Covaris, or similar instrument. DNA should be sheared into fragments several hundred basepairs in length.
- Ligate adapters required for amplification by the sequencing platform used. If necessary, purify ligated product from unligated product and adapters.

Example 2: Use of differentiator tag sequences to detect and correct bias in a MIP-capture reaction of a set of exon targets

The first step in performing the detection/correction is to determine how many differentiator tag sequences are necessary for the given sample. In this example, 1000 genomic targets corresponding to 1000 exons were captured. Since the differentiator tag sequence is part of the probe, it will measure/report biases that occur from the earliest protocol steps. Also, being located in the backbone, the differentiator tag sequence can easily be sequenced from a separate priming site, and therefore not impact the total achievable read-length for the target sequence. MIP probes are synthesized using standard column-based oligonucleotide synthesis by any number of vendors (e.g. IDT), and differentiator tag sequences are introduced as 'degenerate' positions in the backbone. Each degenerate position increases the total number of differentiator tag sequences synthesized by a factor of 4, so a 10 nt degenerate region implies a differentiator tag sequence complexity of ~1e6 species.

Hybridize capture probes to human genomic sample

Petitioner Illumina, Inc., EX1004, p. 196

Assemble hybridization reaction:

- 1.0 ul capture probe mix (~2.5 pmol)
- 2.0 ul 10X Ampligase buffer (Epicentre)
- 6.0 ul 500 ng/ul human genomic DNA (~16.7 fmol)
- 11 ul dH2O

In a thermal cycler, heat reaction to 95C for 5 min to denature genomic DNA, then cool to 60C. Allow to incubate at 60C for 40 hours.

Convert hybridized probes into covalently-closed circular products containing subtargets

Prepare fill-in / ligation reaction mixture:

- 0.25 ul 2 mM dNTP mix (Invitrogen)
- 2.5 ul 10X Ampligase buffer (Epicentre)
- 5.0 ul 5 U/ul Taq Stoffel fragment (Applied Biosystems)
- 12.5 ul 5 U/ul Ampligase (Epicentre)
- 4.75 ul dH2O

Add 1.0 ul of this mix to the hybridized probe reaction, and incubate at 60C for 10 hours.

Purify circularized probe / subtarget products from un-reacted probes and genomic DNA

Prepare exonuclease reaction mixture:

- 21 ul fill-in / ligation reaction product
- 2.0 ul 10X exonuclease I buffer (New England Biolabs)
- 2.0 ul 20 U/ul exonuclease I (New England Biolabs)
- 2.0 ul 100 U/ul exonuclease III (New England Biolabs)

Incubate at 37C for 60 min, then heat-inactivate by incubating at 80C for 15 min. Immediately cool to 4C for storage.

Amplify circular material by PCR using primers specific to the 'constant region' of the probes

Prepare PCR mixture:

Petitioner Illumina, Inc., EX1004, p. 197

- 5.0 ul 10X Accuprime reaction buffer (Invitrogen)
- 1.5 ul 10 uM CP-2-FA (5'-GCACGATCCGACGGTAGTGT-3') (SEQ ID NO: 183)
- 1.5 ul 10 uM CP-2-RA (5'-CCGTAATCGGGAAGCTGAAG-3') (SEQ ID NO: 184)
- 0.4 ul 25 mM dNTP mix (Invitrogen)
- 2.0 ul heat-inactivated exonuclease reaction mix
- 1.5 ul 10X SybrGreen (Invitrogen)
- 0.4 ul 2.5 U/ul Accuprime Pfx polymerase (Invitrogen)
- 37.7 ul dH2O

Thermal cycle in real-time thermal cycler according to the following protocol, but stop cycling before amplification yield plateaus (generally 8-12 cycles):

6. 95C for 5 min
7. 95C for 30 sec
8. 58C for 60 sec
9. 72C for 60 sec
10. goto 2, N more times

Prepare a shotgun next-generation sequencing library for analysis

- Purify desired amplicon population from non-specific amplification products by gel extraction.
- Concatemerize amplicons into high-molecular weight products suitable for shearing
- Mechanically shear, using either a nebulizer, BioRuptor, Hydroshear, Covaris, or similar instrument. DNA should be sheared into fragments several hundred basepairs in lenth.
- Ligate adapters required for amplification by the sequencing platform used. If necessary, purify ligated product from unligated product and adapters.

Petitioner Illumina, Inc., EX1004, p. 198

Perform sequencing of library according to manufacturer's directions (e.g. Illumina,ABI, etc), reading both the target sequence and the differentiator tag sequence.

Analyze data by correcting for any biases detected by quantitation of differentiator tag sequence abundance

Construct a table of target:differentiator tag abundances from the read data, e.g.:

Target ID	Differentiator tag sequence ID	Count
1	3547	1
2	4762	1
1	9637	1
1	1078	5
3	4762	1
1	2984	1

All 'count' entries should be '1', since any particular target:differentiator tag mapping will not occur more than once by chance, and therefore will only be observed if bias was present somewhere in the sample preparation process. For any target:differentiator tag combination observed more than once, all such reads are 'collapsed' into a single read before consensus basecalls are determined. This will cancel the effect of bias on consensus basecall accuracy. FIG. 5 depicts a method for making diploid genotype calls in which repeat target:differentiator tag combination are collapsed.

Example 3: Differentiator tag sequence design for MIP capture reactions

For a set of targets, the number of differentiator tag sequences necessary to be confident (within some statistical bounds) that a certain differentiator tag sequence will not be observed more than once by chance in combination with a certain target sequence was determined. The total number of unique differentiator tag sequences for a certain differentiator tag sequence length is determined as $4^{(\text{Length in nucleotides of the differentiator tag sequence})}$. For a molecular inversion probe capture reaction that uses MIP probes having differentiator tag sequences, the probability of performing the capture reaction and capturing one or more copies of a target sequence having the same differentiator tag sequence is calculated as: p = 1 - [N! / (N-M)!] / [N^M], wherein N is the total number

Petitioner Illumina, Inc., EX1004, p. 199

of possible unique differentiator tag sequences and M is the number of target sequence copies in the capture reaction. Thus, by varying the differentiator tag sequence length it is possible to perform a MIP capture reaction in which the probability of capturing one or more copies of a target sequence having the same differentiator tag sequence is set at a predetermined probability value.

For example, for a differentiator tag sequence of 15 nucleotides in length, there are 1,073,741,824 possible differentiator tag sequences. A MIP capture reaction in which MIP probes, each having a differentiator tag sequence of 15 nucleotides, are combined with 10000 target sequence copies (e.g., genome equivalents), the probability of capturing one or more copies of a target sequence having the same differentiator tag sequence is 0.05. In this example, the MIP reaction will produce very few (usually 0, but occasionally 1 or more) targets where multiple copies are tagged with the same differentiator tag sequence. FIG. 6 depicts results of a simulation for 100000 capture reactions having 15 nucleotide differentiator tag sequences and 10000 target sequences.

Example 4: Assessment of the probability for obtaining enough sequencing reads to make accurate base-calls at multiple independent loci, as a function of sequencing coverage.

Monte Carlo simulations were performed to determine sequencing coverage requirements. The simulations assume 10000 genomic copies of a given locus (target) half mom alleles and half dad alleles. The simulations further assume 1% efficiency of capture for the MIP reaction. The simulation samples from a capture mix 100 times without replacement to create a set of 100 capture products. The simulation then samples from the set of 100 capture products with replacement (assuming unbiased amplification) to generate 'reads' from either mom or dad. The number of reads sampled depends on the coverage. The number of independent reads from both mom and dad necessary to make a high-quality base-call (assumed to be 10 or 20 reads) were then determined. The process was repeated 1000 times for each coverage level, and the fraction of times that enough reads from both parents were successfully obtained was determined. This fraction was raised to the power 1000, assuming we have 1000 independent loci that must obtain successful base-calls, plotted (See FIG. 7). Result show that roughly 50x coverage is required to capture each allele >= 10x with > 0.95 probability.

Petitioner Illumina, Inc., EX1004, p. 200

Example 5: MIP capture of 'target' locus and 'control' loci

In some embodiments, to accurately quantify the efficiency of target locus capture, at least three sets of control loci are captured in parallel that have *a priori* been shown to serve as proxies for various lengths of target locus. For example, if the target locus is expected to have a length between 50 and 1000 bp, then sets of control loci having lengths of 50, 250, and 1000 bp could be captured (e.g. 20 loci per set should provide adequate protection from outliers), and their abundance digitally measured by sequencing. These loci should be chosen such that minimal variation in efficiency between samples and on multiple runs of the same sample is observed (and are therefore 'efficiency invariant'). These will serve as 'reference' points that define the shape of the curve of abundance-vs-length. Determining the length of the target is then simply a matter of 'reading' the length from the appropriate point on the calibration curve.

In some embodiments, the statistical confidence one has in the estimate of target length from this method is driven largely by three factors: 1) reproducibility / variation of the abundance data used to generate the calibration curve; 2) goodness of fit of the regression to the 'control' datapoints; 3) reproducibility of abundance data for the target locus being measured. Statistical bounds on 1) and 2) will be known in advance, having been measured during development of the assay. Additionally, statistical bounds on 3) will be known in general in advance, since assay development should include adequate population sampling and measure of technical reproducibility. Standard statistical methods should be used to combine these three measures into a single *P* value for any given experimental measure of target abundance.

In some embodiments, given the set of calibration observations, and a linear regression fit to that data, the regression can be used to predict the length value for n observations of the target locus whose length is unknown. First, choose an acceptable range for the confidence interval of the length estimate. For example, in the case of distinguishing "normal" (87-93bp) from "premutation" (165-600bp) potential cases of Fragile X, the goal is to measure length to sufficient precision to distinguish 93 bp from 165 bp. The predicted response value, computed when n observations is substituted into the equation for the regressed line, will have arbitrary precision. However, if for example a 95% confidence level is desired, that 95% confidence interval must be sufficiently short that it does not overlap both the "normal" and "premutation" length

Petitioner Illumina, Inc., EX1004, p. 201

ranges. Continuing the example, if one calculates a length of 190 from n=400 MIP observations, and based on the regression from calibration data, the 95% confidence interval is 190 +/- 20 bp, one can conclude the sample represents a "premutation" length with 95% certainty. Conversely, if the calibration data were less robust, error estimates of the regression would be higher, leading to larger confidence intervals on the predicted response value. In some embodiments, if the 95% CI were calculated as 190 +/- 100 bp from n=400, one could not determine whether the predicted response value corresponds to a "normal" or "premutation" length.

In some embodiments, the confidence interval for a predicted response is calculated as:

The estimate for the response $\hat{y}$ *is identical to the estimate for the mean of the response:* $\hat{y} = b_0 + b_1 x^*$. *The confidence interval for the predicted value is given by* $\hat{y} \pm t^* s_{\hat{y}}$, *where* $\hat{y}$ *is the fitted value corresponding to* x^*. *The value* t^* *is the upper (1 - C)/2 critical value for the t(n - 2) distribution.*

In some embodiments, a technique for analyzing a locus of interest can involve the following steps.

Convert hybridized probes into covalently-closed circular products containing subtargets

Prepare fill-in / ligation reaction mixture:

- 0.25 ul 2 mM dNTP mix (Invitrogen)
- 2.5 ul 10X Ampligase buffer (Epicentre)
- 5.0 ul 5 U/ul Taq Stoffel fragment (Applied Biosystems)
- 12.5 ul 5 U/ul Ampligase (Epicentre)
- 4.75 ul dH2O

Add 1.0 ul of this mix to the hybridized probe reaction, and incubate at 60C for 10 hours.

Purify circularized probe / subtarget products from un-reacted probes and genomic DNA

Prepare exonuclease reaction mixture:

Petitioner Illumina, Inc., EX1004, p. 202

- 21 ul fill-in / ligation reaction product
- 2.0 ul 10X exonuclease I buffer (New England Biolabs)
- 2.0 ul 20 U/ul exonuclease I (New England Biolabs)
- 2.0 ul 100 U/ul exonuclease III (New England Biolabs)

Incubte at 37C for 60 min, then heat-inactivate by incubating at 80C for 15 min. Immediately cool to 4C for storage.

Amplify circular material by PCR using primers specific to the 'constant region' of the probes

Prepare PCR mixture:

- 5.0 ul 10X Accuprime reaction buffer (Invitrogen)
- 1.5 ul 10 uM CP-2-FA-Ilmn (platform-specific amplification sequence plus 'circle constant region'-specific sequence)
- 1.5 ul 10 uM CP-2-RA-Ilmn (platform-specific amplification sequence plus 'circle constant region'-specific sequence)
- 0.4 ul 25 mM dNTP mix (Invitrogen)
- 2.0 ul heat-inactivated exonuclease reaction mix
- 1.5 ul 10X SybrGreen (Invitrogen)
- 0.4 ul 2.5 U/ul Accuprime Pfx polymerase (Invitrogen)
- 37.7 ul dH2O

Thermal cycle in real-time thermal cycler according to the following protocol, but stop cycling before amplification yield plateaus (generally 8-12 cycles):

11. 95C for 5 min
12. 95C for 30 sec
13. 58C for 60 sec
14. 72C for 60 sec
15. goto 2, N more times

Perform sequencing (e.g., next-generation sequencing) on sample for digital quantitation according to manufacturer's instructions (e.g., Illumina, ABI)

Example 6: MIP-capture reaction of a set of exon target nucleic acids

Petitioner Illumina, Inc., EX1004, p. 203

MIP probes are synthesized using standard column-based oligonucleotide synthesis by any number of vendors (e.g. IDT).

Hybridize capture probes to human genomic sample

Assemble hybridization reaction:

- 1.0 ul capture probe mix (~2.5 pmol)
- 2.0 ul 10X Ampligase buffer (Epicentre)
- 6.0 ul 500 ng/ul human genomic DNA (~16.7 fmol)
- 11 ul dH2O

In a thermal cycler, heat reaction to 95C for 5 min to denature genomic DNA, then cool to 60C. Allow to incubate at 60C for 40 hours.

Convert hybridized probes into covalently-closed circular products containing target nucleic acids

Prepare fill-in / ligation reaction mixture:

- 0.25 ul 2 mM dNTP mix (Invitrogen)
- 2.5 ul 10X Ampligase buffer (Epicentre)
- 5.0 ul 5 U/ul Taq Stoffel fragment (Applied Biosystems)
- 12.5 ul 5 U/ul Ampligase (Epicentre)
- 4.75 ul dH2O

Add 1.0 ul of this mix to the hybridized probe reaction, and incubate at 60C for 10 hours.

Purify circularized probe/target nucleic acid products from un-reacted probes and genomic DNA

Prepare exonuclease reaction mixture:

- 21 ul fill-in / ligation reaction product
- 2.0 ul 10X exonuclease I buffer (New England Biolabs)
- 2.0 ul 20 U/ul exonuclease I (New England Biolabs)
- 2.0 ul 100 U/ul exonuclease III (New England Biolabs)

Incubate at 37C for 60 min, then heat-inactivate by incubating at 80C for 15 min. Immediately cool to 4C for storage.

Amplify circular material by PCR using primers specific to the 'constant region' of the probes

Prepare PCR mixture:

Petitioner Illumina, Inc., EX1004, p. 204

- 5.0 ul 10X Accuprime reaction buffer (Invitrogen)
- 1.5 ul 10 uM CP-2-FA (5'-GCACGATCCGACGGTAGTGT-3') (SEQ ID NO: 183)
- 1.5 ul 10 uM CP-2-RA (5'-CCGTAATCGGGAAGCTGAAG-3') (SEQ ID NO: 184)
- 0.4 ul 25 mM dNTP mix (Invitrogen)
- 2.0 ul heat-inactivated exonuclease reaction mix
- 1.5 ul 10X SybrGreen (Invitrogen)
- 0.4 ul 2.5 U/ul Accuprime Pfx polymerase (Invitrogen)
- 37.7 ul dH2O

Thermal cycle in real-time thermal cycler according to the following protocol, but stop cycling before amplification yield plateaus (generally 8-12 cycles):

16. 95C for 5 min
17. 95C for 30 sec
18. 58C for 60 sec
19. 72C for 60 sec
20. goto 2, N more times

Prepare a shotgun next-generation sequencing library for analysis

- Purify desired amplicon population from non-specific amplification products by gel extraction.
- Concatemerize amplicons into high-molecular weight products suitable for shearing
- Mechanically shear, using either a nebulizer, BioRuptor, Hydroshear, Covaris, or similar instrument. DNA should be sheared into fragments several hundred basepairs in length.
- Ligate adapters required for amplification by the sequencing platform used. If necessary, purify ligated product from unligated product and adapters.

Perform sequencing of library according to manufacturer's directions (e.g. Illumina,ABI, etc), reading the target sequence to determine abundance of the target nucleic acid.

Petitioner Illumina, Inc., EX1004, p. 205

Example 7: Use of MIPs, hybridization, and mutation-detection MIPs to genotype a set of 1000 targets

MIPs, hybridization, and mutation-detection MIPs are used to genotype a set of 1000 targets. The protocol permits detection of any of 50 specific known point mutations

First, separate MIP, hybridization, and mutation-detection MIP reactions are performed on a biological sample. A MIP capture reaction is performed essentially as described in Turner et al 2009, Nature methods 6:315-6. A set of MIPs is designed such that each probe in the set flanks one of the 1000 targets. Separately, a hybridization enrichment reaction is performed using the Agilent SureSelect procedure. Prior to selection, the genomic DNA to be enriched is converted into a shotgun sequencing library using Illumina's 'Fragment Library' kit and protocol. Agilent's web interface is used to design a set of probes which will hybridize to the target nucleic acids. Separately, a set of probes are designed (mutation-detection MIPs) which will form MIPs only if mutations (e.g., specific polymorphisms) are present. Each mutation-detection MIP has a 3'-most base identity that is specific for a single known mutation. A reaction with this set of mutation-detection MIPs is performed to selectively detect the presence of any mutant alleles.

Once all three reactions have been performed, the two MIP reactions are combined (e.g., at potentially non-equimolar ratios to further increase sensitivity of mutation detection) into a single tube, and run as one sample on the next-generation DNA sequencing instrument. The hybridization-enriched reaction is run as a separate sample on the next-generation DNA sequencing instrument. Reads from each 'sample' are combined by a software algorithm which forms a consensus diploid genotype at each position in the target set by evaluating the total coverage at each position, the origin of each read in that total coverage, the quality score of each individual read, and the presence (or absence) of any reads derived from mutation-specific MIPs overlapping the region.

It should be appreciated that the preceding examples are non-limiting and aspects of the invention may be implemented as described herein using alternative techniques and/or protocols that are available to one or ordinary skill in the art.

Petitioner Illumina, Inc., EX1004, p. 206

It will be clear that the methods may be practiced other than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present disclosure are possible in light of the above teachings and, therefore, are within the scope of the claims. Preferred features of each aspect of the disclosure are as for each of the other aspects mutatis mutandis. The documents including patents, patent applications, journal articles, or other disclosures mentioned herein are hereby incorporated by reference in their entirety. In the event of conflict, the disclosure of present application controls, other than in the event of clear error.

Petitioner Illumina, Inc., EX1004, p. 207

CLAIMS

What is claimed is:

1. A method of analyzing a plurality of genetic loci, the method comprising:
contacting each of a plurality of target nucleic acids with a probe set,
wherein each probe set comprises a plurality of different probes, each probe having a central region flanked by a 5' region and a 3' region that are complementary to nucleic acids flanking the same strand of one of a plurality of subregions of the target nucleic acid, wherein the subregions of the target nucleic acid are different, and wherein each subregion overlaps with at least one other subregion,
isolating a plurality of nucleic acids each having a nucleic acid sequence of a different subregion for each of the plurality of target nucleic acids, and
analyzing the isolated nucleic acids.

2. The method of claim 1, wherein at least one subregion entirely overlaps at least one other subregion.

3. The method of claim 1, wherein the sequences of the 5' region and the 3' region of each probe are, respectively, non-overlapping with the sequences of the 5' region and the 3' region of each other probe.

4. The method of claim 1, wherein the sequences of the 5' region and the 3' region of each probe are, respectively, different than the sequences of the 5' region and the 3' region of each other probe.

5. A method of analyzing a plurality of genetic loci, the method comprising:
contacting each of a plurality of target nucleic acids with a probe set,
wherein each probe set comprises a plurality of different probes, each probe having a central region flanked by a 5' region and a 3' region that are complementary to nucleic acids flanking the same strand of one of a plurality of subregions of the target nucleic acid, wherein the subregions of the target nucleic acid are different, and wherein a portion of the 5' region and a portion of the 3' region of a probe have, respectively, the sequence of the 5' region and the 3' region of a different probe,

Petitioner Illumina, Inc., EX1004, p. 208

isolating a plurality of nucleic acids each having a nucleic acid sequence of a different subregion for each of the plurality of target nucleic acids, and

analyzing the isolated nucleic acids.

6. The method of claim 1 or 5, wherein the target nucleic acids are selected from Table 1.

7. The method of claim 1 or 5, wherein the analyzing comprises genotyping a polymorphism of the target nucleic acid.

8. The method of claim 7, wherein the polymorphism is selected from Table 2.

9. A method of genotyping a subject, the method comprising:

determining the sequence of at least a threshold number of independently isolated nucleic acids, wherein the sequence of each isolated nucleic acid comprises a target nucleic acid sequence and a differentiator tag sequence,

wherein the threshold number is a number of unique combinations of target nucleic acid and differentiator tag sequences, wherein the isolated nucleic acids are identified as independently isolated if they comprise unique combinations of target nucleic acid and differentiator tag sequences, and wherein the target nucleic acid sequence is the sequence of a genomic locus of a subject.

10. The method of claim 9., wherein the isolated nucleic acids are molecular inversion probe capture products.

11. A method of genotyping a subject, the method comprising:

determining the sequence of molecular inversion probe capture products, each comprising a molecular inversion probe and a target nucleic acid, wherein the sequence of the molecular inversion probe comprises a differentiator tag sequence and, optionally, a primer sequence, and wherein the target nucleic acid is a captured genomic locus of a subject, and

Petitioner Illumina, Inc., EX1004, p. 209

genotyping the subject at the captured genomic locus based on the sequence of at least a threshold number of unique combinations of target nucleic acid and differentiator tag sequences of molecular inversion probe capture products.

12. A method of genotyping a subject, the method comprising:

obtaining molecular inversion probe capture products, each comprising a molecular inversion probe and a target nucleic acid, wherein the sequence of the molecular inversion probe comprises a differentiator tag sequence and, optionally, a primer sequence, wherein the target nucleic acid is a captured genomic locus of the subject,

amplifying the molecular inversion probe capture products, and

genotyping the subject by determining, for each target nucleic acid, the sequence of at least a threshold number of unique combinations of target nucleic acid and differentiator tag sequence of molecular inversion probe capture products.

13. The method of claim 12, wherein the obtaining comprises capturing target nucleic acids from a genomic sample of the subject with molecular inversion probes, each comprising a unique differentiator tag sequence.

14. The method of claim 12 or 13, wherein the capturing is performed under conditions wherein the likelihood of obtaining two or more molecular inversion probe capture products with identical combinations of target and differentiator tag sequences is equal to or less than a predetermined value, optionally wherein the predetermined value is about 0.05.

15. The method of any one of claims 12-14, wherein the threshold number for a specific target nucleic acid sequence is selected based on a desired statistical confidence for the genotype.

16. The method of any one of claims 12-14, further comprising determining a statistical confidence for the genotype based on the number of unique combinations of target nucleic acid and differentiator tag sequences.

Petitioner Illumina, Inc., EX1004, p. 210

17. A method of analyzing a plurality of genetic loci, the method comprising:

obtaining a plurality of molecular inversion probe capture products each comprising a molecular inversion probe and a target nucleic acid, wherein the sequence of the molecular inversion probe comprises a differentiator tag sequence and, optionally, a primer sequence,

amplifying the plurality of molecular inversion probe capture products,

determining numbers of occurrence of combinations of target nucleic acid and differentiator tag sequence of molecular inversion probe capture products in the amplified plurality, and

if the number of occurrence of a specific combination of target nucleic acid sequence and differentiator tag sequence exceeds a predetermined value, detecting bias in the amplification of the molecular inversion probe comprising the specific combination.

18. The method of claim 17, further comprising genotyping target sequences in the plurality, wherein the genotyping comprises correcting for bias, if detected.

19. The method of any of claims 9-18, wherein the target nucleic acid sequence is

a sequence of a gene selected from Table 1.

20. A method for determining whether a target nucleic acid has an abnormal length, the method comprising evaluating the capture efficiency of a target nucleic acid in a biological sample from a subject, wherein a capture efficiency that is different from a reference capture efficiency is indicative of the presence, in the biological sample, of a target nucleic acid having an abnormal length.

21. The method of claim 20, comprising determining the capture efficiency of a target region suspected of having a deletion or insertion and comparing the capture efficiency to a reference indicative of a normal capture efficiency.

22. The method of claim 20-21, wherein the capture efficiency is lower than the reference capture efficiency.

Petitioner Illumina, Inc., EX1004, p. 211

23. The method of claim 22, wherein the subject is identified as having an insertion in the target region.

24. The method of claim 20-21, wherein the capture efficiency is higher than the reference capture efficiency.

25. The method of claim 24, wherein the subject is identified as having a deletion in the target region.

26. The method of claim 23, wherein the subject is identified as being heterozygous for the insertion.

27. The method of claim 25, wherein the subject is identified as being heterozygous for the deletion.

28. The method of any prior claim, wherein a sub-target nucleic acid is captured from the target nucleic acid using a molecular inversion probe.

29. The method of claim 28, wherein the molecular inversion probe comprises a first targeting arm at its 5' end and a second targeting arm at its 3' end, wherein the first targeting arm is capable of specifically hybridizing to a first region flanking one end of the sub-target nucleic acid, and wherein the second targeting arm is capable of specifically hybridizing to a second region flanking the other end of the sub-target nucleic acid on the same strand of the target nucleic acid.

30. The method of claim 29, wherein the first and second targeting arms are between about 10 and about 100 nucleotides long.

31. The method of claim 30, wherein the first and second targeting arms are about 10-20, 20-30, 30-40, or 40-50 nucleotides long.

Petitioner Illumina, Inc., EX1004, p. 212

32. The method of claim 31, wherein the first and second targeting arms are about 20 nucleotides long.

33. The method of any one of claims 29-31, wherein the first and second targeting arms have the same length.

34. The method of any one of claims 29-33, wherein the hybridization Tms of the first and second targeting arms are similar.

35. The method of claim 34, wherein the hybridization Tms of the first and second targeting arms are within 2-5 C of each other.

36. The method of claim 34, wherein the hybridization Tms of the first and second targeting arms are identical.

37. The method of any one of claims 28-36, wherein the sub-target nucleic acid contains a nucleic acid repeat.

38. The method of claim 37, wherein the nucleic acid repeat is a dinucleotide or trinucleotide repeat.

39. The method of claim 37, wherein the sub-target nucleic acid contains 10-100 copies of the nucleic acid repeat in the absence of an abnormal increase or decrease in nucleic acid repeats.

40. The method of claim 37, wherein the sub-target nucleic acid is a region of the Fragile-X locus that contains a nucleic acid repeat.

41. The method of any one of claims 29-36, wherein one or both targeting arms hybridize to a region on the target nucleic acid that is immediately adjacent to a region of nucleic acid repeats.

Petitioner Illumina, Inc., EX1004, p. 213

42. The method of any one of claims 29-36, wherein one or both targeting arms hybridize to a region on the target nucleic acid that is separated from a region of nucleic acid repeats by a region that does not contain any nucleic acid repeats.

43. The method of any one of claims 29-36, wherein the molecular inversion probe further comprises a primer-binding region that can be used to sequence the captured sub-target nucleic acid and optionally the first and/or second targeting arm.

44. The method of any of the prior claims, wherein a plurality of different target nucleic acids are analyzed in a biological sample.

45. The method of claim 44, wherein the plurality of target nucleic acids are analyzed using a plurality of different molecular inversion probes.

46. The method of claim 45, wherein each different molecular inversion probe comprises a different pair of first and second targeting arms at each of the 3' and 5' ends.

47. The method of claim 46, wherein each different molecular inversion probe comprises the same primer-binding sequence.

48. The method of any prior claim, wherein the biological sample is a blood sample.

49. The method of any prior claim, wherein the biological sample is a tissue sample.

50. The method of any prior claim, wherein the capture efficiency is evaluated by determining an amount of target nucleic acid that is captured.

51. The method of claim 50, wherein the amount of target nucleic acid that is captured is determined by determining a number of independently captured target nucleic acid molecules.

Petitioner Illumina, Inc., EX1004, p. 214

52. The method of claim 50-51, wherein the amount of target nucleic acid that is captured is compared to a reference amount of captured nucleic acid.

53. The method of claim 52, wherein the reference amount is determined by determining a number of independently captured molecules of a reference nucleic acid.

54. The method of claim 53, wherein the reference nucleic acid is a nucleic acid of a different locus in the biological sample that is not suspected of containing a deletion or insertion.

55. The method of claim 53, wherein the reference nucleic acid is a nucleic acid of known size and amount that is added to the capture reaction.

56. The method of any prior claim, wherein a subject is identified as having an insertion or deletion in one or more alleles of the genetic locus if the capture efficiency is statistically significantly different that the reference capture efficiency.

57. A method for estimating the length of a target nucleic acid, the method comprising:

(i) contacting the target nucleic acid with a plurality of detection probes under conditions that permit hybridization of the detection probes to the target nucleic acid,

wherein each detection probe is a polynucleotide that comprises a first arm that hybridizes to a first region of the target nucleic acid and a second arm that hybridizes to a second region of the target nucleic acid,

wherein the first and second regions are on a common strand of the target nucleic acid, and

wherein the nucleotide sequence between the 5' end of the first region and the 3' end of the second region is the nucleotide sequence of a sub-target nucleic acid; and

(ii) capturing a plurality of sub-target nucleic acids that are hybridized with the plurality of detection probes; and

Petitioner Illumina, Inc., EX1004, p. 215

(iii) measuring the frequency of occurrence of a sub-target nucleic acid in the plurality of sub-target nucleic acids, wherein the frequency of occurrence of the sub-target nucleic acid in the plurality of sub-target nucleic acids is indicative of the length of the sub-target nucleic acid.

58. A method of increasing the sensitivity of a nucleic acid detection assay, the method comprising:

obtaining a first preparation of a target nucleic acid using a first preparative method on a biological sample,

obtaining a second preparation of a target nucleic acid using a second preparative method on the biological sample,

assaying the sequences obtained in both first and second nucleic acid preparations,

using the sequence information from both first and second nucleic acid preparations to determine the genotype of the target nucleic acid in the biological sample, wherein the first and second preparative methods have different systematic sequence biases.

59. The method of claim 58, wherein the first and second nucleic acid preparations are combined prior to performing a sequence assay.

60. The method of claim 58, wherein separate sequence assays are performed on the first and second nucleic acid preparations and the sequence information from both assays are combined to determine the genotype of the target nucleic acid in the biological sample.

61. The method of claim 58, wherein the first preparative method is a PCR-based, a hybridization-based, or a circular probe-based preparative method.

62. The method of claim 61, wherein the second preparative method is a PCR-based, a hybridization-based, or a circular probe-based preparative method, provided that the second method is different from the first method.

Petitioner Illumina, Inc., EX1004, p. 216

63. A method of obtaining a nucleic acid preparation that is representative of the nucleic acid in a biological sample, the method comprising

obtaining a first nucleic acid preparation using a first preparative method on a biological sample,

obtaining a second nucleic acid preparation using a second preparative method on the biological sample, and

combining the first and second nucleic acid preparations to obtain a combined preparation that is representative of nucleic acid in the biological sample.

64. The method of claim 1 or 6, further comprising performing a third nucleic acid preparation using a third preparative method that is different from the first and second preparative methods, wherein the first, second, and third preparative methods all have different systematic sequence biases.

65. The method of claim any of the prior claims, wherein the different preparative methods are used for a plurality of different loci in the biological sample to increase the sensitivity of a multiplex nucleic acid analysis.

66. A method of genotyping a nucleic acid in a biological sample, the method comprising:

obtaining a nucleic acid preparation using a preparative method on a biological sample,

sequencing a target nucleic acid of the nucleic acid preparation,

and performing a molecular inversion probe capture reaction on the biological sample, wherein a molecular inversion probe capture of the target nucleic acid in the biological sample is indicative of the presence of a polymorphism in the target nucleic acid,

genotyping the target nucleic acid based on the results of the sequencing and the capture reaction.

67. The method of any of the prior claims, wherein the target nucleic acid has a sequence of a gene selected from Table 1.

Petitioner Illumina, Inc., EX1004, p. 217

ABSTRACT

Aspects of the invention relates to methods and compositions that are useful to reduce bias and increase the reproducibility of multiplex analysis of genetic loci. In some configurations, predetermined preparative steps and/or nucleic acid sequence analysis techniques are used in multiplex analyses for a plurality of genetic loci in a plurality of samples.

Petitioner Illumina, Inc., EX1004, p. 218

FIG. 1

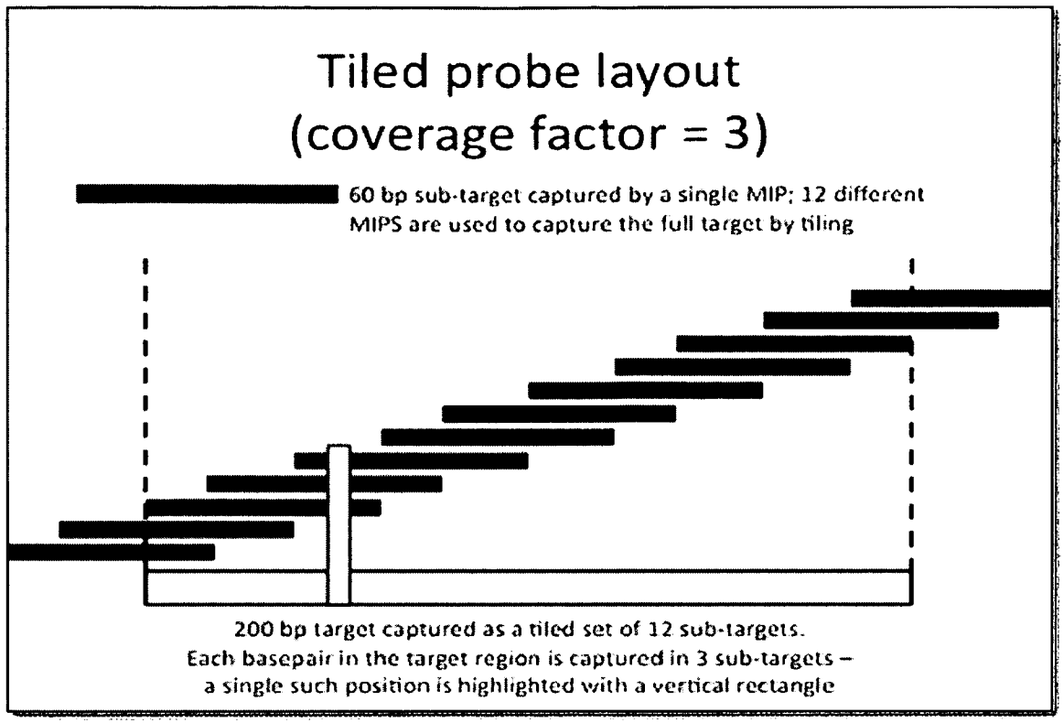

Tiled probe layout
(coverage factor = 3)
60 bp sub-target captured by a single MIP; 12 different MIPS are used to capture the full target by tiling
200 bp target captured as a tiled set of 12 sub-targets.
Each basepair in the target region is captured in 3 sub-targets –
a single such position is highlighted with a vertical rectangle

FIG. 2

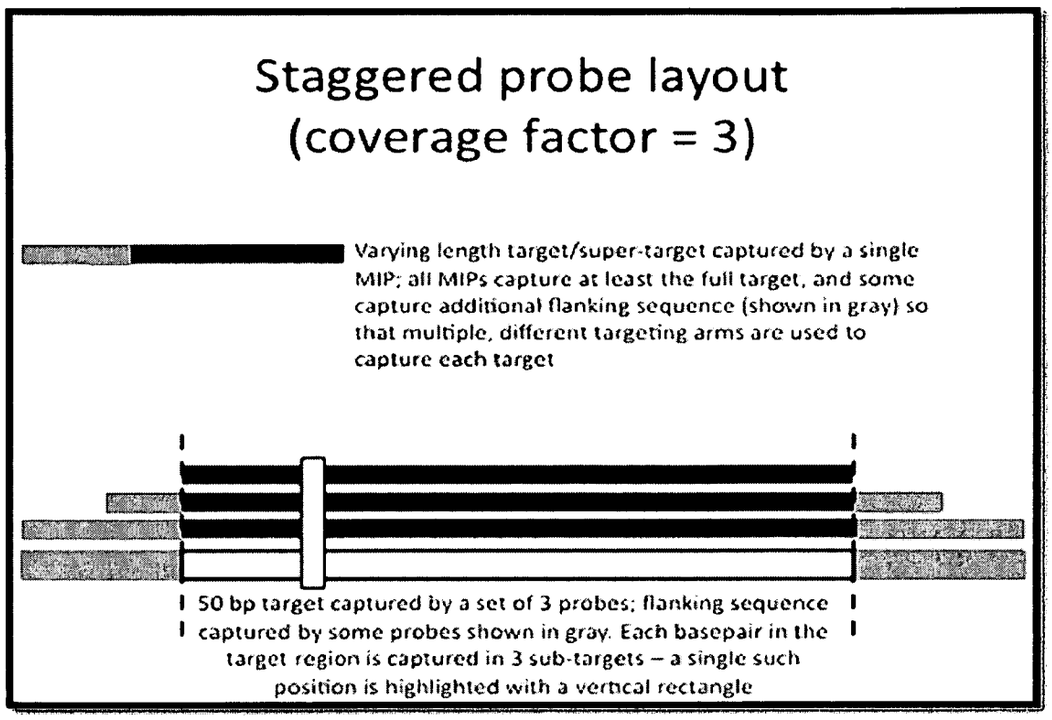

Staggered probe layout
(coverage factor = 3)
Varying length target/super-target captured by a single MIP; all MIPs capture at least the full target, and some capture additional flanking sequence (shown in gray) so that multiple, different targeting arms are used to capture each target
50 bp target captured by a set of 3 probes; flanking sequence captured by some probes shown in gray. Each basepair in the target region is captured in 3 sub-targets – a single such position is highlighted with a vertical rectangle

FIG. 3

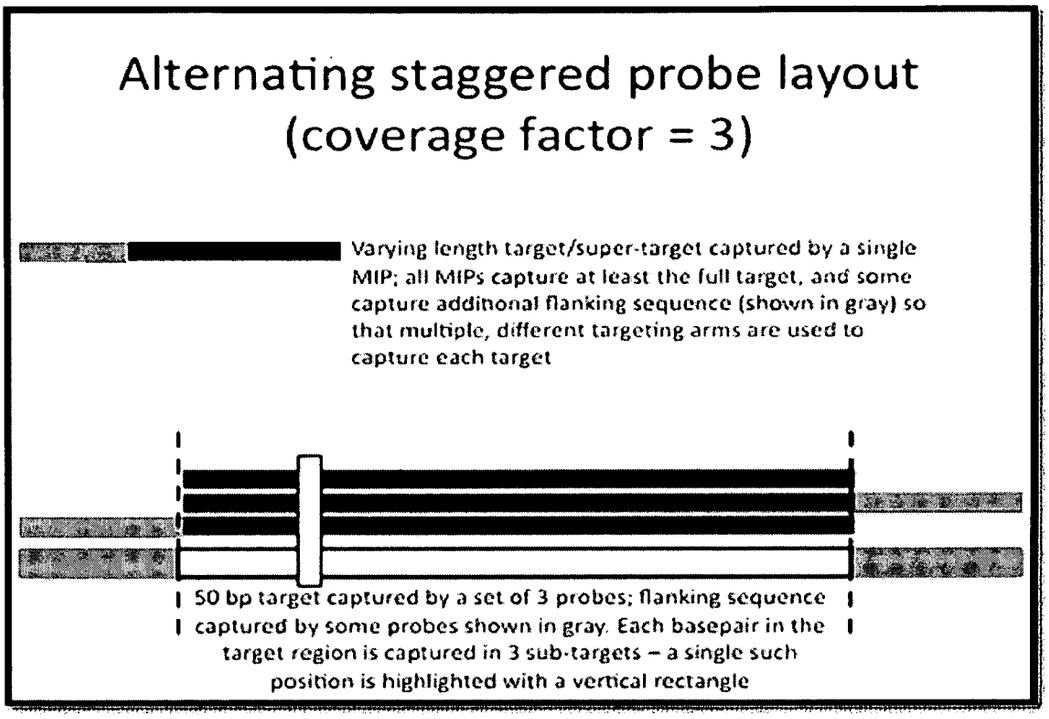

Alternating staggered probe layout
(coverage factor = 3)
Varying length target/super-target captured by a single MIP; all MIPs capture at least the full target, and some capture additional flanking sequence (shown in gray) so that multiple, different targeting arms are used to capture each target
50 bp target captured by a set of 3 probes; flanking sequence captured by some probes shown in gray. Each basepair in the target region is captured in 3 sub-targets – a single such position is highlighted with a vertical rectangle

FIG. 4A

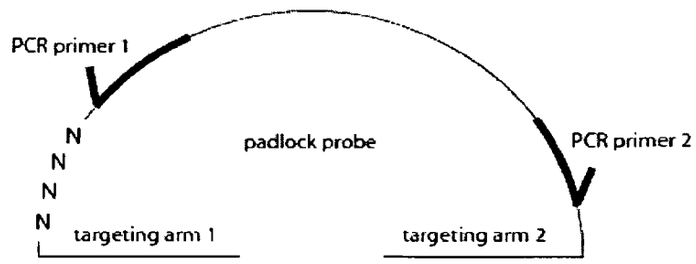


FIG. 4B

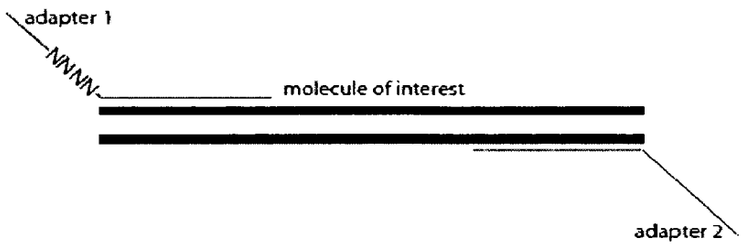


FIG. 4C

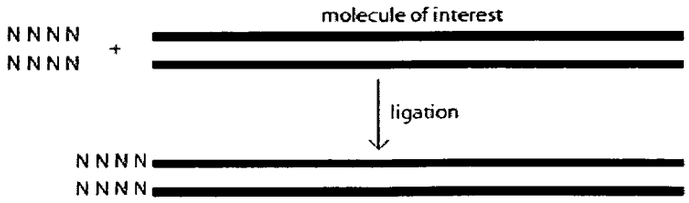


FIG. 5

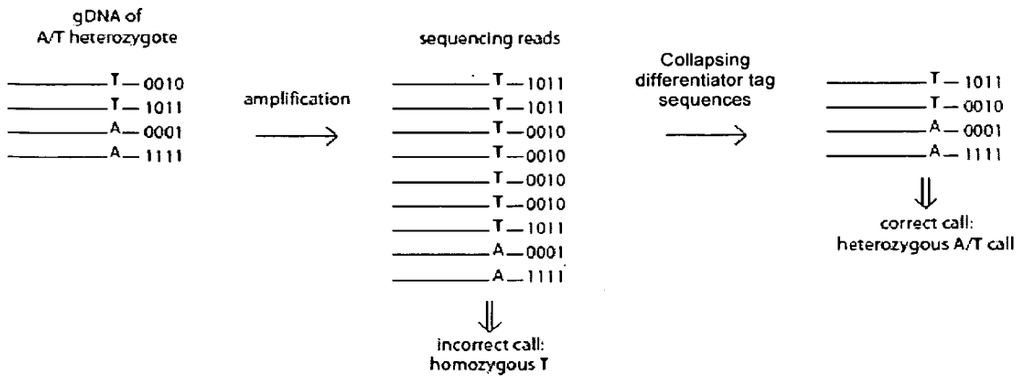

gDNA of
A/T heterozygote
T — 0010
T — 1011
A — 0001
A — 1111
amplification
sequencing reads
T — 1011
T — 1011
T — 0010
T — 0010
T — 0010
T — 0010
T — 1011
A — 0001
A — 1111
incorrect call:
homozygous T
Collapsing
differentiator tag
sequences
T — 1011
T — 0010
A — 0001
A — 1111
correct call:
heterozygous A/T call

FIG. 6

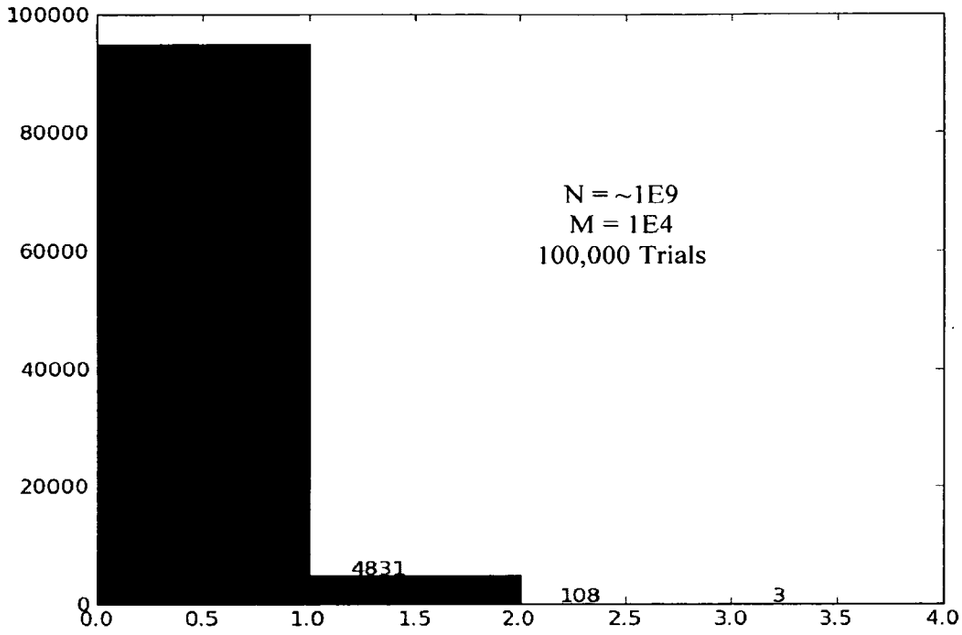

100000
80000
60000
40000
20000
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
N = ~1E9
M = 1E4
100,000 Trials
4831
108
3
Independent Occurrences of
Target:Differentiator Tag Sequence
100,000 Trials

FIG. 7

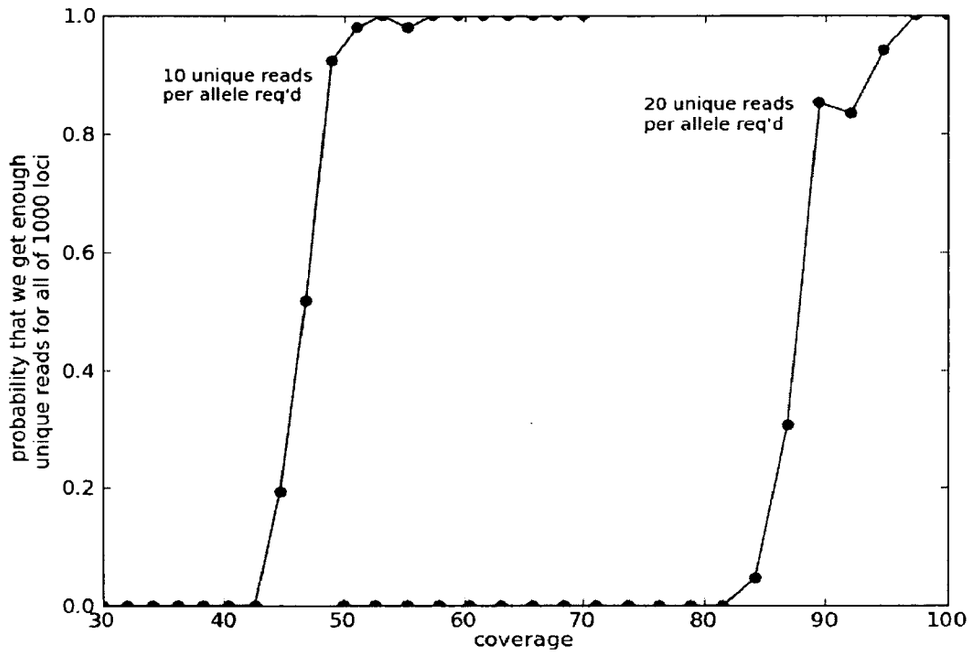

1.0
0.8
0.6
0.4
0.2
0.0
probability that we get enough unique reads for all of 1000 loci
10 unique reads per allele req'd
20 unique reads per allele req'd
30
40
50
60
70
80
90
100
coverage

FIG. 8

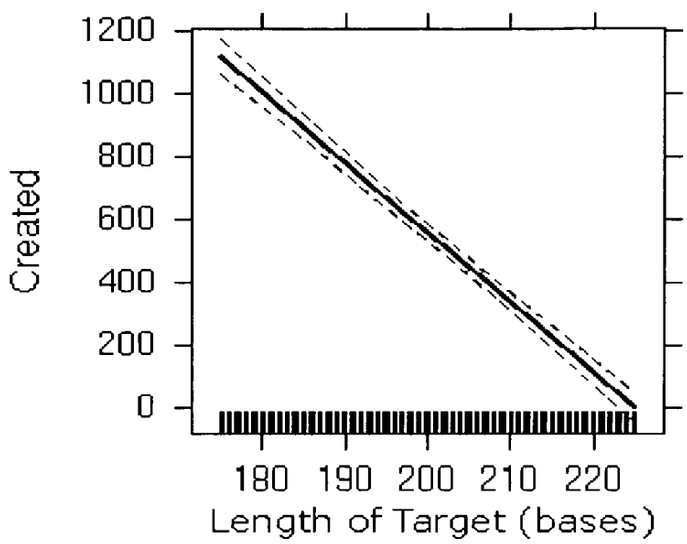

1200
1000
800
600
400
200
0
Created
180
190
200
210
220
Length of Target (bases)

FIG. 9

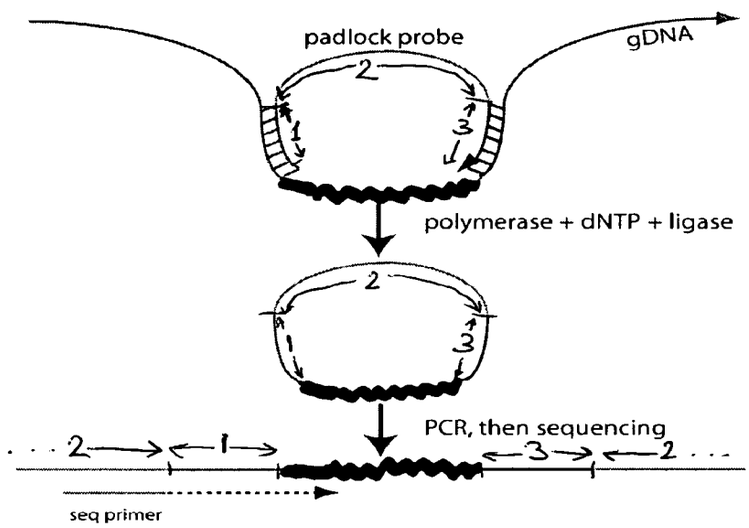

padlock probe
gDNA
2
1
3
polymerase + dNTP + ligase
2
1
3
PCR, then sequencing
2
1
3
2
seq primer

FIG. 10

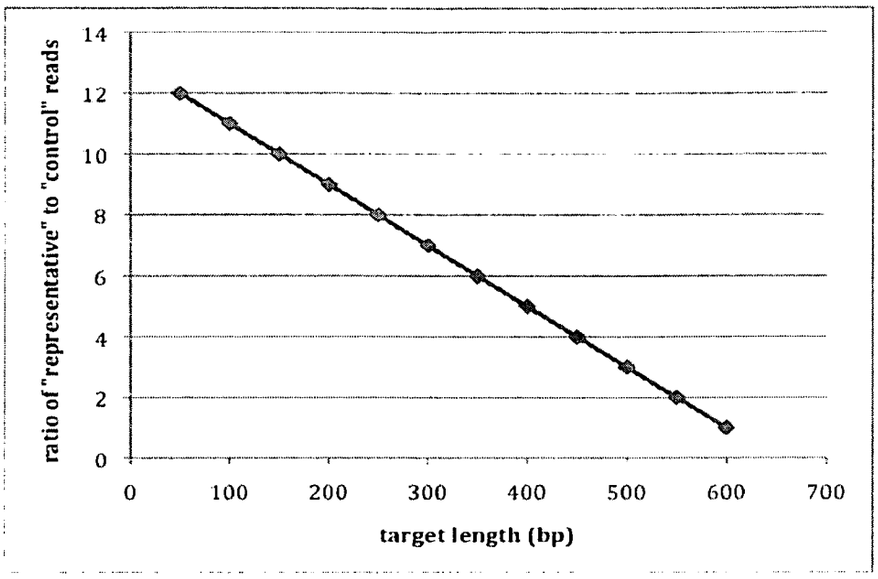

ratio of "representative" to "control" reads
14
12
10
8
6
4
2
0
0
100
200
300
400
500
600
700
target length (bp)

FIG. 11A

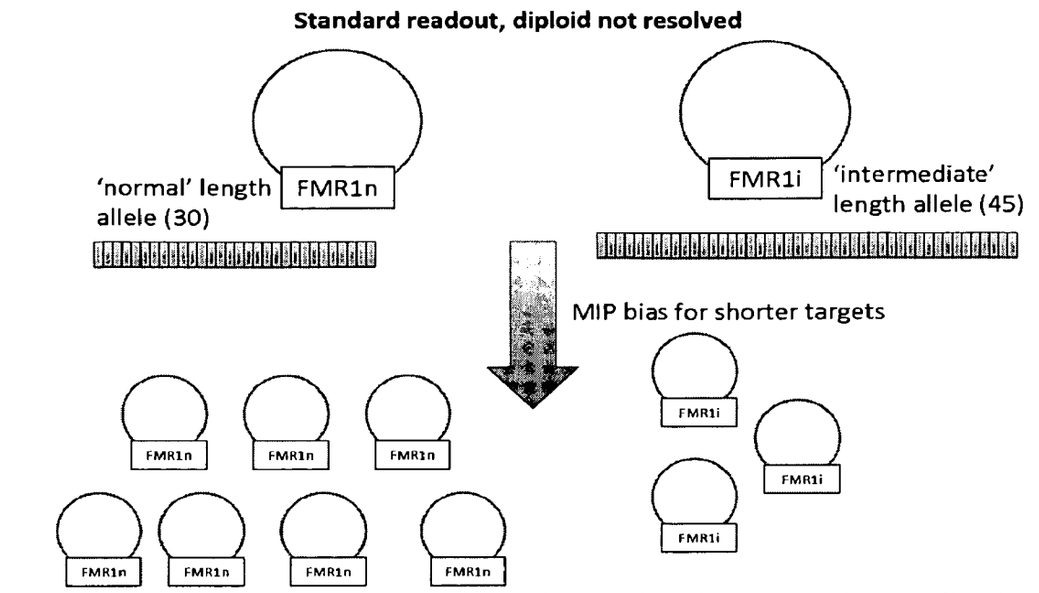

Standard readout, diploid not resolved
'normal' length allele (30)
FMR1n
FMR1i
'intermediate' length allele (45)
MIP bias for shorter targets
FMR1n
FMR1n
FMR1n
FMR1n
FMR1n
FMR1n
FMR1n
FMR1i
FMR1i
FMR1i
Readout: sum of alleles appears as 'long' end of normal length range, not intermediate het

FIG. 11B

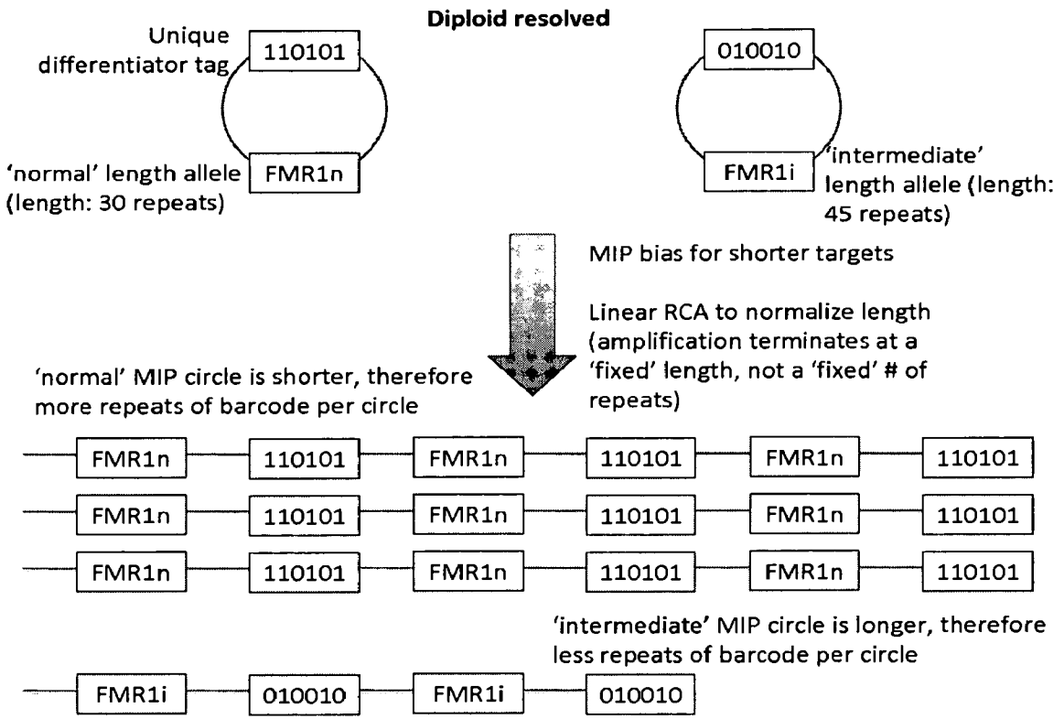

Diploid resolved
Unique differentiator tag
110101
FMR1n
'normal' length allele (length: 30 repeats)
010010
FMR1i
'intermediate' length allele (length: 45 repeats)
MIP bias for shorter targets
Linear RCA to normalize length (amplification terminates at a 'fixed' length, not a 'fixed' # of repeats)
'normal' MIP circle is shorter, therefore more repeats of barcode per circle
FMR1n 110101 FMR1n 110101 FMR1n 110101
FMR1n 110101 FMR1n 110101 FMR1n 110101
FMR1n 110101 FMR1n 110101 FMR1n 110101
'intermediate' MIP circle is longer, therefore less repeats of barcode per circle
FMR1i 010010 FMR1i 010010

FIG. 11C

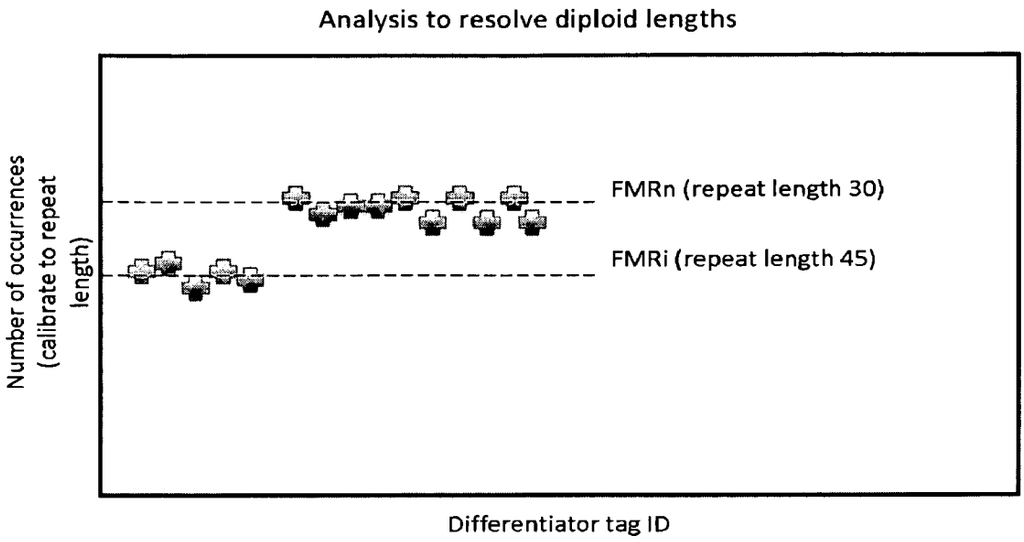

Analysis to resolve diploid lengths
Number of occurrences (calibrate to repeat length)
FMRn (repeat length 30)
FMRi (repeat length 45)
Differentiator tag ID

FIG. 12

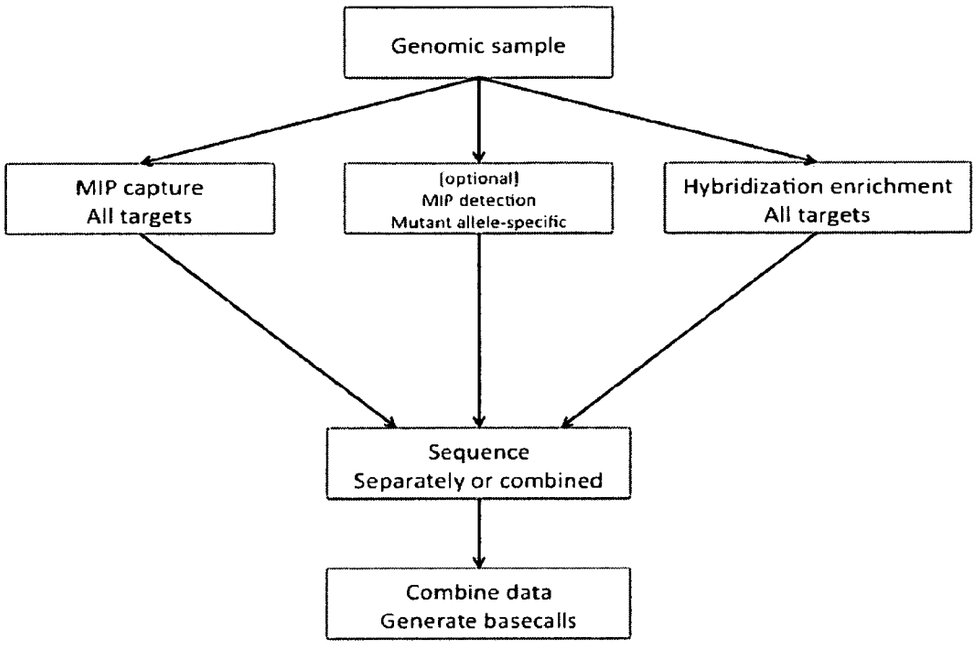

Genomic sample
MIP capture
All targets
(optional)
MIP detection
Mutant allele-specific
Hybridization enrichment
All targets
Sequence
Separately or combined
Combine data
Generate basecalls

FIG. 13

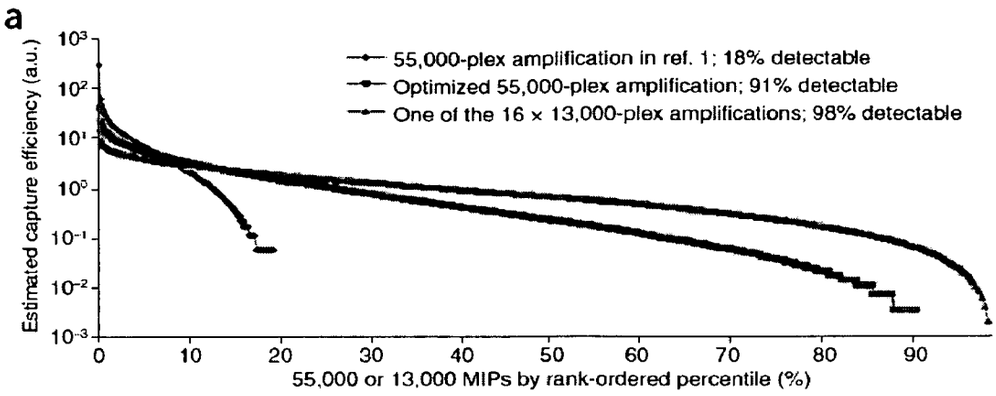
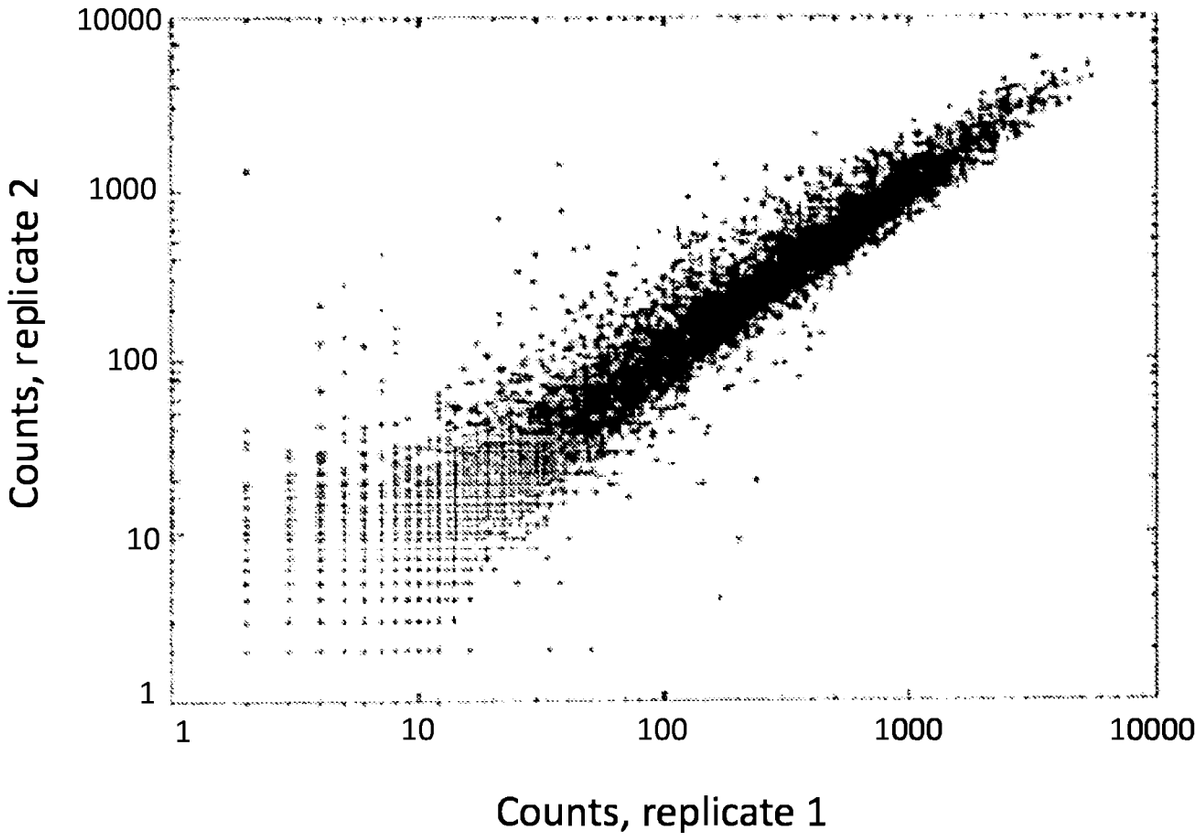

a
Estimated capture efficiency (a.u.)
10³
10²
10¹
10⁰
10⁻¹
10⁻²
10⁻³
0
10
20
30
40
50
60
70
80
90
55,000 or 13,000 MIPs by rank-ordered percentile (%)
55,000-plex amplification in ref. 1; 18% detectable
Optimized 55,000-plex amplification; 91% detectable
One of the 16 × 13,000-plex amplifications; 98% detectable

FIG. 14


10000
1000
100
10
1
Counts, replicate 2
1
10
100
1000
10000
Counts, replicate 1

SEQUENCE LISTING

<110> Good Start Genetics, Inc.
Porreca, Gregory
Laserson, Uri

<120> METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS

<130> GSGE-002/04US 29258/78

<140> US 15/231,687
<141> 2016-08-08

<160> 184

<170> PatentIn version 3.5

<210> 1
<211> 52
<212> DNA
<213> Homo sapiens

<400> 1
tcacatcacc aagttaaaaa aaaaaarggg gcgggggggc agaatgaaaa tt 52

<210> 2
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 2
aaacaaggat gaattaagtt ttttttaaa aaagaaacat ttggtaaggg ga 52

<210> 3
<211> 52
<212> DNA
<213> Homo sapiens

<400> 3
atcaccaagt taaaaaaaaa aaagggrcgg gggggcagaa tgaaaattgc at 52

<210> 4
<211> 52
<212> DNA
<213> Homo sapiens

<400> 4
ctattgaacc agaaccaaac aggaatrcca tagcattttg taaactaaac tg 52

<210> 5
<211> 52
<212> DNA
<213> Homo sapiens

<400> 5
caggagttca agaccagcct actaaamcac acacacacac acacacacac ac 52

<210> 6
<211> 52
<212> DNA
<213> Homo sapiens

<400> 6
gattaaataa tagtgtttat gtacccsgct tataggagaa gagggtgtgt gt 52

<210> 7
<211> 52
<212> DNA
<213> Homo sapiens

<400> 7
attgttatct tttcatataa ggtaacwgag gcccagagag attaaataac at 52

<210> 8
<211> 52
<212> DNA
<213> Homo sapiens

<400> 8
taattttaat taagtaaatt taattgrtag ataaataagt agataaaaaa ta 52

<210> 9
<211> 52
<212> DNA
<213> Homo sapiens

<400> 9
gtataaaaaa aaaaaaaaaa aaagttwgaa tgttttcttg cattcagagc ct 52

<210> 10
<211> 52
<212> DNA
<213> Homo sapiens

<400> 10
atactaaaaa tttaaagttc tcttgcrata tattttctta atatcttaca tc 52

<210> 11
<211> 52
<212> DNA
<213> Homo sapiens

<400> 11
tgctgggatt acaggcgtga gccaccrcgc ctggcctgat gggacatatt tt 52

<210> 12
<211> 52
<212> DNA
<213> Homo sapiens

<400> 12
ctacaatata agtatagtat tgcaaamcca tcaggaaggg tgttaactat tt 52

<210> 13
<211> 54
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(29)
<223> This region may or may not be present in its entirety

<400> 13
gttgtctcca aacttttttt caggtgagaa ggtggccaac cgagcttcgg aaag 54

<210> 14
<211> 52
<212> DNA
<213> Homo sapiens

<400> 14
ttttttcata aaagattata taaaggmtat tgcttttgaa tcacaaacac ta 52

<210> 15
<211> 52
<212> DNA
<213> Homo sapiens

<400> 15
atctagtgag cagtcaggaa agagaayttc cagatcctgg aaatcagggt ta 52

<210> 16
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 16
tagagtataa aaaaaaaaaa aaaaaaagtt tgaatgtttt cttgcattca ga 52

<210> 17
<211> 52
<212> DNA
<213> Homo sapiens

<400> 17
tgcaaatgcc aactatcaaa gatattsgag tatactgtca ataaacttca ta 52

<210> 18
<211> 52
<212> DNA
<213> Homo sapiens

<400> 18
tcctcaacag ttagaaacaa tattttsagt gatttcccat gccaacttta ct 52

<210> 19
<211> 52
<212> DNA

<213> Homo sapiens

<400> 19
tttttggtat tgttgttaaa taagtgrgaa ttcaatacag tataatgtct gt 52

<210> 20
<211> 87
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(72)
<223> This region may or may not be present in its entirety

<400> 20
cttgaaatcg gatatatata tatatatgta tatatatata tatatatata tatatatata 60

tagtattatc cctgttttca cagtttt 87

<210> 21
<211> 52
<212> DNA
<213> Homo sapiens

<400> 21
agaggggctg tgaaggacac caaggargag actaagagcc aggagggaaa ac 52

<210> 22
<211> 52
<212> DNA
<213> Homo sapiens

<400> 22
tagagtttat tagcttttac tactctrctt agttactttg tgttacagaa ta 52

<210> 23
<211> 53
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(28)

<223> This region may or may not be present in its entirety

<400> 23
actagtgatg atgagcttct tttcatatgt ttgttggctg cataaatgtc ttc 53

<210> 24
<211> 52
<212> DNA
<213> Homo sapiens

<400> 24
gcagagaaaa gaggggctgt gaaggasacc aaggaggaga ctaagagcca gg 52

<210> 25
<211> 52
<212> DNA
<213> Homo sapiens

<400> 25
ttcagaggtc taccactggt gcatackcta atcacagtgt cgaaaatttt ac 52

<210> 26
<211> 52
<212> DNA
<213> Homo sapiens

<400> 26
aagaaagagc aaaagagggc aaacttytca tacatttttg atgtcgaaac ca 52

<210> 27
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 27
cctaaagttt aaaaagaaaa aaaaaaagga agaaggaatt aaaaatccaa ag 52

<210> 28
<211> 52

<212> DNA
<213> Homo sapiens

<400> 28
gtgtgtgtgt gtatatatat atatatwtat atatttttt tttcctgagc ca 52

<210> 29
<211> 52
<212> DNA
<213> Homo sapiens

<400> 29
aaactgttga tgttttcatt tatttasatc attggaaaac tttagattct ag 52

<210> 30
<211> 52
<212> DNA
<213> Homo sapiens

<400> 30
tttatccatt cttaaccaga acagacrttt tttcagagct ggtccaggaa aa 52

<210> 31
<211> 52
<212> DNA
<213> Homo sapiens

<400> 31
ttgaaatcgg atatatatat atatatrtat atatatatat atatatatat at 52

<210> 32
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 32
tagtttttta tttcctcata ttattttcag tggctttttc ttccacatct tt 52

<210> 33

<211> 52
<212> DNA
<213> Homo sapiens

<400> 33
acatagttct cagtggtaca actacargtg atttctcttt tcttatttct gg 52

<210> 34
<211> 52
<212> DNA
<213> Homo sapiens

<400> 34
agagcaatgg catcccttgt cttgtgytat acaggatgca gcaatttata gg 52

<210> 35
<211> 52
<212> DNA
<213> Homo sapiens

<400> 35
ttctgtctac ataagatgtc atactaratt atcttttcca gcatgcattc ag 52

<210> 36
<211> 52
<212> DNA
<213> Homo sapiens

<400> 36
cagggtggca tgttaggcag tgcttaraat aaatgagttg gttatacaag ta 52

<210> 37
<211> 53
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(28)
<223> This region may or may not be present in its entirety

<400> 37
aggacacaca cacacacaca cacacacatg cacacacatt taaatagatg cat 53

<210> 38
<211> 52
<212> DNA
<213> Homo sapiens

<400> 38
taaaaaattg gtataatgaa attgcayttg tagtctttgg acatttaaat cc 52

<210> 39
<211> 52
<212> DNA
<213> Homo sapiens

<400> 39
tttcaatact taagaggtac gcagagraaa gaggggctgt gaaggacacc aa 52

<210> 40
<211> 52
<212> DNA
<213> Homo sapiens

<400> 40
cagcaatgaa tattttgagg ctgaggygct gaggggtaaa attgcagcct gg 52

<210> 41
<211> 55
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(30)
<223> This region may or may not be present in its entirety

<400> 41
ttatggttta tatttttgtg tcttctcttt aacacatctt ttctagcaga attca 55

<210> 42
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature

<222> (27)..(27)
<223> May or may not be present

<400> 42
gtattttagt ttttttttt tgtttgtttt gttttgtttt gttttgtttt tg 52

<210> 43
<211> 52
<212> DNA
<213> Homo sapiens

<400> 43
tgggtgactc catttttact tttagtytgg tctgttgagg cctcgtgaga ga 52

<210> 44
<211> 55
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(30)
<223> This region may or may not be present in its entirety

<400> 44
tattttcatg tattttagtt ttttttttt gtttgttttg ttttgttttg ttttg 55

<210> 45
<211> 52
<212> DNA
<213> Homo sapiens

<400> 45
gtgtgtgtgt atatatatat atatatwtat attttttttt tcctgagcca aa 52

<210> 46
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 46
aacctattag catgtctggc agaaaaatag atacttaata aatttcttaa at 52

<210> 47
<211> 52
<212> DNA
<213> Homo sapiens

<400> 47
gaggcttaga cagtttaagt aactcargca tggttacaca actagctagg gc 52

<210> 48
<211> 52
<212> DNA
<213> Homo sapiens

<400> 48
gtgtgagtat tatgagacca tatgttrgga gattttattt ggtattgagg at 52

<210> 49
<211> 52
<212> DNA
<213> Homo sapiens

<400> 49
gaaaccccac cccttctata gttttcyctt taatatttac aatggaacca tt 52

<210> 50
<211> 52
<212> DNA
<213> Homo sapiens

<400> 50
catatatata tagtgtgtgt gtgtgtrtat atatatatat atatatattt tt 52

<210> 51
<211> 53
<212> DNA
<213> Homo sapiens

<400> 51
cgagcgagag agagagagag agagagagga gccggcgcga gaactacgca tgc 53

<210> 52
<211> 52
<212> DNA
<213> Homo sapiens

<400> 52
ggcaggtaat atctagtacc ttacttwtat ttcctgagca cattctacat tt 52

<210> 53
<211> 52
<212> DNA
<213> Homo sapiens

<400> 53
ggccaggaat gggagtgctt aggtgcrgag gtggcactgt tcccgcagct gc 52

<210> 54
<211> 52
<212> DNA
<213> Homo sapiens

<400> 54
tgaaaactcc atcccctcag ggtcatyaga tgaagagaag accacagggg tt 52

<210> 55
<211> 52
<212> DNA
<213> Homo sapiens

<400> 55
tgtaggtaag ggtcacatgt gggagasgca gctgtgggta ggtcagccct gt 52

<210> 56
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 56
ccaagaaggc gccattacac tccagcctgg gcgacagggc gagactccct ca 52

<210> 57
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 57
tgccacaccc agctaatttg tgtgtggtat gtgtgtgtat gtatgtgtgt gt 52

<210> 58
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 58
gttcctccag taatttttt ttttttggt tttgagacag agtcttgccc tg 52

<210> 59
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 59
atcatgccca gataattttt ttttttgta ttttagtaga cacagggttt ca 52

<210> 60
<211> 53
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(28)
<223> This region may or may not be present in its entirety

<400> 60
cgagcgagag agagagagag agagagagga gccggcgcga gaactacgca tgc 53

<210> 61
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 61
cctgtgaggg gcacattcct tagtagctaa ggagttgggg gtgtgaagat cc 52

<210> 62
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 62
acaggctact ggctgggccc aggcaaaggg ggccttggca ggaaaagttc ct 52

<210> 63
<211> 53
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(28)

<223> This region may or may not be present in its entirety

<400> 63
gcgagagaga gagagagaga gagaggagag ccggcgcgag aactacgcat gcg 53

<210> 64
<211> 52
<212> DNA
<213> Homo sapiens

<400> 64
aagaagaaaa ataaaaagaa agtgggycag accgagagaa caggaagcct ga 52

<210> 65
<211> 52
<212> DNA
<213> Homo sapiens

<400> 65
aaggacaaag gcaaagagac aaaggckcaa cactgggggt ccccagagag tg 52

<210> 66
<211> 52
<212> DNA
<213> Homo sapiens

<400> 66
tacctagtca cttcctgcct ccatggygca aaaggggatg ggtgtgcctc tt 52

<210> 67
<211> 52
<212> DNA
<213> Homo sapiens

<400> 67
ctcttccgag gttccaccct gaacacyttc ctgctccctc gtggtgtaga gt 52

<210> 68
<211> 52
<212> DNA
<213> Homo sapiens

<400> 68
ttctgactgg caaccagccc cactctytgg gagccctcag gaatgaactt gc 52

<210> 69
<211> 52
<212> DNA
<213> Homo sapiens

<400> 69
gctcagcctc ccaggctgga gtgcagwggc gcgatctcgg ctcaccgcaa cc 52

<210> 70
<211> 52
<212> DNA
<213> Homo sapiens

<400> 70
catgaaccac atcaaatgag atttagyggg agtggcacac acagtcatga cc 52

<210> 71
<211> 52
<212> DNA
<213> Homo sapiens

<400> 71
aagcagccct ggggagtcgg ggcgggrcct ggattggaaa agagacggtc ac 52

<210> 72
<211> 52
<212> DNA
<213> Homo sapiens

<400> 72
ctccaagttc tgggagcaga gtgtgcrgct aggctcctgg gatcgaggga tg 52

<210> 73
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 73
gttcctccag taatttttt ttttttkgtt ttgagacaga gtcttgccct gt 52

<210> 74
<211> 52
<212> DNA
<213> Homo sapiens

<400> 74
ctagtacctt acttccctca agttcawtca tctcacagat atttcctgag ca 52

<210> 75
<211> 52
<212> DNA
<213> Homo sapiens

<400> 75
aattagccgt gcgtggtggc gggtgcytgt aatcccacgt acttgggagg ct 52

<210> 76
<211> 52
<212> DNA
<213> Homo sapiens

<400> 76
ccatggccag ccggggaggg gacgggmaca cacagaccca cacagagact ca 52

<210> 77
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (77)..(77)
<223> May or may not be present

<400> 77
agcgagagag agagagagag agagagrgag ccggcgcgag aactacgcat gc 52

<210> 78
<211> 52
<212> DNA
<213> Homo sapiens

<400> 78

cgtagcagtt agcagatgat aggcggbgaa atcttatttc acagggcatt aa 52

<210> 79
<211> 52
<212> DNA
<213> Homo sapiens

<400> 79
ctggccctgg tgacagtggg gctgtgygtg gggccagagc cttctcagag gt 52

<210> 80
<211> 52
<212> DNA
<213> Homo sapiens

<400> 80
cagatactgg ccctggtgac agtgggrctg tgcgtggggc cagagccttc tc 52

<210> 81
<211> 52
<212> DNA
<213> Homo sapiens

<400> 81
gacagatact ggccctggtg acagtgkggc tgtgcgtggg gccagagcct tc 52

<210> 82
<211> 53
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(28)
<223> This region may or may not be present in its entirety

<400> 82
ggctcctctc tctctctctc tctctctcgc tcgctctctc gctctctcgc tct 53

<210> 83
<211> 52
<212> DNA
<213> Homo sapiens

<400> 83
gtttcagacc agcctggcca acatagygaa accccatctc tactaaaaat aa 52

<210> 84
<211> 52
<212> DNA
<213> Homo sapiens

<400> 84
agtgggcgat tggatggagc tgagtayggg gcccatccag gctaatcaca cc 52

<210> 85
<211> 52
<212> DNA
<213> Homo sapiens

<400> 85
ccgggctccg tgaatgtttg tcacatstct gaagaacgta tgaattacat aa 52

<210> 86
<211> 52
<212> DNA
<213> Homo sapiens

<400> 86
gaatcccaac cccgacgctc gtcgccsggc tccgtgaatg tttgtcacat gt 52

<210> 87
<211> 52
<212> DNA
<213> Homo sapiens

<400> 87
gcgaatccca accccgacgc tcgtcgycgg gctccgtgaa tgtttgtcac at 52

<210> 88
<211> 52
<212> DNA
<213> Homo sapiens

<400> 88
tgggcagaag tcagggtcca aagaaakggc aaagaaaagt gtcagtggct ca 52

<210> 89

<211> 52
<212> DNA
<213> Homo sapiens

<400> 89
taagaaagac gtttttgatt tttttcrgac ttctctggct ccactaccct gc 52

<210> 90
<211> 52
<212> DNA
<213> Homo sapiens

<400> 90
ctgattcctg gccaggagcg gtggctyacg cctgtaatcc cagcgctttg gg 52

<210> 91
<211> 52
<212> DNA
<213> Homo sapiens

<400> 91
taaaaatgct gattcctggc caggagyggt ggctcacgcc tgtaatccca gc 52

<210> 92
<211> 52
<212> DNA
<213> Homo sapiens

<400> 92
tttaaaaatg ctgattcctg gccaggmgcg gtggctcacg cctgtaatcc ca 52

<210> 93
<211> 52
<212> DNA
<213> Homo sapiens

<400> 93
caagacctgt caaagatctg agaaatwtta cccgacttac aagctaacca tt 52

<210> 94
<211> 52
<212> DNA
<213> Homo sapiens

<400> 94

actgtaataa gtgctgtaaa agaaatrcac aaaataatat agcagagggt at 52

<210> 95
<211> 52
<212> DNA
<213> Homo sapiens

<400> 95
cttgaggtca ggagttcaag accagtytgg gcaacatggg gaaaaccttg tc 52

<210> 96
<211> 52
<212> DNA
<213> Homo sapiens

<400> 96
aggttgcagt gagccgagat catgccrttg cactccagcc ggggcaacaa aa 52

<210> 97
<211> 52
<212> DNA
<213> Homo sapiens

<400> 97
acaagtgtct tgaaattatc tgtgatytgc tatagagcaa tacttttgta aa 52

<210> 98
<211> 52
<212> DNA
<213> Homo sapiens

<400> 98
gtgggtatat gcagctctat gcactayctg ctcatttatt tggtaaatct aa 52

<210> 99
<211> 59
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(34)
<223> This region may or may not be present in its entirety

<400> 99
tgtgtgtgtg tgcgtgtgtg tgtgtgtgtg tgtgatcata agagtggctg cagcaaact 59

<210> 100
<211> 52
<212> DNA
<213> Homo sapiens

<400> 100
agtctggagt gcaatggtgc aatctcrgct cactgcagcc tccacctccg gg 52

<210> 101
<211> 52
<212> DNA
<213> Homo sapiens

<400> 101
ctcctaatgg atatttccta aattttkctg aacagaattt aacttgagct gg 52

<210> 102
<211> 52
<212> DNA
<213> Homo sapiens

<400> 102
atttaaaaat ggatttctag aaaaacratc acatacttga atattttagc aa 52

<210> 103
<211> 52
<212> DNA
<213> Homo sapiens

<400> 103
ctataaatgg gtagcatgag ggattcragg aggtggctga aagaagcacg ta 52

<210> 104
<211> 52
<212> DNA
<213> Homo sapiens

<400> 104
aagaaaccaa gcatagtaga gtgttaraaa accaaagcaa ctaaacaact gt 52

<210> 105

<211> 52
<212> DNA
<213> Homo sapiens

<400> 105
cggggctcag aacttgtaac agaaaawtaa aatatactcc actcaaggga at 52

<210> 106
<211> 57
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(32)
<223> This region may or may not be present in its entirety

<400> 106
tactacactt cacggatact gtacttgtac ttttttttcca aattgaaggt ttttggc 57

<210> 107
<211> 63
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(38)
<223> This region may or may not be present in its entirety

<400> 107
ttgtttttgt ttttgttttt gtttttgttt ttgtttttttg agatggagtc tcgctctgtc 60

gcc 63

<210> 108
<211> 57
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(32)
<223> This region may or may not be present in its entirety

<400> 108
tttgccttac tacacttcac ggatactgta cttgtacttt ttttccaaat tgaaggt 57

<210> 109
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 109
gaggtggctg aaagaagcac gtatccctga tggcatggtt gcgggttata tg 52

<210> 110
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 110
gagaaaagca gttcctggaa caccccacc ccttaacccc ttatctctgc tt 52

<210> 111
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 111
ttacatatgt atacatgtgc catgtttggt gtgccgcacc cattaactcg tc 52

<210> 112
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 112
cttttccag attttttttt tttttttgag acagagtttc actcttgttg cc 52

<210> 113
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 113
tttggaaatc ttaagctttt atttgggtgt cacagagaaa caggatctgt at 52

<210> 114
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 114
tactttaagt tttagggtac atgtgcacca tgtgcaggtt tgttacatat gt 52

<210> 115
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 115
attcatgacc agccacataa atgcacagta ttacttcgca agcatgccaa tg 52

<210> 116
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 116
gtgcactaga attagctaaa gtgggggaaa aaaagatgca tttgatggtc ta 52

<210> 117
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 117
aacctccacc tcccaggttc aagagaattc tcctgcctca gcctcccaag ta 52

<210> 118
<211> 52
<212> DNA
<213> Homo sapiens

<400> 118
cctccctgtg atccgaagta gcagacrtac ttaacttcca tggtggattg tt 52

<210> 119

<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 119
aaaacattat tatatctaga aaaaaaatgt atcttaacca ttgtgggaag tg 52

<210> 120
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 120
ttgaaggtaa aatcataggg agttgggagc tgtcctcttg cgctgaatca gt 52

<210> 121
<211> 53
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(28)
<223> This region may or may not be present in its entirety

<400> 121
acttgtggcc tttttggaga ggttagcaac tctgaaaact ctgtccctgg acc 53

<210> 122
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 122
gaaggagaaa aagagaggaa ataagttaaa ataataaaca caattaataa ag 52

<210> 123
<211> 52
<212> DNA
<213> Homo sapiens

<400> 123
tgtatacatg tgccatgttg gtgtgcygca cccattaact cgtcatttag ca 52

<210> 124
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 124
tcacctgtca cctcctatag aactttcccc tgaccctcct ctatagcatt aa 52

<210> 125
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 125
ataaatgatc atcattcaca gtaggggttt tgttttgttt tttttctgga aa 52

<210> 126
<211> 52

<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 126
acagacatat ctacaaacac acttttttcac atatttgtgt aagtcattta tg 52

<210> 127
<211> 52
<212> DNA
<213> Homo sapiens

<400> 127
aaagacaact aaactaacgc tcaatgmaaa aagtattcgc tgctgtttac at 52

<210> 128
<211> 52
<212> DNA
<213> Homo sapiens

<400> 128
taccgtgtac cccgtgtttg tgaatgmggc cgcatattac gaaaagaaag aa 52

<210> 129
<211> 52
<212> DNA
<213> Homo sapiens

<400> 129
cagggctgga ggtaaaacca tttattrcta accccagagc agtgaagaag tg 52

<210> 130
<211> 52
<212> DNA
<213> Homo sapiens

<400> 130
ttcttcattg cctattgaag agagagygga atgctttggt tgccagatat gg 52

<210> 131

<211> 52
<212> DNA
<213> Homo sapiens

<400> 131
cacaagatct cattactcag gagctgycca agtgtctaat gtacttagtt aa 52

<210> 132
<211> 52
<212> DNA
<213> Homo sapiens

<400> 132
ttctgtgtaa catttcattt aagcaargga ttcggcaaat caaaaattgt ca 52

<210> 133
<211> 52
<212> DNA
<213> Homo sapiens

<400> 133
taaaacgtat tgaaggtatt attgackctg ttgaagcaaa gagaacaaaa ca 52

<210> 134
<211> 52
<212> DNA
<213> Homo sapiens

<400> 134
atctgctctt ccagttggat gacaagyctt gctgtctaac acctgctgca ga 52

<210> 135
<211> 52
<212> DNA
<213> Homo sapiens

<400> 135
ccatttttg ttgtattttt tttttcytga atacttttta tcgcagttgg tt 52

<210> 136
<211> 52
<212> DNA
<213> Homo sapiens

<400> 136

ccctgtctct aaaagaaaaa aaaaaaraaa aaaaaaagaa aacaaaaccc aa 52

<210> 137
<211> 52
<212> DNA
<213> Homo sapiens

<400> 137
agtggctcca aaaaggtcat ggaaccyctt gaggatgatg caaattgact ct 52

<210> 138
<211> 52
<212> DNA
<213> Homo sapiens

<400> 138
taaagttact tttcttttat tgactttccc ttatttttta accttatgct tt 52

<210> 139
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 139
cagagttaaa aaaaaaaaaa aaaaaaagga agtagcagca acagcttgga aa 52

<210> 140
<211> 52
<212> DNA
<213> Homo sapiens

<400> 140
gttgcccagg gttgagtgca gaggcayatc ttggctcaca gcaacctctg cc 52

<210> 141
<211> 55
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(30)
<223> This region may or may not be present in its entirety

<400> 141
aaggcttttt ttttttttt ttttttttt gagacagagt cttgctgtgt caccc 55

<210> 142
<211> 52
<212> DNA
<213> Homo sapiens

<400> 142
gcctacattc tgcaaagagg agggaasatt cacagctcca tacttgaacc ct 52

<210> 143
<211> 52
<212> DNA
<213> Homo sapiens

<400> 143
ccaaaggaga atagctctag gggaggsagg tggatgagta tgcatggggg ag 52

<210> 144
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 144
gactccatct caaaaaaaaa aaaaaaatgc agtctaatgg cagaattaga ct 52

<210> 145
<211> 52
<212> DNA
<213> Homo sapiens

<400> 145
ttatttattt atttatttat ttttgaraca gggtctctgt tgtccaggct gg 52

<210> 146
<211> 52
<212> DNA
<213> Homo sapiens

<400> 146
ttttgaggca gggtctcact ctgttgycca gggttgagtg cagaggcaca tc 52

<210> 147
<211> 52
<212> DNA
<213> Homo sapiens

<400> 147
cgccttgcga aggccccaca gcttgcytgt gacaaacgtt cataggcaaa tg 52

<210> 148
<211> 52
<212> DNA
<213> Homo sapiens

<400> 148
ggaggtctgt acaaagcacc acctacytca tgggtcagtt tccacagcag aa 52

<210> 149
<211> 55
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(30)
<223> This region may or may not be present in its entirety

<400> 149
gaatcttata attcactgtg tacctccctc tgtttcatat tttcgcaatt gaact 55

<210> 150
<211> 52
<212> DNA
<213> Homo sapiens

<400> 150

aacatagtat ctaatatagc tttacaycca aagccaaaat atgaatacac tg 52

<210> 151
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 151
ttgttttgtt ttgtttgggg gggggggttg tttttctgag agggagtctt gc 52

<210> 152
<211> 52
<212> DNA
<213> Homo sapiens

<400> 152
cataccaaag ggcagctgga gggatayaga cggaagtcat gtggagagtg aa 52

<210> 153
<211> 52
<212> DNA
<213> Homo sapiens

<400> 153
caggtgtgag ccaccacgac caccaawtta gctcttttta ctccttccct tc 52

<210> 154
<211> 52
<212> DNA
<213> Homo sapiens

<400> 154
agtggtagct gattttgctt ctggatmctt tgccaccttc ccactcttta at 52

<210> 155
<211> 64
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(39)
<223> This region may or may not be present in its entirety

<400> 155
aaagacctgt ttcttaaaaa aaaaaaagaa aaaaaaaaag aaagaaaaga aaagaaaaaa 60

acag 64

<210> 156
<211> 52
<212> DNA
<213> Homo sapiens

<400> 156
taaaaaatct ttcaatgagg agatgtyccc agagcaagac agctgtagga tg 52

<210> 157
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 157
aaaagaaaaa aaaaaaaaaa aaaaaaagaa aacaaaaccc aaacccataa ag 52

<210> 158
<211> 52
<212> DNA
<213> Homo sapiens

<400> 158
cctgtctcta aaagaaaaaa aaaaaaraaa aaaaaagaaa acaaaaccca aa 52

<210> 159
<211> 53
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(28)
<223> This region may or may not be present in its entirety

<400> 159
gttatcatag aaaaatatca cactctgtct gtatccccac ttccagaaac tgt 53

<210> 160
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 160
caggagctca tagaattaca tacaatcttt ttttttttt tttgagacag cg 52

<210> 161
<211> 55
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(30)
<223> This region may or may not be present in its entirety

<400> 161
ttgagaatct tataattcac tgtgtacctc cctctgtttc atattttcgc aattg 55

<210> 162
<211> 52
<212> DNA
<213> Homo sapiens

<400> 162
ccactaccac agtgcctaga gaacaayatg tgtttaataa tatttaaata at 52

<210> 163

<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 163
ccctgtctct aaaagaaaaa aaaaaaaaaa aaaaaaagaa aacaaaaccc aa 52

<210> 164
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 164
ccattatatc attcatttcc cactcatttt cttcattcca accaagatat at 52

<210> 165
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 165
tccgtctcaa aaaaaaaaaa aaaaagagaa aggaattatt ctcatgtata ca 52

<210> 166
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 166
ctggggcagt taaaaagaaa aacaaacccc tggtccctgc ccttgaggag at 52

<210> 167
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 167
ctccagggtc ccattccagg accacacgcc tgctacctct gcagctcact ca 52

<210> 168
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 168
ggattgacat ataccagtta gacggatttt ttttttccat aaaccaggct ca 52

<210> 169
<211> 52
<212> DNA
<213> Homo sapiens

<400> 169
acaaataatt actacatatc tacaacrttc cagatacaga agaaatggcc aa 52

<210> 170
<211> 54

<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(29)
<223> This region may or may not be present in its entirety

<400> 170
taaacacact tgaaacatca tataaaatga tattactaca agacttaacc gtaa 54

<210> 171
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 171
acacaggtaa tccatgttta ttatagaaaa atgccacatt actctttatt ga 52

<210> 172
<211> 53
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(28)
<223> This region may or may not be present in its entirety

<400> 172
agttatcata gaaaaatatc acactctgtc tgtatcccca cttccagaaa ctg 53

<210> 173
<211> 52
<212> DNA
<213> Homo sapiens

<220>

<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 173
aatgaactta caggaaggta atatatggga aataaacatc ttattgaatt ta 52

<210> 174
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 174
ggacccctga aaggcacaag acaccctttc aggttcacac ttcctgaaag ct 52

<210> 175
<211> 52
<212> DNA
<213> Homo sapiens

<400> 175
ccaccaatca ccagagcctt ctgctcrggg gtacctgagg gaaaacaagc aa 52

<210> 176
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 176
aaagactgaa aaaacattca taactatttt tcttgttatc ctcggaaatg tc 52

<210> 177
<211> 52
<212> DNA

<213> Homo sapiens

<400> 177
tatcttcatc ttggaggaga tgaggtygat ttcacctgct ggaagtccaa cc 52

<210> 178
<211> 52
<212> DNA
<213> Homo sapiens

<400> 178
ttgcctatga acgtttgtca cacttcygct gtgagttgct gaggcgaggt gt 52

<210> 179
<211> 52
<212> DNA
<213> Homo sapiens

<400> 179
gcttgctgtt ggatacatct cgccatyacc tgccactctc tagcatcctg ga 52

<210> 180
<211> 52
<212> DNA
<213> Homo sapiens

<400> 180
ccggggcttg ctgttggata catctckcca ttacctgcca ctctctagca tc 52

<210> 181
<211> 20
<212> DNA
<213> Homo sapiens

<400> 181
ctccgtttcg gtttcacttc 20

<210> 182
<211> 20
<212> DNA
<213> Homo sapiens

<400> 182
atcttctctt cagccctgct 20

<210> 183
<211> 20
<212> DNA
<213> Homo sapiens

<400> 183
gcacgatccg acggtagtgt 20

<210> 184
<211> 20
<212> DNA
<213> Homo sapiens

<400> 184
ccgtaatcgg gaagctgaag 20

SEQUENCE LISTING

<110> Good Start Genetics, Inc.
Porreca, Gregory
Laserson, Uri

<120> METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS

<130> GSGE-002/04US 29258/78

<140> US 16/952,764
<141> 2020-11-19

<150> US 15/231,687
<151> 2016-08-08

<160> 184

<170> PatentIn version 3.5

<210> 1
<211> 52
<212> DNA
<213> Homo sapiens

<400> 1
tcacatcacc aagttaaaaa aaaaaarggg gcgggggggc agaatgaaaa tt 52

<210> 2
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 2
aaacaaggat gaattaagtt ttttttaaa aaagaaacat ttggtaaggg ga 52

<210> 3
<211> 52
<212> DNA
<213> Homo sapiens

<400> 3
atcaccaagt taaaaaaaaa aaagggrcgg gggggcagaa tgaaaattgc at 52

<210> 4
<211> 52
<212> DNA
<213> Homo sapiens

<400> 4
ctattgaacc agaaccaaac aggaatrcca tagcattttg taaactaaac tg 52

<210> 5
<211> 52

<212> DNA
<213> Homo sapiens

<400> 5
caggagttca agaccagcct actaaamcac acacacacac acacacacac ac 52

<210> 6
<211> 52
<212> DNA
<213> Homo sapiens

<400> 6
gattaaataa tagtgtttat gtacccsgct tataggagaa gagggtgtgt gt 52

<210> 7
<211> 52
<212> DNA
<213> Homo sapiens

<400> 7
attgttatct tttcatataa ggtaacwgag gcccagagag attaaataac at 52

<210> 8
<211> 52
<212> DNA
<213> Homo sapiens

<400> 8
taattttaat taagtaaatt taattgrtag ataaataagt agataaaaaa ta 52

<210> 9
<211> 52
<212> DNA
<213> Homo sapiens

<400> 9
gtataaaaaa aaaaaaaaaa aaagttwgaa tgttttcttg cattcagagc ct 52

<210> 10
<211> 52
<212> DNA
<213> Homo sapiens

<400> 10
atactaaaaa tttaaagttc tcttgcrata tattttctta atatcttaca tc 52

<210> 11
<211> 52
<212> DNA
<213> Homo sapiens

<400> 11
tgctgggatt acaggcgtga gccaccrcgc ctggcctgat gggacatatt tt 52

<210> 12
<211> 52
<212> DNA

<213> Homo sapiens

<400> 12
ctacaatata agtatagtat tgcaaamcca tcaggaaggg tgttaactat tt 52

<210> 13
<211> 54
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(29)
<223> This region may or may not be present in its entirety

<400> 13
gttgtctcca aacttttttt caggtgagaa ggtggccaac cgagcttcgg aaag 54

<210> 14
<211> 52
<212> DNA
<213> Homo sapiens

<400> 14
ttttttcata aaagattata taaaggmtat tgcttttgaa tcacaaacac ta 52

<210> 15
<211> 52
<212> DNA
<213> Homo sapiens

<400> 15
atctagtgag cagtcaggaa agagaayttc cagatcctgg aaatcagggt ta 52

<210> 16
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 16
tagagtataa aaaaaaaaaa aaaaaaagtt tgaatgtttt cttgcattca ga 52

<210> 17
<211> 52
<212> DNA
<213> Homo sapiens

<400> 17
tgcaaatgcc aactatcaaa gatattsgag tatactgtca ataaacttca ta 52

<210> 18

<211> 52
<212> DNA
<213> Homo sapiens

<400> 18
tcctcaacag ttagaaacaa tattttsagt gatttcccat gccaacttta ct 52

<210> 19
<211> 52
<212> DNA
<213> Homo sapiens

<400> 19
tttttggtat tgttgttaaa taagtgrgaa ttcaatacag tataatgtct gt 52

<210> 20
<211> 87
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(72)
<223> This region may or may not be present in its entirety

<400> 20
cttgaaatcg gatatatata tatatatgta tatatatata tatatatata tatatatata 60

tagtattatc cctgttttca cagtttt 87

<210> 21
<211> 52
<212> DNA
<213> Homo sapiens

<400> 21
agaggggctg tgaaggacac caaggargag actaagagcc aggagggaaa ac 52

<210> 22
<211> 52
<212> DNA
<213> Homo sapiens

<400> 22
tagagtttat tagcttttac tactctrctt agttactttg tgttacagaa ta 52

<210> 23
<211> 53
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(28)
<223> This region may or may not be present in its entirety

<400> 23

actagtgatg atgagcttct tttcatatgt ttgttggctg cataaatgtc ttc 53

<210> 24
<211> 52
<212> DNA
<213> Homo sapiens

<400> 24
gcagagaaaa gaggggctgt gaaggasacc aaggaggaga ctaagagcca gg 52

<210> 25
<211> 52
<212> DNA
<213> Homo sapiens

<400> 25
ttcagaggtc taccactggt gcatackcta atcacagtgt cgaaaatttt ac 52

<210> 26
<211> 52
<212> DNA
<213> Homo sapiens

<400> 26
aagaaagagc aaaagagggc aaacttytca tacatttttg atgtcgaaac ca 52

<210> 27
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 27
cctaaagttt aaaaagaaaa aaaaaaagga agaaggaatt aaaaatccaa ag 52

<210> 28
<211> 52
<212> DNA
<213> Homo sapiens

<400> 28
gtgtgtgtgt gtatatatat atatatwtat atatttttt tttcctgagc ca 52

<210> 29
<211> 52
<212> DNA
<213> Homo sapiens

<400> 29
aaactgttga tgttttcatt tatttasatc attggaaaac tttagattct ag 52

<210> 30

<211> 52
<212> DNA
<213> Homo sapiens

<400> 30
tttatccatt cttaaccaga acagacrttt tttcagagct ggtccaggaa aa 52

<210> 31
<211> 52
<212> DNA
<213> Homo sapiens

<400> 31
ttgaaatcgg atatatatat atatatrtat atatatatat atatatatat at 52

<210> 32
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 32
tagtttttta tttcctcata ttattttcag tggctttttc ttccacatct tt 52

<210> 33
<211> 52
<212> DNA
<213> Homo sapiens

<400> 33
acatagttct cagtggtaca actacargtg atttctcttt tcttatttct gg 52

<210> 34
<211> 52
<212> DNA
<213> Homo sapiens

<400> 34
agagcaatgg catcccttgt cttgtgytat acaggatgca gcaatttata gg 52

<210> 35
<211> 52
<212> DNA
<213> Homo sapiens

<400> 35
ttctgtctac ataagatgtc atactaratt atcttttcca gcatgcattc ag 52

<210> 36
<211> 52
<212> DNA
<213> Homo sapiens

<400> 36
cagggtggca tgttaggcag tgcttaraat aaatgagttg gttatacaag ta 52

<210> 37
<211> 53
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(28)
<223> This region may or may not be present in its entirety

<400> 37
aggacacaca cacacacaca cacacacatg cacacacatt taaatagatg cat 53

<210> 38
<211> 52
<212> DNA
<213> Homo sapiens

<400> 38
taaaaaattg gtataatgaa attgcayttg tagtctttgg acatttaaat cc 52

<210> 39
<211> 52
<212> DNA
<213> Homo sapiens

<400> 39
tttcaatact taagaggtac gcagagraaa gaggggctgt gaaggacacc aa 52

<210> 40
<211> 52
<212> DNA
<213> Homo sapiens

<400> 40
cagcaatgaa tattttgagg ctgaggygct gaggggtaaa attgcagcct gg 52

<210> 41
<211> 55
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(30)
<223> This region may or may not be present in its entirety

<400> 41
ttatggttta tatttttgtg tcttctcttt aacacatctt ttctagcaga attca 55

<210> 42
<211> 52
<212> DNA

<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 42
gtattttagt ttttttttt tgtttgtttt gttttgtttt gttttgtttt tg 52

<210> 43
<211> 52
<212> DNA
<213> Homo sapiens

<400> 43
tgggtgactc catttttact tttagtytgg tctgttgagg cctcgtgaga ga 52

<210> 44
<211> 55
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(30)
<223> This region may or may not be present in its entirety

<400> 44
tattttcatg tattttagtt ttttttttt gtttgttttg ttttgttttg ttttg 55

<210> 45
<211> 52
<212> DNA
<213> Homo sapiens

<400> 45
gtgtgtgtgt atatatatat atatatwtat attttttttt tcctgagcca aa 52

<210> 46
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 46
aacctattag catgtctggc agaaaaatag atacttaata aatttcttaa at 52

<210> 47
<211> 52
<212> DNA
<213> Homo sapiens

<400> 47
gaggcttaga cagtttaagt aactcargca tggttacaca actagctagg gc 52

<210> 48
<211> 52
<212> DNA
<213> Homo sapiens

<400> 48
gtgtgagtat tatgagacca tatgttrgga gattttattt ggtattgagg at 52

<210> 49
<211> 52
<212> DNA
<213> Homo sapiens

<400> 49
gaaaccccac cccttctata gttttcyctt taatatttac aatggaacca tt 52

<210> 50
<211> 52
<212> DNA
<213> Homo sapiens

<400> 50
catatatata tagtgtgtgt gtgtgtrtat atatatatat atatatattt tt 52

<210> 51
<211> 53
<212> DNA
<213> Homo sapiens

<400> 51
cgagcgagag agagagagag agagagagga gccggcgcga gaactacgca tgc 53

<210> 52
<211> 52
<212> DNA
<213> Homo sapiens

<400> 52
ggcaggtaat atctagtacc ttacttwtat ttcctgagca cattctacat tt 52

<210> 53
<211> 52
<212> DNA
<213> Homo sapiens

<400> 53
ggccaggaat gggagtgctt aggtgcrgag gtggcactgt tcccgcagct gc 52

<210> 54
<211> 52
<212> DNA
<213> Homo sapiens

<400> 54
tgaaaactcc atcccctcag ggtcatyaga tgaagagaag accacagggg tt 52

<210> 55
<211> 52
<212> DNA
<213> Homo sapiens

<400> 55
tgtaggtaag ggtcacatgt gggagasgca gctgtgggta ggtcagccct gt 52

<210> 56
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 56
ccaagaaggc gccattacac tccagcctgg gcgacagggc gagactccct ca 52

<210> 57
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 57
tgccacaccc agctaatttg tgtgtggtat gtgtgtgtat gtatgtgtgt gt 52

<210> 58
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 58
gttcctccag taatttttt ttttttggt tttgagacag agtcttgccc tg 52

<210> 59
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 59
atcatgccca gataattttt tttttttgta ttttagtaga cacagggttt ca 52

<210> 60
<211> 53
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(28)
<223> This region may or may not be present in its entirety

<400> 60
cgagcgagag agagagagag agagagagga gccggcgcga gaactacgca tgc 53

<210> 61
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 61
cctgtgaggg gcacattcct tagtagctaa ggagttgggg gtgtgaagat cc 52

<210> 62
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 62
acaggctact ggctgggccc aggcaaaggg ggccttggca ggaaaagttc ct 52

<210> 63
<211> 53
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(28)
<223> This region may or may not be present in its entirety

<400> 63
gcgagagaga gagagagaga gagaggagag ccggcgcgag aactacgcat gcg 53

<210> 64
<211> 52
<212> DNA
<213> Homo sapiens

<400> 64
aagaagaaaa ataaaaagaa agtgggycag accgagagaa caggaagcct ga 52

<210> 65
<211> 52
<212> DNA
<213> Homo sapiens

<400> 65
aaggacaaag gcaaagagac aaaggckcaa cactgggggt ccccagagag tg 52

<210> 66
<211> 52
<212> DNA
<213> Homo sapiens

<400> 66
tacctagtca cttcctgcct ccatggygca aaaggggatg ggtgtgcctc tt 52

<210> 67
<211> 52
<212> DNA
<213> Homo sapiens

<400> 67
ctcttccgag gttccaccct gaacacyttc ctgctccctc gtggtgtaga gt 52

<210> 68
<211> 52
<212> DNA
<213> Homo sapiens

<400> 68
ttctgactgg caaccagccc cactctytgg gagccctcag gaatgaactt gc 52

<210> 69
<211> 52
<212> DNA
<213> Homo sapiens

<400> 69
gctcagcctc ccaggctgga gtgcagwggc gcgatctcgg ctcaccgcaa cc 52

<210> 70
<211> 52
<212> DNA
<213> Homo sapiens

<400> 70
catgaaccac atcaaatgag atttagyggg agtggcacac acagtcatga cc 52

<210> 71
<211> 52
<212> DNA
<213> Homo sapiens

<400> 71
aagcagccct ggggagtcgg ggcgggrcct ggattggaaa agagacggtc ac 52

<210> 72
<211> 52
<212> DNA
<213> Homo sapiens

<400> 72
ctccaagttc tgggagcaga gtgtgcrgct aggctcctgg gatcgaggga tg 52

<210> 73
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 73
gttcctccag taatttttt ttttttkgtt ttgagacaga gtcttgccct gt 52

<210> 74
<211> 52
<212> DNA
<213> Homo sapiens

<400> 74
ctagtacctt acttccctca agttcawtca tctcacagat atttcctgag ca 52

<210> 75
<211> 52
<212> DNA
<213> Homo sapiens

<400> 75
aattagccgt gcgtggtggc gggtgcytgt aatcccacgt acttgggagg ct 52

<210> 76
<211> 52
<212> DNA
<213> Homo sapiens

<400> 76
ccatggccag ccggggaggg gacgggmaca cacagaccca cacagagact ca 52

<210> 77
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (77)..(77)
<223> May or may not be present

<400> 77
agcgagagag agagagagag agagagrgag ccggcgcgag aactacgcat gc 52

<210> 78
<211> 52
<212> DNA
<213> Homo sapiens

<400> 78
cgtagcagtt agcagatgat aggcggbgaa atcttatttc acagggcatt aa 52

<210> 79
<211> 52
<212> DNA
<213> Homo sapiens

<400> 79
ctggccctgg tgacagtggg gctgtgygtg gggccagagc cttctcagag gt 52

<210> 80
<211> 52
<212> DNA
<213> Homo sapiens

<400> 80
cagatactgg ccctggtgac agtgggrctg tgcgtggggc cagagccttc tc 52

<210> 81
<211> 52
<212> DNA
<213> Homo sapiens

<400> 81
gacagatact ggccctggtg acagtgkggc tgtgcgtggg gccagagcct tc 52

<210> 82
<211> 53
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(28)
<223> This region may or may not be present in its entirety

<400> 82
ggctcctctc tctctctctc tctctctcgc tcgctctctc gctctctcgc tct 53

<210> 83
<211> 52
<212> DNA
<213> Homo sapiens

<400> 83
gtttcagacc agcctggcca acatagygaa accccatctc tactaaaaat aa 52

<210> 84
<211> 52
<212> DNA
<213> Homo sapiens

<400> 84
agtgggcgat tggatggagc tgagtayggg gcccatccag gctaatcaca cc 52

<210> 85
<211> 52
<212> DNA
<213> Homo sapiens

<400> 85
ccgggctccg tgaatgtttg tcacatstct gaagaacgta tgaattacat aa 52

<210> 86
<211> 52
<212> DNA
<213> Homo sapiens

<400> 86
gaatcccaac cccgacgctc gtcgccsggc tccgtgaatg tttgtcacat gt 52

<210> 87
<211> 52
<212> DNA
<213> Homo sapiens

<400> 87
gcgaatccca accccgacgc tcgtcgycgg gctccgtgaa tgtttgtcac at 52

<210> 88
<211> 52
<212> DNA
<213> Homo sapiens

<400> 88
tgggcagaag tcagggtcca aagaaakggc aaagaaaagt gtcagtggct ca 52

<210> 89
<211> 52
<212> DNA
<213> Homo sapiens

<400> 89
taagaaagac gtttttgatt tttttcrgac ttctctggct ccactaccct gc 52

<210> 90
<211> 52
<212> DNA
<213> Homo sapiens

<400> 90
ctgattcctg gccaggagcg gtggctyacg cctgtaatcc cagcgctttg gg 52

<210> 91
<211> 52
<212> DNA
<213> Homo sapiens

<400> 91
taaaaatgct gattcctggc caggagyggt ggctcacgcc tgtaatccca gc 52

<210> 92
<211> 52
<212> DNA
<213> Homo sapiens

<400> 92
tttaaaaatg ctgattcctg gccaggmgcg gtggctcacg cctgtaatcc ca 52

<210> 93
<211> 52
<212> DNA
<213> Homo sapiens

<400> 93
caagacctgt caaagatctg agaaatwtta cccgacttac aagctaacca tt 52

<210> 94
<211> 52
<212> DNA
<213> Homo sapiens

<400> 94
actgtaataa gtgctgtaaa agaaatrcac aaaataatat agcagagggt at 52

<210> 95
<211> 52
<212> DNA
<213> Homo sapiens

<400> 95
cttgaggtca ggagttcaag accagtytgg gcaacatggg gaaaaccttg tc 52

<210> 96
<211> 52
<212> DNA
<213> Homo sapiens

<400> 96
aggttgcagt gagccgagat catgccrttg cactccagcc ggggcaacaa aa 52

<210> 97
<211> 52
<212> DNA
<213> Homo sapiens

<400> 97
acaagtgtct tgaaattatc tgtgatytgc tatagagcaa tacttttgta aa 52

<210> 98
<211> 52
<212> DNA
<213> Homo sapiens

<400> 98
gtgggtatat gcagctctat gcactayctg ctcatttatt tggtaaatct aa 52

<210> 99
<211> 59
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(34)
<223> This region may or may not be present in its entirety

<400> 99
tgtgtgtgtg tgcgtgtgtg tgtgtgtgtg tgtgatcata agagtggctg cagcaaact 59

<210> 100
<211> 52
<212> DNA
<213> Homo sapiens

<400> 100
agtctggagt gcaatggtgc aatctcrgct cactgcagcc tccacctccg gg 52

<210> 101
<211> 52
<212> DNA
<213> Homo sapiens

<400> 101
ctcctaatgg atatttccta aattttkctg aacagaattt aacttgagct gg 52

<210> 102
<211> 52
<212> DNA
<213> Homo sapiens

<400> 102
atttaaaaat ggatttctag aaaaacratc acatacttga atattttagc aa 52

<210> 103
<211> 52
<212> DNA
<213> Homo sapiens

<400> 103
ctataaatgg gtagcatgag ggattcragg aggtggctga aagaagcacg ta 52

<210> 104
<211> 52
<212> DNA
<213> Homo sapiens

<400> 104
aagaaaccaa gcatagtaga gtgttaraaa accaaagcaa ctaaacaact gt 52

<210> 105
<211> 52
<212> DNA
<213> Homo sapiens

<400> 105
cggggctcag aacttgtaac agaaaawtaa aatatactcc actcaaggga at 52

<210> 106
<211> 57
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(32)
<223> This region may or may not be present in its entirety

<400> 106
tactacactt cacggatact gtacttgtac ttttttttcca aattgaaggt ttttggc 57

<210> 107
<211> 63
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(38)
<223> This region may or may not be present in its entirety

<400> 107
ttgtttttgt ttttgttttt gtttttgttt ttgtttttttg agatggagtc tcgctctgtc 60

gcc 63

<210> 108
<211> 57
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(32)

<223> This region may or may not be present in its entirety

<400> 108
tttgccttac tacacttcac ggatactgta cttgtacttt ttttccaaat tgaaggt 57

<210> 109
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 109
gaggtggctg aaagaagcac gtatccctga tggcatggtt gcgggttata tg 52

<210> 110
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 110
gagaaaagca gttcctggaa caccccacc ccttaacccc ttatctctgc tt 52

<210> 111
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 111
ttacatatgt atacatgtgc catgtttggt gtgccgcacc cattaactcg tc 52

<210> 112
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 112
ctttttccag attttttttt ttttttgag acagagtttc actcttgttg cc 52

<210> 113
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 113
tttggaaatc ttaagctttt atttgggtgt cacagagaaa caggatctgt at 52

<210> 114
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 114
tactttaagt tttagggtac atgtgcacca tgtgcaggtt tgttacatat gt 52

<210> 115
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 115
attcatgacc agccacataa atgcacagta ttacttcgca agcatgccaa tg 52

<210> 116
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 116
gtgcactaga attagctaaa gtgggggaaa aaaagatgca tttgatggtc ta 52

<210> 117
<211> 52

<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 117
aacctccacc tcccaggttc aagagaattc tcctgcctca gcctcccaag ta 52

<210> 118
<211> 52
<212> DNA
<213> Homo sapiens

<400> 118
cctccctgtg atccgaagta gcagacrtac ttaacttcca tggtggattg tt 52

<210> 119
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 119
aaaacattat tatatctaga aaaaaaatgt atcttaacca ttgtgggaag tg 52

<210> 120
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 120
ttgaaggtaa aatcataggg agttgggagc tgtcctcttg cgctgaatca gt 52

<210> 121
<211> 53
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(28)
<223> This region may or may not be present in its entirety

<400> 121

acttgtggcc tttttggaga ggttagcaac tctgaaaact ctgtccctgg acc 53

<210> 122
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 122
gaaggagaaa aagagaggaa ataagttaaa ataataaaca caattaataa ag 52

<210> 123
<211> 52
<212> DNA
<213> Homo sapiens

<400> 123
tgtatacatg tgccatgttg gtgtgcygca cccattaact cgtcatttag ca 52

<210> 124
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 124
tcacctgtca cctcctatag aactttcccc tgaccctcct ctatagcatt aa 52

<210> 125
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 125
ataaatgatc atcattcaca gtaggggttt tgttttgttt tttttctgga aa 52

<210> 126
<211> 52
<212> DNA
<213> Homo sapiens

<220>

<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 126
acagacatat ctacaaacac acttttttcac atatttgtgt aagtcattta tg 52

<210> 127
<211> 52
<212> DNA
<213> Homo sapiens

<400> 127
aaagacaact aaactaacgc tcaatgmaaa aagtattcgc tgctgtttac at 52

<210> 128
<211> 52
<212> DNA
<213> Homo sapiens

<400> 128
taccgtgtac cccgtgtttg tgaatgmggc cgcatattac gaaaagaaag aa 52

<210> 129
<211> 52
<212> DNA
<213> Homo sapiens

<400> 129
cagggctgga ggtaaaacca tttattrcta accccagagc agtgaagaag tg 52

<210> 130
<211> 52
<212> DNA
<213> Homo sapiens

<400> 130
ttcttcattg cctattgaag agagagygga atgctttggt tgccagatat gg 52

<210> 131
<211> 52
<212> DNA
<213> Homo sapiens

<400> 131
cacaagatct cattactcag gagctgycca agtgtctaat gtacttagtt aa 52

<210> 132
<211> 52
<212> DNA
<213> Homo sapiens

<400> 132
ttctgtgtaa catttcattt aagcaargga ttcggcaaat caaaaattgt ca 52

<210> 133
<211> 52

<212> DNA
<213> Homo sapiens

<400> 133
taaaacgtat tgaaggtatt attgackctg ttgaagcaaa gagaacaaaa ca 52

<210> 134
<211> 52
<212> DNA
<213> Homo sapiens

<400> 134
atctgctctt ccagttggat gacaagyctt gctgtctaac acctgctgca ga 52

<210> 135
<211> 52
<212> DNA
<213> Homo sapiens

<400> 135
ccatttttg ttgtattttt tttttcytga atacttttta tcgcagttgg tt 52

<210> 136
<211> 52
<212> DNA
<213> Homo sapiens

<400> 136
ccctgtctct aaaagaaaaa aaaaaaraaa aaaaaaagaa aacaaaaccc aa 52

<210> 137
<211> 52
<212> DNA
<213> Homo sapiens

<400> 137
agtggctcca aaaaggtcat ggaaccyctt gaggatgatg caaattgact ct 52

<210> 138
<211> 52
<212> DNA
<213> Homo sapiens

<400> 138
taaagttact tttcttttat tgactttccc ttatttttta accttatgct tt 52

<210> 139
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 139

cagagttaaa aaaaaaaaaa aaaaaaagga agtagcagca acagcttgga aa 52

<210> 140
<211> 52
<212> DNA
<213> Homo sapiens

<400> 140
gttgcccagg gttgagtgca gaggcayatc ttggctcaca gcaacctctg cc 52

<210> 141
<211> 55
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(30)
<223> This region may or may not be present in its entirety

<400> 141
aaggcttttt ttttttttt ttttttttt gagacagagt cttgctgtgt caccc 55

<210> 142
<211> 52
<212> DNA
<213> Homo sapiens

<400> 142
gcctacattc tgcaaagagg agggaasatt cacagctcca tacttgaacc ct 52

<210> 143
<211> 52
<212> DNA
<213> Homo sapiens

<400> 143
ccaaaggaga atagctctag gggaggsagg tggatgagta tgcatggggg ag 52

<210> 144
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 144
gactccatct caaaaaaaaa aaaaaaatgc agtctaatgg cagaattaga ct 52

<210> 145
<211> 52
<212> DNA
<213> Homo sapiens

<400> 145
ttatttattt atttatttat ttttgaraca gggtctctgt tgtccaggct gg 52

<210> 146
<211> 52
<212> DNA
<213> Homo sapiens

<400> 146
ttttgaggca gggtctcact ctgttgycca gggttgagtg cagaggcaca tc 52

<210> 147
<211> 52
<212> DNA
<213> Homo sapiens

<400> 147
cgccttgcga aggccccaca gcttgcytgt gacaaacgtt cataggcaaa tg 52

<210> 148
<211> 52
<212> DNA
<213> Homo sapiens

<400> 148
ggaggtctgt acaaagcacc acctacytca tgggtcagtt tccacagcag aa 52

<210> 149
<211> 55
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(30)
<223> This region may or may not be present in its entirety

<400> 149
gaatcttata attcactgtg tacctccctc tgtttcatat tttcgcaatt gaact 55

<210> 150
<211> 52
<212> DNA
<213> Homo sapiens

<400> 150
aacatagtat ctaatatagc tttacaycca aagccaaaat atgaatacac tg 52

<210> 151
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature

<222> (27)..(27)
<223> May or may not be present

<400> 151
ttgttttgtt ttgtttgggg gggggggttg tttttctgag agggagtctt gc 52

<210> 152
<211> 52
<212> DNA
<213> Homo sapiens

<400> 152
cataccaaag ggcagctgga gggatayaga cggaagtcat gtggagagtg aa 52

<210> 153
<211> 52
<212> DNA
<213> Homo sapiens

<400> 153
caggtgtgag ccaccacgac caccaawtta gctcttttta ctccttccct tc 52

<210> 154
<211> 52
<212> DNA
<213> Homo sapiens

<400> 154
agtggtagct gattttgctt ctggatmctt tgccaccttc ccactcttta at 52

<210> 155
<211> 64
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(39)
<223> This region may or may not be present in its entirety

<400> 155
aaagacctgt ttcttaaaaa aaaaaaagaa aaaaaaaaag aaagaaaaga aaagaaaaaa 60

acag 64

<210> 156
<211> 52
<212> DNA
<213> Homo sapiens

<400> 156
taaaaaatct ttcaatgagg agatgtyccc agagcaagac agctgtagga tg 52

<210> 157
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 157
aaaagaaaaa aaaaaaaaaa aaaaaaagaa aacaaaaccc aaacccataa ag 52

<210> 158
<211> 52
<212> DNA
<213> Homo sapiens

<400> 158
cctgtctcta aaagaaaaaa aaaaaaraaa aaaaaagaaa acaaaaccca aa 52

<210> 159
<211> 53
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(28)
<223> This region may or may not be present in its entirety

<400> 159
gttatcatag aaaaatatca cactctgtct gtatccccac ttccagaaac tgt 53

<210> 160
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 160
caggagctca tagaattaca tacaatcttt ttttttttt tttgagacag cg 52

<210> 161
<211> 55
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(30)
<223> This region may or may not be present in its entirety

<400> 161
ttgagaatct tataattcac tgtgtacctc cctctgtttc atattttcgc aattg 55

<210> 162
<211> 52
<212> DNA
<213> Homo sapiens

<400> 162
ccactaccac agtgcctaga gaacaayatg tgtttaataa tatttaaata at 52

<210> 163
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 163
ccctgtctct aaaagaaaaa aaaaaaaaaa aaaaaaagaa aacaaaaccc aa 52

<210> 164
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 164
ccattatatc attcatttcc cactcatttt cttcattcca accaagatat at 52

<210> 165
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 165
tccgtctcaa aaaaaaaaaa aaaaagagaa aggaattatt ctcatgtata ca 52

<210> 166
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)

<223> May or may not be present

<400> 166
ctggggcagt taaaaagaaa aacaaacccc tggtccctgc ccttgaggag at 52

<210> 167
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 167
ctccagggtc ccattccagg accacacgcc tgctacctct gcagctcact ca 52

<210> 168
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 168
ggattgacat ataccagtta gacggatttt ttttttccat aaaccaggct ca 52

<210> 169
<211> 52
<212> DNA
<213> Homo sapiens

<400> 169
acaaataatt actacatatc tacaacrttc cagatacaga agaaatggcc aa 52

<210> 170
<211> 54
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(29)
<223> This region may or may not be present in its entirety

<400> 170
taaacacact tgaaacatca tataaaatga tattactaca agacttaacc gtaa 54

<210> 171
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 171
acacaggtaa tccatgttta ttatagaaaa atgccacatt actctttatt ga 52

<210> 172
<211> 53
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(28)
<223> This region may or may not be present in its entirety

<400> 172
agttatcata gaaaaatatc acactctgtc tgtatcccca cttccagaaa ctg 53

<210> 173
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 173
aatgaactta caggaaggta atatatggga aataaacatc ttattgaatt ta 52

<210> 174
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 174
ggacccctga aaggcacaag acacccttтc aggttcacac ttcctgaaag ct 52

<210> 175
<211> 52
<212> DNA
<213> Homo sapiens

<400> 175
ccaccaatca ccagagcctt ctgctcrggg gtacctgagg gaaaacaagc aa 52

<210> 176
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 176
aaagactgaa aaaacattca taactatttt tcttgttatc ctcggaaatg tc 52

<210> 177
<211> 52
<212> DNA
<213> Homo sapiens

<400> 177
tatcttcatc ttggaggaga tgaggtygat ttcacctgct ggaagtccaa cc 52

<210> 178
<211> 52
<212> DNA
<213> Homo sapiens

<400> 178
ttgcctatga acgtttgtca cacttcygct gtgagttgct gaggcgaggt gt 52

<210> 179
<211> 52
<212> DNA
<213> Homo sapiens

<400> 179
gcttgctgtt ggatacatct cgccatyacc tgccactctc tagcatcctg ga 52

<210> 180
<211> 52
<212> DNA
<213> Homo sapiens

<400> 180
ccggggcttg ctgttggata catctckcca ttacctgcca ctctctagca tc 52

<210> 181
<211> 20
<212> DNA
<213> Homo sapiens

<400> 181
ctccgtttcg gtttcacttc 20

<210> 182
<211> 20
<212> DNA

<213> Homo sapiens

<400> 182
atcttctctt cagccctgct 20

<210> 183
<211> 20
<212> DNA
<213> Homo sapiens

<400> 183
gcacgatccg acggtagtgt 20

<210> 184
<211> 20
<212> DNA
<213> Homo sapiens

<400> 184
ccgtaatcgg gaagctgaag 20

<210> 1
<211> 52
<212> DNA
<213> Homo sapiens

<400> 1
tcacatcacc aagttaaaaa aaaaaarggg gcgggggggc agaatgaaaa tt 52

<210> 2
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 2
aaacaaggat gaattaagtt ttttttaaa aaagaaacat ttggtaaggg ga 52

<210> 3
<211> 52
<212> DNA
<213> Homo sapiens

<400> 3
atcaccaagt taaaaaaaaa aaagggrcgg gggggcagaa tgaaaattgc at 52

<210> 4
<211> 52
<212> DNA
<213> Homo sapiens

<400> 4
ctattgaacc agaaccaaac aggaatrcca tagcattttg taaactaaac tg 52

<210> 5
<211> 52
<212> DNA
<213> Homo sapiens

<400> 5
caggagttca agaccagcct actaaamcac acacacacac acacacacac ac 52

<210> 6
<211> 52
<212> DNA
<213> Homo sapiens

<400> 6
gattaaataa tagtgtttat gtacccsgct tataggagaa gagggtgtgt gt 52

<210> 7
<211> 52
<212> DNA
<213> Homo sapiens

<400> 7
attgttatct tttcatataa ggtaacwgag gcccagagag attaaataac at 52

<210> 8
<211> 52
<212> DNA
<213> Homo sapiens

<400> 8
taattttaat taagtaaatt taattgrtag ataaataagt agataaaaaa ta 52

<210> 9
<211> 52
<212> DNA
<213> Homo sapiens

<400> 9
gtataaaaaa aaaaaaaaaa aaagttwgaa tgttttcttg cattcagagc ct 52

<210> 10
<211> 52
<212> DNA
<213> Homo sapiens

<400> 10
atactaaaaa tttaaagttc tcttgcrata tattttctta atatcttaca tc 52

<210> 11
<211> 52
<212> DNA
<213> Homo sapiens

<400> 11
tgctgggatt acaggcgtga gccaccrcgc ctggcctgat gggacatatt tt 52

<210> 12
<211> 52
<212> DNA
<213> Homo sapiens

<400> 12
ctacaatata agtatagtat tgcaaamcca tcaggaaggg tgttaactat tt 52

<210> 13
<211> 54
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(29)
<223> This region may or may not be present in its entirety

<400> 13
gttgtctcca aacttttttt caggtgagaa ggtggccaac cgagcttcgg aaag 54

<210> 14
<211> 52
<212> DNA
<213> Homo sapiens

<400> 14
tttttttcata aaagattata taaaggmtat tgcttttgaa tcacaaacac ta 52

<210> 15
<211> 52
<212> DNA
<213> Homo sapiens

<400> 15
atctagtgag cagtcaggaa agagaayttc cagatcctgg aaatcagggt ta 52

<210> 16
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 16
tagagtataa aaaaaaaaaa aaaaaaagtt tgaatgtttt cttgcattca ga 52

<210> 17
<211> 52
<212> DNA
<213> Homo sapiens

<400> 17
tgcaaatgcc aactatcaaa gatattsgag tatactgtca ataaacttca ta 52

<210> 18
<211> 52
<212> DNA
<213> Homo sapiens

<400> 18
tcctcaacag ttagaaacaa tattttsagt gatttcccat gccaacttta ct 52

<210> 19
<211> 52
<212> DNA
<213> Homo sapiens

<400> 19
tttttggtat tgttgttaaa taagtgrgaa ttcaatacag tataatgtct gt 52

<210> 20
<211> 87
<212> DNA

<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(72)
<223> This region may or may not be present in its entirety

<400> 20
cttgaaatcg gatatatata tatatatgta tatatatata tatatatata tatatatata 60

tagtattatc cctgttttca cagtttt 87

<210> 21
<211> 52
<212> DNA
<213> Homo sapiens

<400> 21
agaggggctg tgaaggacac caaggargag actaagagcc aggagggaaa ac 52

<210> 22
<211> 52
<212> DNA
<213> Homo sapiens

<400> 22
tagagtttat tagcttttac tactctrctt agttactttg tgttacagaa ta 52

<210> 23
<211> 53
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(28)
<223> This region may or may not be present in its entirety

<400> 23
actagtgatg atgagcttct tttcatatgt ttgttggctg cataaatgtc ttc 53

<210> 24
<211> 52
<212> DNA
<213> Homo sapiens

<400> 24
gcagagaaaa gaggggctgt gaaggasacc aaggaggaga ctaagagcca gg 52

<210> 25
<211> 52
<212> DNA
<213> Homo sapiens

<400> 25
ttcagaggtc taccactggt gcatackcta atcacagtgt cgaaaatttt ac 52

<210> 26
<211> 52
<212> DNA
<213> Homo sapiens

<400> 26
aagaaagagc aaaagagggc aaacttytca tacatttttg atgtcgaaac ca 52

<210> 27
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 27
cctaaagttt aaaaagaaaa aaaaaaagga agaaggaatt aaaaatccaa ag 52

<210> 28
<211> 52
<212> DNA
<213> Homo sapiens

<400> 28
gtgtgtgtgt gtatatatat atatatwtat atatttttt tttcctgagc ca 52

<210> 29
<211> 52
<212> DNA
<213> Homo sapiens

<400> 29
aaactgttga tgttttcatt tatttasatc attggaaaac tttagattct ag 52

<210> 30
<211> 52
<212> DNA
<213> Homo sapiens

<400> 30
tttatccatt cttaaccaga acagacrttt tttcagagct ggtccaggaa aa 52

<210> 31
<211> 52
<212> DNA
<213> Homo sapiens

<400> 31
ttgaaatcgg atatatatat atatatrtat atatatatat atatatatat at 52

<210> 32
<211> 52
<212> DNA

<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 32
tagtttttta tttcctcata ttattttcag tggctttttc ttccacatct tt 52

<210> 33
<211> 52
<212> DNA
<213> Homo sapiens

<400> 33
acatagttct cagtggtaca actacargtg atttctcttt tcttatttct gg 52

<210> 34
<211> 52
<212> DNA
<213> Homo sapiens

<400> 34
agagcaatgg catcccttgt cttgtgytat acaggatgca gcaatttata gg 52

<210> 35
<211> 52
<212> DNA
<213> Homo sapiens

<400> 35
ttctgtctac ataagatgtc atactaratt atcttttcca gcatgcattc ag 52

<210> 36
<211> 52
<212> DNA
<213> Homo sapiens

<400> 36
cagggtggca tgttaggcag tgcttaraat aaatgagttg gttatacaag ta 52

<210> 37
<211> 53
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(28)
<223> This region may or may not be present in its entirety

<400> 37
aggacacaca cacacacaca cacacacatg cacacacatt taaatagatg cat 53

<210> 38

<211> 52
<212> DNA
<213> Homo sapiens

<400> 38
taaaaaattg gtataatgaa attgcayttg tagtctttgg acatttaaat cc 52

<210> 39
<211> 52
<212> DNA
<213> Homo sapiens

<400> 39
tttcaatact taagaggtac gcagagraaa gaggggctgt gaaggacacc aa 52

<210> 40
<211> 52
<212> DNA
<213> Homo sapiens

<400> 40
cagcaatgaa tattttgagg ctgaggygct gaggggtaaa attgcagcct gg 52

<210> 41
<211> 55
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(30)
<223> This region may or may not be present in its entirety

<400> 41
ttatggttta tatttttgtg tcttctcttt aacacatctt ttctagcaga attca 55

<210> 42
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 42
gtattttagt ttttttttt tgtttgtttt gttttgtttt gttttgtttt tg 52

<210> 43
<211> 52
<212> DNA
<213> Homo sapiens

<400> 43
tgggtgactc catttttact tttagtytgg tctgttgagg cctcgtgaga ga 52

<210> 44
<211> 55
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(30)
<223> This region may or may not be present in its entirety

<400> 44
tattttcatg tattttagtt ttttttttt gtttgttttg ttttgttttg ttttg 55

<210> 45
<211> 52
<212> DNA
<213> Homo sapiens

<400> 45
gtgtgtgtgt atatatatat atatatwtat attttttttt tcctgagcca aa 52

<210> 46
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 46
aacctattag catgtctggc agaaaaatag atacttaata aatttcttaa at 52

<210> 47
<211> 52
<212> DNA
<213> Homo sapiens

<400> 47
gaggcttaga cagtttaagt aactcargca tggttacaca actagctagg gc 52

<210> 48
<211> 52
<212> DNA
<213> Homo sapiens

<400> 48
gtgtgagtat tatgagacca tatgttrgga gattttattt ggtattgagg at 52

<210> 49
<211> 52
<212> DNA
<213> Homo sapiens

<400> 49

gaaaccccac cccttctata gttttcyctt taatatttac aatggaacca tt 52

<210> 50
<211> 52
<212> DNA
<213> Homo sapiens

<400> 50
catatatata tagtgtgtgt gtgtgtrtat atatatatat atatatattt tt 52

<210> 51
<211> 53
<212> DNA
<213> Homo sapiens

<400> 51
cgagcgagag agagagagag agagagagga gccggcgcga gaactacgca tgc 53

<210> 52
<211> 52
<212> DNA
<213> Homo sapiens

<400> 52
ggcaggtaat atctagtacc ttacttwtat ttcctgagca cattctacat tt 52

<210> 53
<211> 52
<212> DNA
<213> Homo sapiens

<400> 53
ggccaggaat gggagtgctt aggtgcrgag gtggcactgt tcccgcagct gc 52

<210> 54
<211> 52
<212> DNA
<213> Homo sapiens

<400> 54
tgaaaactcc atcccctcag ggtcatyaga tgaagagaag accacagggg tt 52

<210> 55
<211> 52
<212> DNA
<213> Homo sapiens

<400> 55
tgtaggtaag ggtcacatgt gggagasgca gctgtgggta ggtcagccct gt 52

<210> 56
<211> 52
<212> DNA
<213> Homo sapiens

<220>

<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 56
ccaagaaggc gccattacac tccagcctgg gcgacagggc gagactccct ca 52

<210> 57
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 57
tgccacaccc agctaatttg tgtgtggtat gtgtgtgtat gtatgtgtgt gt 52

<210> 58
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 58
gttcctccag taatttttt ttttttggt tttgagacag agtcttgccc tg 52

<210> 59
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 59
atcatgccca gataattttt ttttttgta ttttagtaga cacagggttt ca 52

<210> 60
<211> 53
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(28)
<223> This region may or may not be present in its entirety

<400> 60
cgagcgagag agagagagag agagagagga gccggcgcga gaactacgca tgc 53

<210> 61
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 61
cctgtgaggg gcacattcct tagtagctaa ggagttgggg gtgtgaagat cc 52

<210> 62
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 62
acaggctact ggctgggccc aggcaaaggg ggccttggca ggaaaagttc ct 52

<210> 63
<211> 53
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(28)
<223> This region may or may not be present in its entirety

<400> 63
gcgagagaga gagagagaga gagaggagag ccggcgcgag aactacgcat gcg 53

<210> 64
<211> 52
<212> DNA
<213> Homo sapiens

<400> 64
aagaagaaaa ataaaaagaa agtgggycag accgagagaa caggaagcct ga 52

<210> 65
<211> 52
<212> DNA
<213> Homo sapiens

<400> 65

aaggacaaag gcaaagagac aaaggckcaa cactgggggt ccccagagag tg 52

<210> 66
<211> 52
<212> DNA
<213> Homo sapiens

<400> 66
tacctagtca cttcctgcct ccatggygca aaaggggatg ggtgtgcctc tt 52

<210> 67
<211> 52
<212> DNA
<213> Homo sapiens

<400> 67
ctcttccgag gttccaccct gaacacyttc ctgctccctc gtggtgtaga gt 52

<210> 68
<211> 52
<212> DNA
<213> Homo sapiens

<400> 68
ttctgactgg caaccagccc cactctytgg gagccctcag gaatgaactt gc 52

<210> 69
<211> 52
<212> DNA
<213> Homo sapiens

<400> 69
gctcagcctc ccaggctgga gtgcagwggc gcgatctcgg ctcaccgcaa cc 52

<210> 70
<211> 52
<212> DNA
<213> Homo sapiens

<400> 70
catgaaccac atcaaatgag atttagyggg agtggcacac acagtcatga cc 52

<210> 71
<211> 52
<212> DNA
<213> Homo sapiens

<400> 71
aagcagccct ggggagtcgg ggcgggrcct ggattggaaa agagacggtc ac 52

<210> 72
<211> 52
<212> DNA
<213> Homo sapiens

<400> 72
ctccaagttc tgggagcaga gtgtgcrgct aggctcctgg gatcgaggga tg 52

<210> 73
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 73
gttcctccag taatttttt ttttttkgtt ttgagacaga gtcttgccct gt 52

<210> 74
<211> 52
<212> DNA
<213> Homo sapiens

<400> 74
ctagtacctt acttccctca agttcawtca tctcacagat atttcctgag ca 52

<210> 75
<211> 52
<212> DNA
<213> Homo sapiens

<400> 75
aattagccgt gcgtggtggc gggtgcytgt aatcccacgt acttgggagg ct 52

<210> 76
<211> 52
<212> DNA
<213> Homo sapiens

<400> 76
ccatggccag ccggggaggg gacgggmaca cacagaccca cacagagact ca 52

<210> 77
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (77)..(77)
<223> May or may not be present

<400> 77
agcgagagag agagagagag agagagrgag ccggcgcgag aactacgcat gc 52

<210> 78
<211> 52
<212> DNA
<213> Homo sapiens

<400> 78
cgtagcagtt agcagatgat aggcggbgaa atcttatttc acagggcatt aa 52

<210> 79
<211> 52
<212> DNA
<213> Homo sapiens

<400> 79
ctggccctgg tgacagtggg gctgtgygtg gggccagagc cttctcagag gt 52

<210> 80
<211> 52
<212> DNA
<213> Homo sapiens

<400> 80
cagatactgg ccctggtgac agtgggrctg tgcgtggggc cagagccttc tc 52

<210> 81
<211> 52
<212> DNA
<213> Homo sapiens

<400> 81
gacagatact ggccctggtg acagtgkggc tgtgcgtggg gccagagcct tc 52

<210> 82
<211> 53
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(28)
<223> This region may or may not be present in its entirety

<400> 82
ggctcctctc tctctctctc tctctctcgc tcgctctctc gctctctcgc tct 53

<210> 83
<211> 52
<212> DNA
<213> Homo sapiens

<400> 83
gtttcagacc agcctggcca acatagygaa accccatctc tactaaaaat aa 52

<210> 84
<211> 52
<212> DNA
<213> Homo sapiens

<400> 84
agtgggcgat tggatggagc tgagtayggg gcccatccag gctaatcaca cc 52

<210> 85
<211> 52
<212> DNA
<213> Homo sapiens

<400> 85
ccgggctccg tgaatgtttg tcacatstct gaagaacgta tgaattacat aa 52

<210> 86
<211> 52
<212> DNA
<213> Homo sapiens

<400> 86
gaatcccaac cccgacgctc gtcgccsggc tccgtgaatg tttgtcacat gt 52

<210> 87
<211> 52
<212> DNA
<213> Homo sapiens

<400> 87
gcgaatccca accccgacgc tcgtcgycgg gctccgtgaa tgtttgtcac at 52

<210> 88
<211> 52
<212> DNA
<213> Homo sapiens

<400> 88
tgggcagaag tcagggtcca aagaaakggc aaagaaaagt gtcagtggct ca 52

<210> 89
<211> 52
<212> DNA
<213> Homo sapiens

<400> 89
taagaaagac gtttttgatt tttttcrgac ttctctggct ccactaccct gc 52

<210> 90
<211> 52
<212> DNA
<213> Homo sapiens

<400> 90
ctgattcctg gccaggagcg gtggctyacg cctgtaatcc cagcgctttg gg 52

<210> 91
<211> 52
<212> DNA
<213> Homo sapiens

<400> 91
taaaaatgct gattcctggc caggagyggt ggctcacgcc tgtaatccca gc 52

<210> 92

<211> 52
<212> DNA
<213> Homo sapiens

<400> 92
tttaaaaatg ctgattcctg gccaggmgcg gtggctcacg cctgtaatcc ca 52

<210> 93
<211> 52
<212> DNA
<213> Homo sapiens

<400> 93
caagacctgt caaagatctg agaaatwtta cccgacttac aagctaacca tt 52

<210> 94
<211> 52
<212> DNA
<213> Homo sapiens

<400> 94
actgtaataa gtgctgtaaa agaaatrcac aaaataatat agcagagggt at 52

<210> 95
<211> 52
<212> DNA
<213> Homo sapiens

<400> 95
cttgaggtca ggagttcaag accagtytgg gcaacatggg gaaaaccttg tc 52

<210> 96
<211> 52
<212> DNA
<213> Homo sapiens

<400> 96
aggttgcagt gagccgagat catgccrttg cactccagcc ggggcaacaa aa 52

<210> 97
<211> 52
<212> DNA
<213> Homo sapiens

<400> 97
acaagtgtct tgaaattatc tgtgatytgc tatagagcaa tacttttgta aa 52

<210> 98
<211> 52
<212> DNA
<213> Homo sapiens

<400> 98
gtgggtatat gcagctctat gcactayctg ctcatttatt tggtaaatct aa 52

<210> 99
<211> 59

<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(34)
<223> This region may or may not be present in its entirety

<400> 99
tgtgtgtgtg tgcgtgtgtg tgtgtgtgtg tgtgatcata agagtggctg cagcaaact 59

<210> 100
<211> 52
<212> DNA
<213> Homo sapiens

<400> 100
agtctggagt gcaatggtgc aatctcrgct cactgcagcc tccacctccg gg 52

<210> 101
<211> 52
<212> DNA
<213> Homo sapiens

<400> 101
ctcctaatgg atatttccta aattttkctg aacagaattt aacttgagct gg 52

<210> 102
<211> 52
<212> DNA
<213> Homo sapiens

<400> 102
atttaaaaat ggatttctag aaaaacratc acatacttga atattttagc aa 52

<210> 103
<211> 52
<212> DNA
<213> Homo sapiens

<400> 103
ctataaatgg gtagcatgag ggattcragg aggtggctga aagaagcacg ta 52

<210> 104
<211> 52
<212> DNA
<213> Homo sapiens

<400> 104
aagaaaccaa gcatagtaga gtgttaraaa accaaagcaa ctaaacaact gt 52

<210> 105
<211> 52
<212> DNA
<213> Homo sapiens

<400> 105

cggggctcag aacttgtaac agaaaawtaa aatatactcc actcaaggga at 52

<210> 106
<211> 57
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(32)
<223> This region may or may not be present in its entirety

<400> 106
tactacactt cacggatact gtacttgtac ttttttttcca aattgaaggt ttttggc 57

<210> 107
<211> 63
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(38)
<223> This region may or may not be present in its entirety

<400> 107
ttgtttttgt ttttgttttt gtttttgttt ttgtttttttg agatggagtc tcgctctgtc 60

gcc 63

<210> 108
<211> 57
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(32)
<223> This region may or may not be present in its entirety

<400> 108
tttgccttac tacacttcac ggatactgta cttgtacttt ttttccaaat tgaaggt 57

<210> 109
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 109
gaggtggctg aaagaagcac gtatccctga tggcatggtt gcgggttata tg 52

<210> 110
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 110
gagaaaagca gttcctggaa caccccacc ccttaacccc ttatctctgc tt 52

<210> 111
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 111
ttacatatgt atacatgtgc catgtttggt gtgccgcacc cattaactcg tc 52

<210> 112
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 112
ctttttccag attttttttt ttttttgag acagagtttc actcttgttg cc 52

<210> 113
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 113
tttggaaatc ttaagctttt atttgggtgt cacagagaaa caggatctgt at 52

<210> 114
<211> 52
<212> DNA

<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 114
tactttaagt tttagggtac atgtgcacca tgtgcaggtt tgttacatat gt 52

<210> 115
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 115
attcatgacc agccacataa atgcacagta ttacttcgca agcatgccaa tg 52

<210> 116
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 116
gtgcactaga attagctaaa gtgggggaaa aaaagatgca tttgatggtc ta 52

<210> 117
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 117
aacctccacc tcccaggttc aagagaattc tcctgcctca gcctcccaag ta 52

<210> 118
<211> 52
<212> DNA
<213> Homo sapiens

<400> 118
cctccctgtg atccgaagta gcagacrtac ttaacttcca tggtggattg tt 52

<210> 119
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 119
aaaacattat tatatctaga aaaaaaatgt atcttaacca ttgtgggaag tg 52

<210> 120
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 120
ttgaaggtaa aatcataggg agttgggagc tgtcctcttg cgctgaatca gt 52

<210> 121
<211> 53
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(28)
<223> This region may or may not be present in its entirety

<400> 121
acttgtggcc tttttggaga ggttagcaac tctgaaaact ctgtccctgg acc 53

<210> 122
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 122
gaaggagaaa aagagaggaa ataagttaaa ataataaaca caattaataa ag 52

<210> 123
<211> 52

<212> DNA
<213> Homo sapiens

<400> 123
tgtatacatg tgccatgttg gtgtgcygca cccattaact cgtcatttag ca 52

<210> 124
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 124
tcacctgtca cctcctatag aactttcccc tgaccctcct ctatagcatt aa 52

<210> 125
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 125
ataaatgatc atcattcaca gtaggggttt tgttttgttt tttttctgga aa 52

<210> 126
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 126
acagacatat ctacaaacac acttttttcac atatttgtgt aagtcattta tg 52

<210> 127
<211> 52
<212> DNA
<213> Homo sapiens

<400> 127
aaagacaact aaactaacgc tcaatgmaaa aagtattcgc tgctgtttac at 52

<210> 128
<211> 52
<212> DNA

<213> Homo sapiens

<400> 128
taccgtgtac cccgtgtttg tgaatgmggc cgcatattac gaaaagaaag aa 52

<210> 129
<211> 52
<212> DNA
<213> Homo sapiens

<400> 129
cagggctgga ggtaaaacca tttattrcta accccagagc agtgaagaag tg 52

<210> 130
<211> 52
<212> DNA
<213> Homo sapiens

<400> 130
ttcttcattg cctattgaag agagagygga atgctttggt tgccagatat gg 52

<210> 131
<211> 52
<212> DNA
<213> Homo sapiens

<400> 131
cacaagatct cattactcag gagctgycca agtgtctaat gtacttagtt aa 52

<210> 132
<211> 52
<212> DNA
<213> Homo sapiens

<400> 132
ttctgtgtaa catttcattt aagcaargga ttcggcaaat caaaaattgt ca 52

<210> 133
<211> 52
<212> DNA
<213> Homo sapiens

<400> 133
taaaacgtat tgaaggtatt attgackctg ttgaagcaaa gagaacaaaa ca 52

<210> 134
<211> 52
<212> DNA
<213> Homo sapiens

<400> 134
atctgctctt ccagttggat gacaagyctt gctgtctaac acctgctgca ga 52

<210> 135
<211> 52
<212> DNA
<213> Homo sapiens

<400> 135
ccatttttg ttgtattttt tttttcytga atacttttta tcgcagttgg tt 52

<210> 136
<211> 52
<212> DNA
<213> Homo sapiens

<400> 136
ccctgtctct aaaagaaaaa aaaaaaraaa aaaaaaagaa aacaaaaccc aa 52

<210> 137
<211> 52
<212> DNA
<213> Homo sapiens

<400> 137
agtggctcca aaaaggtcat ggaaccyctt gaggatgatg caaattgact ct 52

<210> 138
<211> 52
<212> DNA
<213> Homo sapiens

<400> 138
taaagttact tttcttttat tgactttccc ttatttttta accttatgct tt 52

<210> 139
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 139
cagagttaaa aaaaaaaaaa aaaaaaagga agtagcagca acagcttgga aa 52

<210> 140
<211> 52
<212> DNA
<213> Homo sapiens

<400> 140
gttgcccagg gttgagtgca gaggcayatc ttggctcaca gcaacctctg cc 52

<210> 141
<211> 55
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature

<222> (27)..(30)
<223> This region may or may not be present in its entirety

<400> 141
aaggcttttt ttttttttt ttttttttt gagacagagt cttgctgtgt caccc 55

<210> 142
<211> 52
<212> DNA
<213> Homo sapiens

<400> 142
gcctacattc tgcaaagagg agggaasatt cacagctcca tacttgaacc ct 52

<210> 143
<211> 52
<212> DNA
<213> Homo sapiens

<400> 143
ccaaaggaga atagctctag gggaggsagg tggatgagta tgcatggggg ag 52

<210> 144
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 144
gactccatct caaaaaaaaa aaaaaaatgc agtctaatgg cagaattaga ct 52

<210> 145
<211> 52
<212> DNA
<213> Homo sapiens

<400> 145
ttatttattt atttatttat ttttgaraca gggtctctgt tgtccaggct gg 52

<210> 146
<211> 52
<212> DNA
<213> Homo sapiens

<400> 146
ttttgaggca gggtctcact ctgttgycca gggttgagtg cagaggcaca tc 52

<210> 147
<211> 52
<212> DNA
<213> Homo sapiens

<400> 147

cgccttgcga aggccccaca gcttgcytgt gacaaacgtt cataggcaaa tg 52

<210> 148
<211> 52
<212> DNA
<213> Homo sapiens

<400> 148
ggaggtctgt acaaagcacc acctacytca tgggtcagtt tccacagcag aa 52

<210> 149
<211> 55
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(30)
<223> This region may or may not be present in its entirety

<400> 149
gaatcttata attcactgtg tacctccctc tgtttcatat tttcgcaatt gaact 55

<210> 150
<211> 52
<212> DNA
<213> Homo sapiens

<400> 150
aacatagtat ctaatatagc tttacaycca aagccaaaat atgaatacac tg 52

<210> 151
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 151
ttgttttgtt ttgtttgggg ggggggttg tttttctgag agggagtctt gc 52

<210> 152
<211> 52
<212> DNA
<213> Homo sapiens

<400> 152
cataccaaag ggcagctgga gggatayaga cggaagtcat gtggagagtg aa 52

<210> 153
<211> 52
<212> DNA
<213> Homo sapiens

<400> 153
caggtgtgag ccaccacgac caccaawtta gctcttttta ctccttccct tc 52

<210> 154
<211> 52
<212> DNA
<213> Homo sapiens

<400> 154
agtggtagct gattttgctt ctggatmctt tgccaccttc ccactcttta at 52

<210> 155
<211> 64
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(39)
<223> This region may or may not be present in its entirety

<400> 155
aaagacctgt ttcttaaaaa aaaaaaagaa aaaaaaaaag aaagaaaaga aaagaaaaaa 60

acag 64

<210> 156
<211> 52
<212> DNA
<213> Homo sapiens

<400> 156
taaaaaatct ttcaatgagg agatgtyccc agagcaagac agctgtagga tg 52

<210> 157
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 157
aaaagaaaaa aaaaaaaaaa aaaaaaagaa aacaaaaccc aaacccataa ag 52

<210> 158
<211> 52
<212> DNA
<213> Homo sapiens

<400> 158
cctgtctcta aaagaaaaaa aaaaaaraaa aaaaaagaaa acaaaaccca aa 52

<210> 159
<211> 53
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(28)
<223> This region may or may not be present in its entirety

<400> 159
gttatcatag aaaaatatca cactctgtct gtatccccac ttccagaaac tgt 53

<210> 160
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 160
caggagctca tagaattaca tacaatcttt ttttttttt tttgagacag cg 52

<210> 161
<211> 55
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(30)
<223> This region may or may not be present in its entirety

<400> 161
ttgagaatct tataattcac tgtgtacctc cctctgtttc atattttcgc aattg 55

<210> 162
<211> 52
<212> DNA
<213> Homo sapiens

<400> 162
ccactaccac agtgcctaga gaacaayatg tgtttaataa tatttaaata at 52

<210> 163
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 163
ccctgtctct aaaagaaaaa aaaaaaaaaa aaaaaaagaa aacaaaaccc aa 52

<210> 164
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 164
ccattatatc attcatttcc cactcatttt cttcattcca accaagatat at 52

<210> 165
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 165
tccgtctcaa aaaaaaaaaa aaaaagagaa aggaattatt ctcatgtata ca 52

<210> 166
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 166
ctggggcagt taaaaagaaa aacaaacccc tggtccctgc ccttgaggag at 52

<210> 167
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 167
ctccagggtc ccattccagg accacacgcc tgctacctct gcagctcact ca 52

<210> 168
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 168
ggattgacat ataccagtta gacggatttt ttttttccat aaaccaggct ca 52

<210> 169
<211> 52
<212> DNA
<213> Homo sapiens

<400> 169
acaaataatt actacatatc tacaacrttc cagatacaga agaaatggcc aa 52

<210> 170
<211> 54
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(29)
<223> This region may or may not be present in its entirety

<400> 170
taaacacact tgaaacatca tataaaatga tattactaca agacttaacc gtaa 54

<210> 171
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 171
acacaggtaa tccatgttta ttatagaaaa atgccacatt actctttatt ga 52

<210> 172
<211> 53
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(28)

<223> This region may or may not be present in its entirety

<400> 172
agttatcata gaaaaatatc acactctgtc tgtatcccca cttccagaaa ctg 53

<210> 173
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 173
aatgaactta caggaaggta atatatggga aataaacatc ttattgaatt ta 52

<210> 174
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 174
ggacccctga aaggcacaag acaccctttc aggttcacac ttcctgaaag ct 52

<210> 175
<211> 52
<212> DNA
<213> Homo sapiens

<400> 175
ccaccaatca ccagagcctt ctgctcrggg gtacctgagg gaaaacaagc aa 52

<210> 176
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 176
aaagactgaa aaaacattca taactatttt tcttgttatc ctcggaaatg tc 52

<210> 177
<211> 52
<212> DNA
<213> Homo sapiens

<400> 177
tatcttcatc ttggaggaga tgaggtygat ttcacctgct ggaagtccaa cc 52

<210> 178
<211> 52
<212> DNA
<213> Homo sapiens

<400> 178
ttgcctatga acgtttgtca cacttcygct gtgagttgct gaggcgaggt gt 52

<210> 179
<211> 52
<212> DNA
<213> Homo sapiens

<400> 179
gcttgctgtt ggatacatct cgccatyacc tgccactctc tagcatcctg ga 52

<210> 180
<211> 52
<212> DNA
<213> Homo sapiens

<400> 180
ccggggcttg ctgttggata catctckcca ttacctgcca ctctctagca tc 52

<210> 181
<211> 20
<212> DNA
<213> Homo sapiens

<400> 181
ctccgtttcg gtttcacttc 20

<210> 182
<211> 20
<212> DNA
<213> Homo sapiens

<400> 182
atcttctctt cagccctgct 20

<210> 183
<211> 20
<212> DNA
<213> Homo sapiens

<400> 183
gcacgatccg acggtagtgt 20

<210> 184
<211> 20
<212> DNA
<213> Homo sapiens

<400> 184
ccgtaatcgg gaagctgaag 20

SEQUENCE LISTING

<110> Good Start Genetics, Inc.
Porreca, Gregory
Laserson, Uri

<120> METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS

<130> GSGE-002/04US 29258/78

<140> US 16/952,764
<141> 2020-11-19

<150> US 15/231,687
<151> 2016-08-08

<160> 184

<170> PatentIn version 3.5

PTO/AIA/01 (06-12)
Approved for use through 01/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 CFR 1.76)

Title of Invention	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS

As the below named inventor, I hereby declare that:

This declaration is directed to:

- [ ] The attached application, or
- [x] United States application or PCT international application number 15/231,687 filed on 08/08/2016.

The above-identified application was made or authorized to be made by me.

I believe that I am the original inventor or an original joint inventor of a claimed invention in the application.

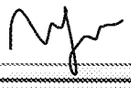
I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.

WARNING:

Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.

LEGAL NAME OF INVENTOR

Inventor: Uri Laserson Date (Optional):

Signature: 

Note: An application data sheet (PTO/SB/14 or equivalent), including naming the entire inventive entity, must accompany this form or must have been previously filed. Use an additional PTO/AIA/01 form for each additional inventor.

62684164 v1

PTO/AIA/01 (06-12)
Approved for use through 01/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 CFR 1.76)

Title of Invention	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS

As the below named inventor, I hereby declare that:

This declaration is directed to:

☐ The attached application, or

☒ United States application or PCT international application number 15/231,687 filed on 08/08/2016

The above-identified application was made or authorized to be made by me.

I believe that I am the original inventor or an original joint inventor of a claimed invention in the application.

I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.

WARNING:

Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.

LEGAL NAME OF INVENTOR

Inventor: Gregory Porreca Date (Optional): 1-27-17

Signature: Gregory Porreca

Note: An application data sheet (PTO/SB/14 or equivalent), including naming the entire inventive entity, must accompany this form or must have been previously filed. Use an additional PTO/AIA/01 form for each additional inventor.

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	MLB-002/06US 34468/40
	Application Number	
Title of Invention	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS	

The application data sheet is part of the provisional or nonprovisional application for which it is being submitted. The following form contains the bibliographic data arranged in a format specified by the United States Patent and Trademark Office as outlined in 37 CFR 1.76.
This document may be completed electronically and submitted to the Office in electronic format using the Electronic Filing System (EFS) or the document may be printed and included in a paper filed application.

Secrecy Order 37 CFR 5.2:

☐ Portions or all of the application associated with this Application Data Sheet may fall under a Secrecy Order pursuant to 37 CFR 5.2 (Paper filers only. Applications that fall under Secrecy Order may not be filed electronically.)

Inventor Information:

Inventor 1 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Gregory		Porreca	

Residence Information (Select One) ● US Residency Non US Residency Active US Military Service

City	State/Province	Country of Residence
Cambridge	MA	US

Mailing Address of Inventor:

Address 1	57 Reservoir Street
Address 2	
City	Cambridge
State/Province	MA
Postal Code	02138
Country	US

Inventor 2 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Uri		Laserson	

Residence Information (Select One) ◉ US Residency Non US Residency Active US Military Service

City	State/Province	Country of Residence
Boston	MA	US

Mailing Address of Inventor:

Address 1	645 Beacon Street
Address 2	#215
City	Boston
State/Province	MA
Postal Code	02215
Country	US

All Inventors Must Be Listed - Additional Inventor Information blocks may be generated within this form by selecting the **Add** button. Add

Correspondence Information:

Enter either Customer Number or complete the Correspondence Information section below.
For further information see 37 CFR 1.33(a).

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	MLB-002/06US 34468/40
	Application Number	
Title of Invention	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS	

☐ **An Address is being provided for the correspondence Information of this application.**

Customer Number	21710		
Email Address	ip@brownrudnick.com	Add Email	Remove Email

Application Information:

Title of the Invention	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS		
Attorney Docket Number	MLB-002/06US 34468/40	**Small Entity Status Claimed**	☒
Application Type	Nonprovisional		
Subject Matter	Utility		
Total Number of Drawing Sheets (if any)	16	**Suggested Figure for Publication (if any)**	

Filing By Reference:

Only complete this section when filing an application by reference under 35 U.S.C. 111(c) and 37 CFR 1.57(a). Do not complete this section if application papers including a specification and any drawings are being filed. Any domestic benefit or foreign priority information must be provided in the appropriate section(s) below (i.e., "Domestic Benefit/National Stage Information" and "Foreign Priority Information").

For the purposes of a filing date under 37 CFR 1.53(b), the description and any drawings of the present application are replaced by this reference to the previously filed application, subject to conditions and requirements of 37 CFR 1.57(a).

Application number of the previously filed application	Filing date (YYYY-MM-DD)	Intellectual Property Authority or Country

Publication Information:

☐ Request Early Publication (Fee required at time of Request 37 CFR 1.219)

☒ **Request Not to Publish.** I hereby request that the attached application not be published under 35 U.S.C. 122(b) and certify that the invention disclosed in the attached application **has not and will not** be the subject of an application filed in another country, or under a multilateral international agreement, that requires publication at eighteen months after filing.

Representative Information:

Representative information should be provided for all practitioners having a power of attorney in the application. Providing this information in the Application Data Sheet does not constitute a power of attorney in the application (see 37 CFR 1.32).
Either enter Customer Number or complete the Representative Name section below. If both sections are completed the customer Number will be used for the Representative Information during processing.

Please Select One:	● Customer Number	US Patent Practitioner	○ Limited Recognition (37 CFR 11.9)
Customer Number	21710		

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	MLB-002/06US 34468/40
	Application Number	
Title of Invention	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS	

Domestic Benefit/National Stage Information:

This section allows for the applicant to either claim benefit under 35 U.S.C. 119(e), 120, 121, 365(c), or 386(c) or indicate National Stage entry from a PCT application. Providing benefit claim information in the Application Data Sheet constitutes the specific reference required by 35 U.S.C. 119(e) or 120, and 37 CFR 1.78.
When referring to the current application, please leave the "Application Number" field blank.

Prior Application Status	Pending		Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
	Continuation of	15/231687	2016-08-08
Prior Application Status	Pending		Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
15/231687	Continuation of	13/266862	2012-03-13
Prior Application Status			Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
13/266862	a 371 of international	US10/01293	2010-04-30
Prior Application Status	Expired		Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
US10/01293	Claims benefit of provisional	61/182089	2009-05-28
Prior Application Status	Expired		Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
US10/01293	Claims benefit of provisional	61/179358	2009-05-18
Prior Application Status	Expired		Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
US10/01293	Claims benefit of provisional	61/178923	2009-05-15
Prior Application Status	Expired		Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
US10/01293	Claims benefit of provisional	61/174470	2009-04-30

Additional Domestic Benefit/National Stage Data may be generated within this form by selecting the **Add** button. Add

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	MLB-002/06US 34468/40
	Application Number	
Title of Invention	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS	

Foreign Priority Information:

This section allows for the applicant to claim priority to a foreign application. Providing this information in the application data sheet constitutes the claim for priority as required by 35 U.S.C. 119(b) and 37 CFR 1.55. When priority is claimed to a foreign application that is eligible for retrieval under the priority document exchange program (PDX)[i] the information will be used by the Office to automatically attempt retrieval pursuant to 37 CFR 1.55(i)(1) and (2). Under the PDX program, applicant bears the ultimate responsibility for ensuring that a copy of the foreign application is received by the Office from the participating foreign intellectual property office, or a certified copy of the foreign priority application is filed, within the time period specified in 37 CFR 1.55(g)(1).

Remove

Application Number	Country[i]	Filing Date (YYYY-MM-DD)	Access Code[i] (if applicable)

Additional Foreign Priority Data may be generated within this form by selecting the **Add** button. Add

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications

☐ This application (1) claims priority to or the benefit of an application filed before March 16, 2013 and (2) also contains, or contained at any time, a claim to a claimed invention that has an effective filing date on or after March 16, 2013.
NOTE: By providing this statement under 37 CFR 1.55 or 1.78, this application, with a filing date on or after March 16, 2013, will be examined under the first inventor to file provisions of the AIA.

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	MLB-002/06US 34468/40
	Application Number	
Title of Invention	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS	

Authorization or Opt-Out of Authorization to Permit Access:

When this Application Data Sheet is properly signed and filed with the application, applicant has provided written authority to permit a participating foreign intellectual property (IP) office access to the instant application-as-filed (see paragraph A in subsection 1 below) and the European Patent Office (EPO) access to any search results from the instant application (see paragraph B in subsection 1 below).

Should applicant choose not to provide an authorization identified in subsection 1 below, applicant **must opt-out** of the authorization by checking the corresponding box A or B or both in subsection 2 below.

NOTE: This section of the Application Data Sheet is **ONLY** reviewed and processed with the **INITIAL** filing of an application. After the initial filing of an application, an Application Data Sheet cannot be used to provide or rescind authorization for access by a foreign IP office(s). Instead, Form PTO/SB/39 or PTO/SB/69 must be used as appropriate.

1. Authorization to Permit Access by a Foreign Intellectual Property Office(s)

A. Priority Document Exchange (PDX) - Unless box A in subsection 2 (opt-out of authorization) is checked, the undersigned hereby **grants the USPTO authority** to provide the European Patent Office (EPO), the Japan Patent Office (JPO), the Korean Intellectual Property Office (KIPO), the State Intellectual Property Office of the People's Republic of China (SIPO), the World Intellectual Property Organization (WIPO), and any other foreign intellectual property office participating with the USPTO in a bilateral or multilateral priority document exchange agreement in which a foreign application claiming priority to the instant patent application is filed, access to: (1) the instant patent application-as-filed and its related bibliographic data, (2) any foreign or domestic application to which priority or benefit is claimed by the instant application and its related bibliographic data, and (3) the date of filing of this Authorization. See 37 CFR 1.14(h)(1).

B. Search Results from U.S. Application to EPO - Unless box B in subsection 2 (opt-out of authorization) is checked, the undersigned hereby **grants the USPTO authority** to provide the EPO access to the bibliographic data and search results from the instant patent application when a European patent application claiming priority to the instant patent application is filed. See 37 CFR 1.14(h)(2).

The applicant is reminded that the EPO's Rule 141(1) EPC (European Patent Convention) requires applicants to submit a copy of search results from the instant application without delay in a European patent application that claims priority to the instant application.

2. Opt-Out of Authorizations to Permit Access by a Foreign Intellectual Property Office(s)

☐ A. Applicant **DOES NOT** authorize the USPTO to permit a participating foreign IP office access to the instant application-as-filed. If this box is checked, the USPTO will not be providing a participating foreign IP office with any documents and information identified in subsection 1A above.

☐ B. Applicant **DOES NOT** authorize the USPTO to transmit to the EPO any search results from the instant patent application. If this box is checked, the USPTO will not be providing the EPO with search results from the instant application.

NOTE: Once the application has published or is otherwise publicly available, the USPTO may provide access to the application in accordance with 37 CFR 1.14.

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	MLB-002/06US 34468/40
	Application Number	
Title of Invention	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS	

Applicant Information:

Providing assignment information in this section does not substitute for compliance with any requirement of part 3 of Title 37 of CFR to have an assignment recorded by the Office.

Applicant 1 — Remove

If the applicant is the inventor (or the remaining joint inventor or inventors under 37 CFR 1.45), this section should not be completed. The information to be provided in this section is the name and address of the legal representative who is the applicant under 37 CFR 1.43; or the name and address of the assignee, person to whom the inventor is under an obligation to assign the invention, or person who otherwise shows sufficient proprietary interest in the matter who is the applicant under 37 CFR 1.46. If the applicant is an applicant under 37 CFR 1.46 (assignee, person to whom the inventor is obligated to assign, or person who otherwise shows sufficient proprietary interest) together with one or more joint inventors, then the joint inventor or inventors who are also the applicant should be identified in this section.

Clear

● Assignee	Legal Representative under 35 U.S.C. 117	Joint Inventor
Person to whom the inventor is obligated to assign.	Person who shows sufficient proprietary interest	

If applicant is the legal representative, indicate the authority to file the patent application, the inventor is:

Name of the Deceased or Legally Incapacitated Inventor:

If the Applicant is an Organization check here. ☒

Organization Name	Molecular Loop Biosolutions, LLC		
Mailing Address Information For Applicant:			
Address 1	67 Reservoir Street #1		
Address 2			
City	Cambridge	**State/Province**	MA
Country	US	Postal Code	02138
Phone Number		Fax Number	
Email Address			

Additional Applicant Data may be generated within this form by selecting the Add button. Add

Assignee Information including Non-Applicant Assignee Information:

Providing assignment information in this section does not substitute for compliance with any requirement of part 3 of Title 37 of CFR to have an assignment recorded by the Office.

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	MLB-002/06US 34468/40
	Application Number	
Title of Invention	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS	

Assignee 1

Complete this section if assignee information, including non-applicant assignee information, is desired to be included on the patent application publication. An assignee-applicant identified in the "Applicant Information" section will appear on the patent application publication as an applicant. For an assignee-applicant, complete this section only if identification as an assignee is also desired on the patent application publication.

Remove

If the Assignee or Non-Applicant Assignee is an Organization check here. ☐

Prefix	**Given Name**	Middle Name	**Family Name**	Suffix

Mailing Address Information For Assignee including Non-Applicant Assignee:

Address 1			
Address 2			
City		**State/Province**	
Country i		Postal Code	
Phone Number		Fax Number	
Email Address			

Additional Assignee or Non-Applicant Assignee Data may be generated within this form by selecting the Add button. Add

Signature:

Remove

NOTE: This Application Data Sheet must be signed in accordance with 37 CFR 1.33(b). **However, if this Application Data Sheet is submitted with the INITIAL filing of the application and either box A or B is not checked in subsection 2 of the "Authorization or Opt-Out of Authorization to Permit Access" section, then this form must also be signed in accordance with 37 CFR 1.14(c).**

This Application Data Sheet **must** be signed by a patent practitioner if one or more of the applicants is a **juristic entity** (e.g., corporation or association). If the applicant is two or more joint inventors, this form must be signed by a patent practitioner, **all** joint inventors who are the applicant, or one or more joint inventor-applicants who have been given power of attorney (e.g., see USPTO Form PTO/AIA/81) on behalf of **all** joint inventor-applicants.

See 37 CFR 1.4(d) for the manner of making signatures and certifications.

Signature	/Thomas C. Meyers/			Date (YYYY-MM-DD)	2020-11-19
First Name	Thomas	Last Name	Meyers	Registration Number	36,989

Additional Signature may be generated within this form by selecting the Add button. Add

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	MLB-002/06US 34468/40
	Application Number	
Title of Invention	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS	

This collection of information is required by 37 CFR 1.76. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 23 minutes to complete, including gathering, preparing, and submitting the completed application data sheet form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1 The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these records.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3 A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent CooperationTreaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

Attorney Docket No.: MLB-002/06US 34468/40

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANT:	Molecular Loop Biosolutions, LLC	ART UNIT:	Not Assigned
SERIAL NUMBER:	--	CONF. NO.:	Not Assigned
FILING DATE:	--	EXAMINER:	Not Assigned
CONTINUITY:	Continuation of Application No. 15/231,687, filed August 8, 2016		
TITLE:	**METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS**		

FILED ELECTRONICALLY

Commissioner for Patents
P. O. Box 1450
Alexandria, VA 22313-1450

PRELIMINARY AMENDMENT

Please enter the following amendments to the specification and claims of the accompanying continuation application.

Applicant believes that no additional fees are due with respect to this Preliminary Amendment.

Amendments to the Specification begin on page 2.

Amendments to the Claims begin on page 3.

Remarks begin on page 5.

Amendments to the Specification:

Please amend the "Related Applications" section at page 1 of the as-filed specification as follows:

RELATED APPLICATIONS

This application is a continuation of U.S. Application No. 15/231,687, filed August 8, 2016, which is a continuation of U.S. Application No. 13/266,862, filed March 13, 2012, which is a national phase application that claims the benefit of, and priority to, international PCT Application No. PCT/US2010/001293, filed April 30, 2010, which claims the benefit under 35 U.S.C. §119(e) of U.S. Provisional Application No. ~~provisional application serial number~~ 61/174,470, filed April 30, 2009, U.S. Provisional Application No. ~~provisional application serial number~~ 61/178,923, filed May 15, 2009, U.S. Provisional Application No. ~~provisional application serial number~~ 61/179,358, filed May 18, 2009, and U.S. Provisional Application No. ~~provisional application serial number~~ 61/182,089, filed May 28, 2009, the entire contents of each of which are incorporated herein by reference.

Please amend page 27, lines 1-3 of the specification as follows:

~~FIGs. 4A, B, and C~~ FIGS. 4A-C depict various non-limiting methods for combining differentiator tag sequence and target sequences (NNNN depicts a differentiator tag sequence).

AMENDMENTS TO THE CLAIMS

Please amend the claims as follows. The following listing of claims will replace all prior versions of claims in the application.

1-67. (Canceled)

68. (New) A method for correcting for errors or bias introduced during nucleic acid analysis workflow, the method comprising the steps of:

obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin;

introducing a set of differentiator tags, wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag;

amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons;

sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and

correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count.

69. (New) The method of claim 68, wherein said correcting step further comprises collapsing all sequence reads comprising a same differentiator tag into a single read before determining a consensus base call for the target nucleic acid molecule sequences.

Petitioner Illumina, Inc., EX1004, p. 356

70. (New) The method of claim 69, further comprising the step of determining the presence of a sequencing or amplification error based upon a number of said differentiator tag sequences.

71. (New) The method of claim 68, wherein said biological sample is selected from the group consisting of blood and biopsy tissue.

72. (New) The method of claim 68, wherein said plurality of target nucleic acid molecules comprises bacterial or viral nucleic acids isolated from said biological sample.

73. (New) The method of claim 68, wherein said plurality of target nucleic acid molecules are circulating tumor nucleic acid molecules.

74. (New) The method of claim 68, further comprising the step of determining the copy number of a genomic region or transcript in a patient from whom the sample is obtained.

75. (New) The method of claim 68, wherein said plurality of target nucleic acid molecules are maternally-circulating fetal or placental nucleic acid molecules.

76. (New) A method for correcting for errors or bias introduced during nucleic acid analysis workflow, the method comprising the steps of:

obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin;

introducing a set of unique differentiator tags, wherein members of said set of unique differentiator tags are randomly associated with members of said plurality of target nucleic acid molecules, such that any given tag is uniquely associated with said locus of origin;

amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons;

sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and

correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule.

Petitioner Illumina, Inc., EX1004, p. 358

Remarks

With entry of the amendment, claims 68-76 will be pending in the application. Claims 1-67 are cancelled and claims 68-76 are new with support for the new claims being found throughout the as-filed application, for example at page 6 (lines 21-24), page 39 (21-26), page 40 (7-9), page 41 (3-22) and page 42 (lines 13-27). No new matter is added.

The specification is amended to update the "Related Applications" section and to update the figure labeling for FIGS. 4A-C.

The paper copy of the Sequence Listing in this application (Continuation of 15/231,687), is identical to the computer readable copy of the Sequence Listing filed in application 15/231,687. In accordance with 37 CFR 1.821(e), please use the computer-readable form filed in that application as the computer readable form for the instant application. It is understood that the Patent and Trademark Office will make the necessary change in application number and filing date for the instant application. A paper copy of the Sequence Listing filed in application 15/231,687 is included with this filing for reference.

If there are any questions regarding the above remarks, the Examiner is invited and encouraged to contact Applicant's representative.

Dated: November 19, 2020

Respectfully submitted,
BROWN RUDNICK LLP

BROWN RUDNICK LLP
One Financial Center
Boston, MA 02111
Tel: (617) 856-8483
Fax: (617) 856-8201

/Thomas C. Meyers/
Thomas C. Meyers, Reg. No. 36,989
Attorney for Applicant
Email: tmeyers@brownrudnick.com

63911571 v1

Electronic Patent Application Fee Transmittal	
Application Number:	
Filing Date:	
Title of Invention:	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS
First Named Inventor/Applicant Name:	Gregory Porreca
Filer:	Thomas C. Meyers/Sandra Richards
Attorney Docket Number:	MLB-002/06US 34468/40

Filed as Small Entity

Filing Fees for Track I Prioritized Examination - Nonprovisional Application under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
UTILITY FILING FEE (ELECTRONIC FILING)	4011	1	80	80
UTILITY SEARCH FEE	2111	1	350	350
UTILITY EXAMINATION FEE	2311	1	400	400
REQUEST FOR PRIORITIZED EXAMINATION	2817	1	2100	2100
Pages:				
Claims:				
Miscellaneous-Filing:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
PUBL. FEE- EARLY, VOLUNTARY, OR NORMAL	1504	1	0	0
PROCESSING FEE, EXCEPT PROV. APPLS.	2830	1	70	70
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				
Miscellaneous:				
		Total in USD ($)		**3000**

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
16/952,764	11/19/2020	Gregory Porreca	MLB-002/06US 34468/40

CONFIRMATION NO. 7812

21710
BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

FORMALITIES LETTER

OC000000121663696

Date Mailed: 12/01/2020

NOTICE TO FILE CORRECTED APPLICATION PAPERS

Filing Date Granted

An application number and filing date have been accorded to this application. The application is informal since it does not comply with the regulations for the reason(s) indicated below. Applicant is given **TWO MONTHS** from the date of this Notice within which to correct the informalities indicated below. Extensions of time may be obtained by filing a petition accompanied by the extension fee under the provisions of 37 CFR 1.136(a).

The required item(s) identified below must be timely submitted to avoid abandonment:

- A substitute specification excluding claims in compliance with 37 CFR 1.52, 1.121(b)(3), and 1.125 is required. The substitute specification must be submitted with markings and be accompanied by a clean version (without markings) as set forth in 37 CFR 1.125(c) and a statement that the substitute specification contains no new matter (see 37 CFR 1.125(b)). Since a preliminary amendment was present on the filing date of the application and such amendment is part of the original disclosure of the application, the substitute specification must include all of the desired changes made in the preliminary amendment. See 37 CFR 1.115 and 1.215.

Items 1) and 2) provide general guidance related to requirements for sequence disclosures.

1) 37 CFR 1.821(c) requires that patent applications which contain disclosures of nucleotide and/or amino acid sequences that fall within the definitions of 37 CFR 1.821(a) must contain a "Sequence Listing", *as a separate part of the disclosure*, which presents the nucleotide and/or amino acid sequences and associated information using the symbols and format in accordance with the requirements of 37 CFR 1.821-1.825. This "Sequence Listing" part of the disclosure may be submitted:
   a) via EFS-Web (see Section I.1 of the Legal Framework for EFS-Web (https://www.uspto.gov/patents-application-process/filing-online/legal-framework-efs-web), hereinafter "Legal Framework") as an ASCII text file, together with an incorporation-by-reference of the material in the ASCII text file in a separate paragraph of the specification identifying:
      i) the name of the ASCII text file;
      ii) the date of creation; and
      iii) the size of the ASCII text file in bytes;
   b) on compact disc(s) in duplicate according to 37 CFR 1.52(e)(1)(ii) and (4), labeled according to 37 CFR 1.52(e)(6), together with a statement that the duplicate compact discs are identical and an incorporation-by-reference of the material in the ASCII text file according to 37 CFR 1.52(e)(5) in a separate paragraph of the specification identifying:
      i) the name of the ASCII text file;

Petitioner Illumina, Inc., EX1004, p. 362

ii) the date of creation; and
iii) the size of the ASCII text file in bytes;
c) via EFS-Web as a PDF (not recommended); or
d) on paper.

2) 37 CFR 1.821(e) requires that a *copy* of the "Sequence Listing" must also be submitted in computer readable form (CRF) in accordance with the requirements of 37 CFR 1.824.
   a) If a "Sequence Listing" ASCII text file submitted via EFS-Web on the application filing date complies with the requirements of 37 CFR 1.824(a)(2)-(6) and (b), and applicant has not filed a "Sequence Listing" as a PDF, the text file will serve as both the "Sequence Listing" required by 37 CFR 1.821(c) and the CRF required by 37 CFR 1.821(e), and the statement of identity under the "Legal Framework" is not required.
   b) If the "Sequence Listing" required by 37 CFR 1.821(c) is filed via EFS-Web as a PDF, then the "Legal Framework" requires submission of a statement that the "Sequence Listing" content of the PDF copy and the ASCII text file copy submitted via EFS-Web are identical.
   c) If the "Sequence Listing" required by 37 CFR 1.821(c) is filed on paper or compact disc, then 37 CFR 1.821(f) requires submission of a statement that the "Sequence Listing" content of the paper or compact disc copy and the CRF are identical.

Specific deficiencies and the required response to this notice are as follows:

- A copy of the "Sequence Listing" in computer readable form (CRF) has not been submitted as required by 37 CFR 1.821(e) as indicated in item 2) above. Applicant must provide a CRF and statement according to item 2) b) or c) above. If applicant desires the CRF in this application to be identical with the CRF of another application on file in the U.S. Patent and Trademark Office, such request in accordance with 37 CFR 1.821(e) may be submitted in lieu of a new CRF (Form PTO/SB/93 may be used for this purpose).

For questions regarding compliance to these requirements, please contact:

- **For Rules interpretation, call (571) 272-2510**
- **For CRF submission help, call (571) 272-2510**
- **For PatentIn software program support technical assistance, call (866) 217-9197.**
- **Send e-mail correspondence for PatentIn software program help to ebc@uspto.gov**
- **PatentIn software is available at http://www.uspto.gov/patents/resources/tools/checker/patentinrel.jsp .**

Applicant is cautioned that correction of the above items may cause the specification and drawings page count to exceed 100 pages. If the specification and drawings exceed 100 pages, applicant will need to submit the required application size fee.

Replies must be received in the USPTO within the set time period or must include a proper Certificate of Mailing or Transmission under 37 CFR 1.8 with a mailing or transmission date within the set time period. For more information and a suggested format, see Form PTO/SB/92 and MPEP 512.

Replies should be mailed to:

Mail Stop Missing Parts
Commissioner for Patents
P.O. Box 1450
Alexandria VA 22313-1450

Registered users of EFS-Web may alternatively submit their reply to this notice via EFS-Web, including a copy of this Notice and selecting the document description "Applicant response to Pre-Exam Formalities Notice". https://sportal.uspto.gov/authenticate/AuthenticateUserLocalEPF.html

For more information about EFS-Web please call the USPTO Electronic Business Center at 1-866-217-9197 or visit our website at http://www.uspto.gov/ebc.

If you are not using EFS-Web to submit your reply, you must include a copy of this notice.

Questions about the contents of this notice and the requirements it sets forth should be directed to the Office of Data Management, Application Assistance Unit, at **(571) 272-4000** or **(571) 272-4200** or **1-888-786-0101**.

/afessehaye/

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING or 371(c) DATE	GRP ART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
16/952,764	11/19/2020		830	MLB-002/06US 34468/40	9	2

CONFIRMATION NO. 7812

21710
BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

FILING RECEIPT



Date Mailed: 12/01/2020

Receipt is acknowledged of this non-provisional utility patent application. The application will be taken up for examination in due course. Applicant will be notified as to the results of the examination. Any correspondence concerning the application must include the following identification information: the U.S. APPLICATION NUMBER, FILING DATE, NAME OF FIRST INVENTOR, and TITLE OF INVENTION. Fees transmitted by check or draft are subject to collection.

Please verify the accuracy of the data presented on this receipt. If an error is noted on this Filing Receipt, please submit a written request for a corrected Filing Receipt, including a properly marked-up ADS showing the changes with strike-through for deletions and underlining for additions. If you received a "Notice to File Missing Parts" or other Notice requiring a response for this application, please submit any request for correction to this Filing Receipt with your reply to the Notice. When the USPTO processes the reply to the Notice, the USPTO will generate another Filing Receipt incorporating the requested corrections provided that the request is grantable.

Inventor(s)

Gregory Porreca, Cambridge, MA;
Uri Laserson, Boston, MA;

Applicant(s)

Molecular Loop Biosolutions, LLC, Cambridge, MA;

Power of Attorney: None

Domestic Priority data as claimed by applicant

This application is a CON of 15/231,687 08/08/2016
which is a CON of 13/266,862 03/13/2012
which is a 371 of PCT/US10/01293 04/30/2010
which claims benefit of 61/182,089 05/28/2009
and claims benefit of 61/179,358 05/18/2009
and claims benefit of 61/178,923 05/15/2009
and claims benefit of 61/174,470 04/30/2009

Foreign Applications for which priority is claimed (You may be eligible to benefit from the **Patent Prosecution Highway** program at the USPTO. Please see http://www.uspto.gov for more information.) - None.
Foreign application information must be provided in an Application Data Sheet in order to constitute a claim to foreign priority. See 37 CFR 1.55 and 1.76.

Permission to Access Application via Priority Document Exchange: Yes

Permission to Access Search Results: Yes

Applicant may provide or rescind an authorization for access using Form PTO/SB/39 or Form PTO/SB/69 as appropriate.

If Required, Foreign Filing License Granted: 11/30/2020

The country code and number of your priority application, to be used for filing abroad under the Paris Convention, is **US 16/952,764**

Projected Publication Date: Request for Non-Publication Acknowledged

Non-Publication Request: Yes

**********WARNING**********

Applicant has requested non-publication, certifying that the invention disclosed in this application has not been and will not be the subject of an application filed in another country, or under a multilateral agreement, that requires publication at eighteen months after filing; **and has also claimed** the benefit of an application filed in another country or under a multilateral agreement. Because the non-publication request and corresponding certification appears to be inconsistent with the priority claim, applicant is reminded that he/she has an obligation to promptly notify the Office and **rescind the non-publication request** if the invention disclosed in this application is the subject of an application filed in another country, or under a multilateral agreement, that requires eighteen-month publication.

**********WARNING**********

Early Publication Request: No
**** SMALL ENTITY ****
Title

METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS

Preliminary Class

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications: No

PROTECTING YOUR INVENTION OUTSIDE THE UNITED STATES

Since the rights granted by a U.S. patent extend only throughout the territory of the United States and have no effect in a foreign country, an inventor who wishes patent protection in another country must apply for a patent in a specific country or in regional patent offices. Applicants may wish to consider the filing of an international application under the Patent Cooperation Treaty (PCT). An international (PCT) application generally has the same effect as a regular national patent application in each PCT-member country. The PCT process **simplifies** the filing of patent applications on the same invention in member countries, but **does not result** in a grant of "an international patent" and does not eliminate the need of applicants to file additional documents and fees in countries where patent protection is desired.

Almost every country has its own patent law, and a person desiring a patent in a particular country must make an application for patent in that country in accordance with its particular laws. Since the laws of many countries differ in various respects from the patent law of the United States, applicants are advised to seek guidance from specific foreign countries to ensure that patent rights are not lost prematurely.

Applicants also are advised that in the case of inventions made in the United States, the Director of the USPTO must issue a license before applicants can apply for a patent in a foreign country. The filing of a U.S. patent application

serves as a request for a foreign filing license. The application's filing receipt contains further information and guidance as to the status of applicant's license for foreign filing.

Applicants may wish to consult the USPTO booklet, "General Information Concerning Patents" (specifically, the section entitled "Treaties and Foreign Patents") for more information on timeframes and deadlines for filing foreign patent applications. The guide is available either by contacting the USPTO Contact Center at 800-786-9199, or it can be viewed on the USPTO website at http://www.uspto.gov/web/offices/pac/doc/general/index.html.

For information on preventing theft of your intellectual property (patents, trademarks and copyrights), you may wish to consult the U.S. Government website, http://www.stopfakes.gov. Part of a Department of Commerce initiative, this website includes self-help "toolkits" giving innovators guidance on how to protect intellectual property in specific countries such as China, Korea and Mexico. For questions regarding patent enforcement issues, applicants may call the U.S. Government hotline at 1-866-999-HALT (1-866-999-4258).

LICENSE FOR FOREIGN FILING UNDER

Title 35, United States Code, Section 184

Title 37, Code of Federal Regulations, 5.11 & 5.15

GRANTED

The applicant has been granted a license under 35 U.S.C. 184, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" followed by a date appears on this form. Such licenses are issued in all applications where the conditions for issuance of a license have been met, regardless of whether or not a license may be required as set forth in 37 CFR 5.15. The scope and limitations of this license are set forth in 37 CFR 5.15(a) unless an earlier license has been issued under 37 CFR 5.15(b). The license is subject to revocation upon written notification. The date indicated is the effective date of the license, unless an earlier license of similar scope has been granted under 37 CFR 5.13 or 5.14.

This license is to be retained by the licensee and may be used at any time on or after the effective date thereof unless it is revoked. This license is automatically transferred to any related applications(s) filed under 37 CFR 1.53(d). This license is not retroactive.

The grant of a license does not in any way lessen the responsibility of a licensee for the security of the subject matter as imposed by any Government contract or the provisions of existing laws relating to espionage and the national security or the export of technical data. Licensees should apprise themselves of current regulations especially with respect to certain countries, of other agencies, particularly the Office of Defense Trade Controls, Department of State (with respect to Arms, Munitions and Implements of War (22 CFR 121-128)); the Bureau of Industry and Security, Department of Commerce (15 CFR parts 730-774); the Office of Foreign AssetsControl, Department of Treasury (31 CFR Parts 500+) and the Department of Energy.

NOT GRANTED

No license under 35 U.S.C. 184 has been granted at this time, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" DOES NOT appear on this form. Applicant may still petition for a license under 37 CFR 5.12, if a license is desired before the expiration of 6 months from the filing date of the application. If 6 months has lapsed from the filing date of this application and the licensee has not received any indication of a secrecy order under 35 U.S.C. 181, the licensee may foreign file the application pursuant to 37 CFR 5.15(b).

Petitioner Illumina, Inc., EX1004, p. 367

SelectUSA

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The U.S. offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to promote and facilitate business investment. SelectUSA provides information assistance to the international investor community; serves as an ombudsman for existing and potential investors; advocates on behalf of U.S. cities, states, and regions competing for global investment; and counsels U.S. economic development organizations on investment attraction best practices. To learn more about why the United States is the best country in the world to develop technology, manufacture products, deliver services, and grow your business, visit http://www.SelectUSA.gov or call +1-202-482-6800.

Petitioner Illumina, Inc., EX1004, p. 368

Electronic Acknowledgement Receipt

EFS ID:	41287470
Application Number:	16952764
International Application Number:	
Confirmation Number:	7812
Title of Invention:	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS
First Named Inventor/Applicant Name:	Gregory Porreca
Customer Number:	21710
Filer:	Thomas C. Meyers/Sandra Richards
Filer Authorized By:	Thomas C. Meyers
Attorney Docket Number:	MLB-002/06US 34468/40
Receipt Date:	03-DEC-2020
Filing Date:	19-NOV-2020
Time Stamp:	15:25:11
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Applicant Response to Pre-Exam Formalities Notice	MLB-002-06US-Response-NTFCAP.pdf	15628 ac130cf4faf54771379969962f278655da1ed738	no	2

Warnings:

Information:					
2	Request for Transfer of a Computer Readable Form	MLB-002-06-CRF-Transfer-Req.pdf	255540 b151b4723dc02f5a08ad2aebf0c778e03fe5bc19	no	2
Warnings:					
Information:					
3	Specification	MLB-002-06US-Marked-Specification.pdf	2930349 5423fe1b6bac83665099b9e259f4a33dfe4b13d5	no	87
Warnings:					
Information:					
4	Specification	MLB-002-06US-Clean-Specification.pdf	2930161 99ba1e8479fb80a0c9cbbaa4a1b8b876ccbc0189	no	87
Warnings:					
Information:					
		Total Files Size (in bytes):	6131678		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

<u>New Applications Under 35 U.S.C. 111</u>
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.
<u>National Stage of an International Application under 35 U.S.C. 371</u>
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.
<u>New International Application Filed with the USPTO as a Receiving Office</u>
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Attorney Docket No.: MLB-002/06US 34468/40

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANT:	Molecular Loop Biosolutions, LLC	ART UNIT:	Not Assigned
SERIAL NUMBER:	16/952,764	CONF. NO.:	7812
FILING DATE:	November 19, 2020	EXAMINER:	Not Assigned
TITLE:	**METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS**		

FILED ELECTRONICALLY

Commissioner for Patents
P. O. Box 1450
Alexandria, VA 22313-1450

RESPONSE TO NOTICE TO FILE CORRECTED APPLICATION PAPERS

This paper is in response to the Notice to File Corrected Application Papers mailed from the United States Patent and Trademark Office on December 1, 2020. Applicant submits herewith both marked-up and clean copies of the specification as amended in the preliminary amendment submitted on November 19, 2020.

Regarding the sequence listing requirements of 37 CFR 1.821(e), as stated in the preliminary amendment submitted on November 19, 2020, Applicant wishes the computer readable form (CRF) of the sequence listing in this application to be identical with the CRF filed in application 15/231,687. To that end, Applicant submits herewith a Request for Transfer of a Computer Readable Form Under 37 CFR 1.821(e) (PTO/SB/93).

If there are any questions regarding the above remarks, the Examiner is invited and encouraged to contact Applicant's representative.

Dated: December 3, 2020

Respectfully submitted,

BROWN RUDNICK LLP

BROWN RUDNICK LLP
One Financial Center
Boston, MA 02111
Tel: (617) 856-8483
Fax: (617) 856-8201

/Thomas C. Meyers/
Thomas C. Meyers, Reg. No. 36,989
Attorney for Applicant
Email: tmeyers@brownrudnick.com

63918234 v1

Doc Code: CRF.TRNS.REQ
Document Description: Request for Transfer of a Computer Readable Form

PTO/SB/93 (11-14)
Approved for use through 09/30/2022. OMB 0651-0024
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

REQUEST FOR TRANSFER OF A COMPUTER READABLE FORM UNDER 37 CFR 1.821(e)

Application No.:	16/952,764	First Named Inventor:	Gregory Porreca
Filing Date:	November 19, 2020	Practitioner Docket No.:	MLB-002/06US 34468/40
Title of the Invention:	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS		

The sequence information in the paper copy or PDF file of the Sequence Listing submitted in this application in accordance with 37 CFR 1.821(c),

- ■ as part of the originally-filed specification of this application
- ☐ as a separate amendment of the specification filed
  - ☐ herewith
  - ☐ on _____

is identical to the sequence information in the

- ☐ only computer readable form
- ■ last filed computer readable form
- ☐ other computer readable form, which was filed on _____

in application number 15/231,687, which was filed on August 8, 2016.

This computer readable form was compliant with 37 CFR 1.821-1.825, and applicant hereby requests that it be used as the computer readable form for the present application, in accordance with 37 CFR 1.821(e).

The above referenced paper copy or PDF file of the Sequence Listing contains no new matter.

A sequence listing text file submitted via EFS-Web that complies with the requirements of 37 CFR 1.824(a) (2)-(6) and (b) (i.e., is a compliant sequence listing ASCII text file), serves as both the paper copy required by 37 CFR 1.821(c) and the CRF required by 37 CFR 1.821(e). If a user submits a compliant sequence listing ASCII text file via EFS-Web, the U.S. Patent and Trademark Office will not carry out a request to use a compliant computer readable "Sequence Listing" that is already on file for another application pursuant to 37 CFR 1.821(e) but will use the sequence listing submitted with the application as originally filed via EFS-Web.

It is understood that upon the transfer of a copy of the computer readable form to this application, the U.S. Patent and Trademark Office will update the copy of the computer readable form to reflect the application number and filing date for this application.

Signature /Thomas C. Meyers/	Date December 3, 2020
Name (Print/Typed) Thomas C. Meyers	Practitioner Registration Number 36,989

Note: *This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4(d) for signature requirements and certifications. Submit multiple forms if more than one signature is required.**

1 *Total of 1 forms are submitted.

This collection of information is required by 35 U.S.C. 22 and 37 CFR 1.821-1.825. The information is required to obtain or retain a benefit by the public, which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 6 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS.

Privacy Act Statement

The **Privacy Act of 1974 (P.L. 93-579)** requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether disclosure of these records is required by the Freedom of Information Act.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (*i.e.*, GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspection or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS

RELATED APPLICATIONS

This application is a continuation of U.S. Application No. 15/231,687, filed August 8, 2016, which is a continuation of U.S. Application No. 13/266,862, filed March 13, 2012, which is a national phase application that claims the benefit of, and priority to, international PCT Application No. PCT/US2010/001293, filed April 30, 2010, which claims the benefit under 35 U.S.C. §119(e) of U.S. Provisional Application No. ~~provisional application serial number~~ 61/174,470, filed April 30, 2009, U.S. Provisional Application No. ~~provisional application serial number~~ 61/178,923, filed May 15, 2009, U.S. Provisional Application No. ~~provisional application serial number~~ 61/179,358, filed May 18, 2009, and U.S. Provisional Application No. ~~provisional application serial number~~ 61/182,089, filed May 28, 2009, the entire contents of each of which are incorporated herein by reference.

FIELD OF INVENTION

The invention relates to methods and compositions for determining genotypes in patient samples.

BACKGROUND OF THE INVENTION

Information about the genotype of a subject is becoming more important and relevant for a range of healthcare decisions as the genetic basis for many diseases, disorders, and physiological characteristics is further elucidated. Medical advice is increasingly personalized, with individual decisions and recommendations being based on specific genetic information. Information about the type and number of alleles at one or more genetic loci impacts disease risk, prognosis, therapeutic options, and genetic counseling amongst other healthcare considerations.

For cost-effective and reliable medical and reproductive counseling on a large scale, it is important to be able correctly and unambiguously identify the allelic status for many different genetic loci in many subjects.

Petitioner Illumina, Inc., EX1004, p. 375

Numerous technologies have been developed for detecting and analyzing nucleic acid sequences from biological samples. These technologies can be used to genotype subjects and determine the allelic status of any locus of interest. However, they are not sufficiently robust and cost-effective to be scaled up for reliable high throughput analysis of many genetic loci in large numbers of patients. The frequency of incorrect or ambiguous calls is too high for current technology to manage large numbers of patient samples without involving expensive and time-consuming steps to resolve uncertainties and provide confidence in the information output.

SUMMARY OF THE INVENTION

Aspects of the invention relate to preparative and analytical methods and compositions for evaluating genotypes, and in particular, for determining the allelic identity (or identities in a diploid organism) of one or more genetic loci in a subject.

Aspects of the invention are based, in part, on the identification of different sources of ambiguity and error in genetic analyses, and, in part, on the identification of one or more approaches to avoid, reduce, recognize, and/or resolve these errors and ambiguities at different stages in a genetic analysis.

According to aspects of the invention, certain types of genetic information can be under-represented or over-represented in a genetic analysis due to a combination of stochastic variation and systematic bias in any of the preparative stages (e.g., capture, amplification, etc.), determining stages (e.g., allele-specific detection, sequencing, etc.), data interpretation stages (e.g., determining whether the assay information is sufficient to identify a subject as homozygous or heterozygous), and/or other stages.

According to aspects of the invention, error or ambiguity may be apparent in a genetic analysis, but not readily resolved without running additional samples or more expensive assays (*e.g.*, array-based assays may report no-calls due to noisy/low signal). According to further aspects of the invention, error or ambiguity may not be accounted for in a genetic analysis and incorrect base calls may be made even when the evidence for them is limited and/or not statistically significant (*e.g.*, next-generation sequencing technologies may report base calls even if the evidence for them is not statistically significant). According to further aspects of the invention error or ambiguity may be problematic for a multi-step genetic analysis because it is

Petitioner Illumina, Inc., EX1004, p. 376

apparent but not readily resolved in one or more steps of the analysis and not apparent or accounted for in other steps of the analysis.

In some embodiments, sources of error and ambiguity in one or more steps can be addressed by capturing and/or interrogating each target locus of interest with one or more sets of overlapping probes that are designed to overcome any systematic bias or stochastic effects that may impact the complexity and/or fidelity of the genetic information that is generated.

In some embodiments, sources of error and ambiguity in one or more steps can be addressed by capturing and/or interrogating each target locus of interest with at least one set of probes, wherein different probes are labeled with different identifiers that can be used to track the assay reactions and determine whether certain types of genetic information are under-represented or over-represented in the information that is generated.

In some embodiments, errors and ambiguities associated with the analysis of regions containing large numbers of sequence repeats are addressed by systematically analyzing frequencies of certain nucleic acids at particular stages in an assay (e.g., at a capture, sequencing, or detection stage). It should be appreciated that such techniques may be particularly useful in the context of a standardized protocol that is designed to allow many different loci to be evaluated in parallel without requiring different assay procedures for each locus. In some embodiments, the use of a single detection modality (*e.g.*, sequencing) to assay multiple types of genetic lesions (*e.g.*, point mutations, insertions/deletions, length polymorphisms) is advantageous in the clinical setting. In some embodiments of the invention, methods are provided that facilitate the use of multiple sample preparation steps in parallel, coupled with multiple analytical processes following sequence detection. Thus, in some embodiments of the invention, an improved workflow is provided that reduces error and uncertainty when simultaneously assaying different types of genetic lesions across multiple loci in multiple patients.

In some embodiments, aspects of the invention provide methods for overcoming preparative and/or analytical bias by combining two or more techniques, each having a different bias (e.g., a known bias towards under-representation or over-representation of one or more types of sequences), and using the resulting data to determine a genetic call for a subject with greater confidence.

Petitioner Illumina, Inc., EX1004, p. 377

It should be appreciated that in some embodiments, aspects of the invention relate to multiplex diagnostic methods. In some embodiments, multiplex diagnostic methods comprise capturing a plurality of genetic loci in parallel (e.g., one or more genetic loci from Table 1). In some embodiments, the genetic loci possess one or more polymorphisms (e.g., one or more polymorphisms from Table 2) the genotypes of which correspond to disease causing alleles. Accordingly, in some embodiments, the disclosure provides methods for assessing multiple heritable disorders in parallel. In some embodiments, methods are provided for diagnosing multiple heritable disorders in parallel at a pre-implantation, prenatal, perinatal, or postnatal stage. In some embodiments, the disclosure provides methods for analyzing multiple genetic loci (e.g., a plurality of target nucleic acids selected from Table 1) from a patient sample, such as a blood, pre-implantation embryo, chorionic villus or amniotic fluid sample, or other sample (e.g., other biological fluid or tissue sample such as a biopsy sample) as aspects of the invention are not limited in this respect.

Other samples may include tumor tissue or circulating tumor cells. In some embodiments, a patient sample (*e.g.*, a tumor tissue or cell sample) is mosaic for one or more mutations of interest, and thus, may require higher sensitivity than is needed for a germline mutation analysis. In some embodiments, a sample comprises cells from a non-host organism (*e.g.*, bacterial or viral infections in a human subject) or a sample for environmental monitoring (*e.g.*, bacterial, viral, fungal composition of a soil, water, or air sample).

Accordingly, in some embodiments, aspects of the methods disclosed herein relate to genotyping a polymorphism of a target nucleic acid. In some embodiments, the genotyping may comprise determining that one or more alleles of the target nucleic acid are heterozygous or homozygous. In further embodiments, the genotyping may comprise determining the sequence of a polymorphism and comparing that sequence to a control sequence that is indicative of a disease risk. In some embodiments, the polymorphism is selected from a locus in Table 1 or Table 2. However, it should be appreciated that any locus associated with a disease or condition of interest may be used.

In some embodiments, a diagnosis, prognosis, or disease risk assessment is provided to a subject based on a genotype determined for that subject at one or more genetic loci (e.g., based on the analysis of a biological sample obtained from that subject). In some embodiments, an assessment is provided to a couple, based on their respective genotypes at one or more genetic

Petitioner Illumina, Inc., EX1004, p. 378

loci, of the risk of their having one or more children having a genotype associated with a disease or condition (e.g., a homozygous or heterozygous genotype associated with a disease or condition). In some embodiments, a subject or a couple may seek genetic or reproductive counseling in connection with a genotype determined according to embodiments of the invention. In some embodiments, genetic information from a tumor or circulating tumor cells is used to determine prognosis and guide selection of appropriate drugs/treatments.

It should be appreciated that any of the methods or compositions described herein may be used in combination with any of the medical evaluations associated with one or more genetic loci as described herein.

In some embodiments, aspects of the invention provide effective methods for overcoming challenges associated with systematic errors (bias) and/or stochastic effects in multiplex genomic capture and/or analysis (including sequencing analysis). In some embodiments, aspects of the invention are useful to avoid, reduce and/or account for variability in one or more sampling and/or analytical steps. For example, in some embodiments, variability in target nucleic acid representation and unequal sampling of heterozygous alleles in pools of captured target nucleic acids can be overcome.

Accordingly, in some embodiments, the disclosure provides methods that reduce variability in the detection of target nucleic acids in multiplex capture methods. In other embodiments, methods improve allelic representation in a capture pool and, thus, improve variant detection outcomes. In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve the use of different sets of multiple probes (e.g., molecular inversion probes MIPs) that capture overlapping regions of a target nucleic acid to achieve a more uniform representation of the target nucleic acids in a capture pool compared with methods of the prior art. In other embodiments, methods reduce bias, or the risk of bias, associated with large scale parallel capture of genetic loci, e.g., for diagnostic purposes. In other embodiments, methods are provided for increasing reproducibility (e.g., by reducing the effect of polymorphisms on target nucleic acid capture) in the detection of a plurality of genetic loci in parallel. In further embodiments, methods are provided for reducing the effect of probe synthesis and/or probe amplification variability on the analysis of a plurality of genetic loci in parallel.

Petitioner Illumina, Inc., EX1004, p. 379

According to some aspects, methods of analyzing a plurality of genetic loci are provided. In some embodiments, the methods comprise contacting each of a plurality of target nucleic acids with a probe set, wherein each probe set comprises a plurality of different probes, each probe having a central region flanked by a 5' region and a 3' region that are complementary to nucleic acids flanking the same strand of one of a plurality of subregions of the target nucleic acid, wherein the subregions of the target nucleic acid are different, and wherein each subregion overlaps with at least one other subregion, isolating a plurality of nucleic acids each having a nucleic acid sequence of a different subregion for each of the plurality of target nucleic acids, and analyzing the isolated nucleic acids.

In other embodiments, methods comprise contacting each of a plurality of target nucleic acids with a probe set, wherein each probe set comprises a plurality of different probes, each probe having a central region flanked by a 5' region and a 3' region that are complementary to nucleic acids flanking the same strand of one of a plurality of subregions of the target nucleic acid, wherein the subregions of the target nucleic acid are different, and wherein a portion of the 5' region and a portion of the 3' region of a probe have, respectively, the sequence of the 5' region and the sequence of the 3' region of a different probe, isolating a plurality of nucleic acids each having a nucleic acid sequence of a different subregion for each of the plurality of target nucleic acids, and analyzing the isolated nucleic acids.

Aspects of the disclosure are based, in part, on the discovery of methods for overcoming problems associated with systematic and random errors (bias) in genome capture, amplification and sequencing methods, namely high variability in the capture and amplification of nucleic acids and disproportionate representation of heterozygous alleles in sequencing libraries. Accordingly, in some embodiments, the disclosure provides methods that reduce errors associated with the variability in the capture and amplification of nucleic acids. In other embodiments, the methods improve allelic representation in sequencing libraries and, thus, improve variant detection outcomes. In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve the use of differentiator tag sequences to uniquely tag individual nucleic acid molecules. In some embodiments, the differentiator tag sequence permit the detection of bias based on the occurrence of combinations of differentiator tag and target sequences observed in a sequencing reaction. In other embodiments, the methods reduce errors caused by bias, or the risk of bias,

Petitioner Illumina, Inc., EX1004, p. 380

associated with the capture, amplification and sequencing of genetic loci, e.g., for diagnostic purposes.

Aspects of the invention relate to providing sequence tags (referred to as differentiator tags) that are useful to determine whether target nucleic acid sequences identified in an assay are from independently isolated target nucleic acids or from multiple copies of the same target nucleic acid molecule (e.g., due to bias in a preparative step, for example, amplification). This information can be used to help analyze a threshold number of independently isolated target nucleic acids from a biological sample in order to obtain sequence information that is reliable and can be used to make a genotype conclusion (e.g., call) with a desired degree of confidence. This information also can be used to detect bias in one or more nucleic acid preparative steps.

In some embodiments, the methods disclosed herein are useful for any application where reduction of bias, e.g., associated with genomic isolation, amplification, sequencing, is important. For example, detection of cancer mutations in a heterogeneous tissue sample, detection of mutations in maternally-circulating fetal DNA, and detection of mutations in cells isolated during a preimplantation genetic diagnostic procedure.

Accordingly, in some aspects, methods of genotyping a subject are provided. In some embodiments, the methods comprise determining the sequence of at least a threshold number of independently isolated nucleic acids, wherein the sequence of each isolated nucleic acid comprises a target nucleic acid sequence and a differentiator tag sequence, wherein the threshold number is a number of unique combinations of target nucleic acid and differentiator tag sequences, wherein the isolated nucleic acids are identified as independently isolated if they comprise unique combinations of target nucleic acid and differentiator tag sequences, and wherein the target nucleic acid sequence is the sequence of a genomic locus of a subject.

In some embodiments, the isolated nucleic acids are products of a circularization selection-based preparative method, e.g., molecular inversion probe capture products. In other embodiments, the isolated nucleic acids are products of an amplification-based preparative methods. In other embodiments, the isolated nucleic acids are products of hybridization-based preparative methods.

Circularization selection-based preparative methods selectively convert regions of interest (target nucleic acids) into a covalently-closed circular molecule which is then isolated typically by removal (usually enzymatic, e.g. with exonuclease) of any non-circularized linear

Petitioner Illumina, Inc., EX1004, p. 381

nucleic acid. Oligonucleotide probes (e.g., molecular inversion probes) are designed which have ends that flank the region of interest (target nucleic acid) and, optionally, primer sites, e.g., sequencing primer sites. The probes are allowed to hybridize to the genomic target, and enzymes are used to first (optionally) fill in any gap between probe ends and second ligate the probe closed. Following circularization, any remaining (non-target) linear nucleic acid is typically removed, resulting in isolation (capture) of target nucleic acid. Circularization selection-based preparative methods include molecular inversion probe capture reactions and 'selector' capture reactions. In some embodiments, molecular inversion probe capture of a target nucleic acid is indicative of the presence of a polymorphism in the target nucleic acid.

In amplification-based (e.g., PCR-based or LCR-based, etc.) preparative methods, genomic loci (target nucleic acids) are isolated directly by means of a polymerase chain reaction or ligase chain reaction (or other amplification method) that selectively amplifies each locus using one or more oligonucleotide primers. It is to be understood that primers will be sufficiently complementary to the target sequence to hybridize with and prime amplification of the target nucleic acid. Any one of a variety of art known methods may be utilized for primer design and synthesis. One or more of the primers may be perfectly complementary to the target sequence. Degenerate primers may also be used. Primers may also include additional nucleic acids that are not complementary to target sequences but that facilitate downstream applications, including for example restriction sites and differentiator tag sequences. Amplification-based methods include amplification of a single target nucleic acid and multiplex amplification (amplification of multiple target nucleic acids in parallel).

Hybridization-based preparative methods involve selectively immobilizing target nucleic acids for further manipulation. It is to be understood that one or more oligonucleotides (immobilization oligonucleotides), which comprise differentiator tag sequences, and which may be from 15 to 170 nucleotides in length, are used which hybridize along the length of a target region of a genetic locus to immobilize it. In some embodiments, immobilization oligonucleotides, are either immobilized before hybridization is performed (e.g., Roche/Nimblegen 'sequence capture'), or are prepared such that they include a moiety (e.g. biotin) which can be used to selectively immobilize the target nucleic acid after hybridization by binding to e.g., streptavidin-coated microbeads (e.g. Agilent 'SureSelect').

Petitioner Illumina, Inc., EX1004, p. 382

It should be appreciated that any of the circularization, amplification, and/or hybridization based methods described herein may be used in connection with one or more of the tiling/staggering, tagging, size-detection, and/or sensitivity enhancing algorithms described herein.

In some embodiments, the methods disclosed herein comprise determining the sequence of molecular inversion probe capture products, each comprising a molecular inversion probe and a target nucleic acid, wherein the sequence of the molecular inversion probe comprises a differentiator tag sequence and, optionally, a primer sequence, and wherein the target nucleic acid is a captured genomic locus of a subject, and genotyping the subject at the captured genomic locus based on the sequence of at least a threshold number of unique combinations of target nucleic acid and differentiator tag sequences of molecular inversion probe capture products.

In some embodiments, the methods disclosed herein comprise obtaining molecular inversion probe capture products, each comprising a molecular inversion probe and a target nucleic acid, wherein the sequence of the molecular inversion probe comprises a differentiator tag sequence and, optionally, a primer sequence, wherein the target nucleic acid is a captured genomic locus of the subject, amplifying the molecular inversion probe capture products, and genotyping the subject by determining, for each target nucleic acid, the sequence of at least a threshold number of unique combinations of target nucleic acid and differentiator tag sequence of molecular inversion probe capture products. In certain embodiments, obtaining comprises capturing target nucleic acids from a genomic sample of the subject with molecular inversion probes, each comprising a unique differentiator tag sequence. In specific embodiments, capturing is performed under conditions wherein the likelihood of obtaining two or more molecular inversion probe capture products with identical combinations of target and differentiator tag sequences is equal to or less than a predetermined value, optionally wherein the predetermined value is about 0.05.

In one embodiment, the threshold number for a specific target nucleic acid sequence is selected based on a desired statistical confidence for the genotype. In some embodiments, the methods further comprising determining a statistical confidence for the genotype based on the number of unique combinations of target nucleic acid and differentiator tag sequences.

According to some aspects, methods of analyzing a plurality of genetic loci are provided. In some embodiments, the methods comprise obtaining a plurality of molecular inversion probe

Petitioner Illumina, Inc., EX1004, p. 383

capture products each comprising a molecular inversion probe and a target nucleic acid, wherein the sequence of the molecular inversion probe comprises a differentiator tag sequence and, optionally, a primer sequence (e.g., a sequence that is complementary to the sequence of a nucleic acid that is used as a primer for sequencing or other extension reaction), amplifying the plurality of molecular inversion probe capture products, determining numbers of occurrence of combinations of target nucleic acid and differentiator tag sequence of molecular inversion probe capture products in the amplified plurality, and if the number of occurrence of a specific combination of target nucleic acid sequence and differentiator tag sequence exceeds a predetermined value, detecting bias in the amplification of the molecular inversion probe comprising the specific combination. In some embodiments, the methods further comprise genotyping target sequences in the plurality, wherein the genotyping comprises correcting for bias, if detected.

In some embodiments, the target nucleic acid is a gene (or portion thereof) selected from Table 1. In some embodiments, the genotyping comprises determining the sequence of a target nucleic acid (e.g., a polymorphic sequence) at one or more (both) alleles of a genome (a diploid genome) of a subject. In certain embodiments, the genotyping comprises determining the sequence of a target nucleic acid at both alleles of a diploid genome of a subject, wherein in the target nucleic acid comprises, or consists of, a sequence of Table 1, Table 2, or other locus of interest.

In some embodiments, aspects of the invention provide methods and compositions for identifying nucleic acid insertions or deletions in genomic regions of interest without determining the nucleotide sequences of these regions. Aspects of the invention are particularly useful for detecting nucleic acid insertions or deletions in genomic regions containing nucleic acid sequence repeats (e.g., di- or tri-nucleotide repeats). However, the invention is not limited to analyzing nucleic acid repeats and may be used to detect insertions or deletions in any target nucleic acid of interest. Aspects of the invention are particularly useful for analyzing multiple loci in a multiplex assay.

In some embodiments, aspects of the invention relate to determining whether an amount of target nucleic acid that is captured in a genomic capture assay is higher or lower than expected. In some embodiments, a statistically significant deviation from an expected amount (e.g., higher or lower) is indicative of the presence of a nucleic acid insertion or deletion in the

Petitioner Illumina, Inc., EX1004, p. 384

genomic region of interest. In some embodiments, the amount is a number of nucleic acid molecules that are captured. In some embodiments, the amount is a number of independently captured nucleic acid molecules in a sample. It should be appreciated that the captured nucleic acids may be literally captured from a sample, or their sequences may be captured without actually capturing the original nucleic acids in the sample. For example, nucleic acid sequences may be captured in an assay that involves a template-based extension of nucleic acids having the region of interest, in the sample.

Aspects of the invention are based on the recognition that the efficiency of certain capture techniques is affected by the length of the nucleic acid being captured.

Accordingly, an increase or decrease in the length of a target nucleic acid (e.g., due to an insertion or deletion of a repeated sequence) can alter the capture efficiency of that nucleic acid. In some embodiments, a difference in the capture efficiency (e.g., a statistically significant difference in the capture efficiency) of a target nucleic acid is indicative of an insertion or deletion in the target nucleic acid. It should be appreciated that the capture efficiency for a target nucleic acid may be evaluated based on an amount of captured nucleic acid (e.g., number of captured nucleic acid molecules) relative to a control amount (e.g., based on an amount of control nucleic acid that is captured). However, the invention is not limited in this respect and other techniques for evaluating capture efficiency also may be used.

According to aspects of the invention, evaluating the capture efficiency as opposed to determining the sequence of the entire repeat region reduces errors associated with sequencing through repeat regions. Repeat sequences often give rise to stutters or skips in sequencing reactions that make it very difficult to accurately determine the number of repeats in a target region without running multiple sequencing reactions under different conditions and carefully analyzing the results. Such procedures are cumbersome and not readily scalable in a manner that is consistent with high throughput analyses of target nucleic acids. In some embodiments, repeat regions may be longer than the length of the individual sequence read, making length determination on the basis of a single read impossible. For example, when using next-generation sequencing the repeat regions may be longer than the length of the individual sequence read, making length determination on the basis of a single read impossible. Accordingly, aspects of the invention are useful to increase the sensitivity of detecting insertions or deletions in target regions, particularly target regions containing repeated sequences.

Petitioner Illumina, Inc., EX1004, p. 385

In some embodiments, aspects of the invention relate to capturing genomic nucleic acid sequences using a molecular inversion probe (e.g., MIP or Padlock probe) technique, and determining whether the amount (e.g., number) of captured sequences is higher or lower than expected. In some embodiments, the amount (e.g., number) of captured sequences is compared to an amount (e.g., number) of sequences captured in a control assay. The control assay may involve analyzing a control sample that contains a nucleic acid from the same genetic locus having a known sequence length (e.g., a known number of nucleic acid repeats). However, a control may involve analyzing a second (e.g., different) genetic locus that is not expected to contain any insertions or deletions. The second genetic locus may be analyzed in the same sample as the locus being interrogated or in a different sample where its length has been previously determined. The second genetic locus may be a locus that is not characterized by the presence of nucleic acid repeats (and thus not expected to contain insertions or deletions of the repeat sequence).

In some embodiments, a target nucleic acid region that is being evaluated may be determined by the identity of the targeting arms of a probe that is designed to capture the target region (or sequence thereof). For example, the targeting arms of a MIP probe may be designed to be complementary (e.g., sufficiently complementary for selective hybridization and//or polymerase extension and/or ligation) to genomic regions flanking a target region suspected of containing an insertion or deletion. It should be appreciated that two targeting arms may be designed to be complementary (e.g., sufficiently complementary for selective hybridization and/or polymerase extension and/or ligation) to the two flanking regions that are immediately adjacent (e.g., immediately 5' and 3', respectively) to a region of a sequence repeat on one strand of a genomic nucleic acid. However, one or both targeting arms may be designed to hybridize several bases (e.g., 1-5, 5-10, 10-25, 25-50, or more) upstream or downstream from the repeat region in such a way that the captured sequence includes a region of unique genomic sequence that on one or both sides of the repeat region. This unique region can then be used to identify the captured target (e.g., based on sequence or hybridization information).

In some embodiments, two or more (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10 or more) different loci may be interrogated in parallel in a single assay (e.g., in a multiplex assay). In some embodiments, the ratio of captured nucleic acids for each locus may be used to determine whether a nucleic acid insertion or deletion is present in one locus relative to the other. For

Petitioner Illumina, Inc., EX1004, p. 386

example, the ratio may be compared to a control ratio that is representative of the two loci when neither one has an insertion or deletion relative to control sequences (e.g., sequences that are normal or known to be associated with healthy phenotypes for those loci). However, the amount of captured nucleic acids may be compared to any suitable control as discussed herein.

The locus of a captured sequence may be identified by determining a portion of unique sequence 5' and/or 3' to the repeat region in the target nucleic acid suspected of containing a deletion or insertion. This does not require sequencing the captured repeat region itself. However, some or all of the repeat region also could be sequenced as aspects of the invention are not limited in this respect.

Aspects of the invention may be combined with one or more sequence-based assays (e.g., SNP detection assays), for example in a multiplex format, to determine the genotype of one or more regions of a subject.

In some embodiments, methods of detecting a polymorphism in a nucleic acid in a biological sample are provided. In some embodiments, the methods comprise evaluating the efficiency of capture at one or more loci and determining whether one or both alleles at that locus contain an insertion or deletion relative to a control locus (e.g., a locus indicative of a length of repeat sequence that is associated with a healthy phenotype).

Accordingly, aspects of the invention relate to methods for determining whether a target nucleic acid has an abnormal length by evaluating the capture efficiency of a target nucleic acid in a biological sample from a subject, wherein a capture efficiency that is different from a reference capture efficiency is indicative of the presence, in the biological sample, of a target nucleic acid having an abnormal length. It should be appreciated that the term "abnormal" is a relative term based on a comparison to a "normal" length. In some embodiments, a normal length is a length that is associated with a normal (e.g., healthy or non-carrier phenotype). Accordingly, an abnormal length is a length that is either shorter or longer than the normal length. In some embodiments, the presence of an abnormal length is indicative of an increased risk that the locus is associated with a disease or a disease carrier phenotype. In some embodiments, the abnormal length is indicative that the subject is either has a disease or condition or is a carrier of a disease or condition (e.g., associated with the locus). However, it should be appreciated that the description of embodiments relating to detecting the presence of

Petitioner Illumina, Inc., EX1004, p. 387

an abnormal length also support detecting the presence of a length that is different from an expected or control length.

In some embodiments, aspects of the invention relate to estimating the length of a target nucleic acid (e.g., of a sub-target region within a target nucleic acid). In some embodiments, aspects of the invention relate to methods for estimating the length of a target nucleic acid by contacting the target nucleic acid with a plurality of detection probes under conditions that permit hybridization of the detection probes to the target nucleic acid, wherein each detection probe is a polynucleotide that comprises a first arm that hybridizes to a first region of the target nucleic acid and a second arm that hybridizes to a second region of the target nucleic acid, wherein the first and second regions are on a common strand of the target nucleic acid, and wherein the nucleotide sequence of the target between the 5' end of the first region and the 3' end of the second region is the nucleotide sequence of a sub-target nucleic acid; and capturing a plurality of sub-target nucleic acids that are hybridized with the plurality of detection probes; and measuring the frequency of occurrence of a sub-target nucleic acid in the plurality of sub-target nucleic acids, wherein the frequency of occurrence of the sub-target nucleic acid in the plurality of sub-target nucleic acids is indicative of the length of the sub-target nucleic acid. It should be appreciated that methods for estimating a nucleic acid length may involve comparing a capture efficiency for a target nucleic acid region to two or more reference efficiencies for known nucleic acid lengths in order to determine whether the target nucleic acid region is smaller, intermediate, or larger in size than the known control lengths. In some embodiments, a series of nucleic acids of known different lengths may be used to provide a calibration curve for evaluating the length of a target nucleic acid region of interest.

In some embodiments, the capture efficiency of a target region suspected of having a deletion or insertion is determined by comparing the capture efficiency to a reference indicative of a normal capture efficiency. In some embodiments, the capture efficiency is lower than the reference capture efficiency. In some embodiments, the subject is identified as having an insertion in the target region. In some embodiments, the capture efficiency is higher than the reference capture efficiency. In some embodiments, the subject is identified as having a deletion in the target region. In some embodiments, the subject is identified as being heterozygous for the insertion. In some embodiments, the subject is identified as being heterozygous for the deletion.

Petitioner Illumina, Inc., EX1004, p. 388

In some embodiments of any of the methods described herein (e.g., tiling/staggering, tagging, size-detection, and/or sensitivity enhancement) aspects of the invention relate to capturing a sub-target nucleic acid (or a sequence of a sub-target nucleic acid). In some embodiments, a molecular inversion probe technique is used. In some embodiments, a molecular inversion probe is a single linear strand of nucleic acid that comprises a first targeting arm at its 5' end and a second targeting arm at its 3' end, wherein the first targeting arm is capable of specifically hybridizing to a first region flanking one end of the sub-target nucleic acid, and wherein the second targeting arm is capable of specifically hybridizing to a second region flanking the other end of the sub target nucleic acid on the same strand of the target nucleic acid. In some embodiments, the first and second targeting arms are between about 10 and about 100 nucleotides long. In some embodiments, the first and second targeting arms are about 10-20, 20-30, 30-40, or 40-50 nucleotides long. In some embodiments, the first and second targeting arms are about 20 nucleotides long. In some embodiments, the first and second targeting arms have the same length. In some embodiments, the first and second targeting arms have different lengths. In some embodiments, each pair of first and second targeting arms in a set of probes has the same length. Accordingly, if one of the targeting arms is longer, the other one is correspondingly shorter. This allows for a quality control step in some embodiments to confirm that all captured probe/target sequence products have the same length after a multiplexed plurality of capture reactions. In some embodiments, a set of probes may be designed to have the same length if the intervening region is varied to accommodate any differences in the length of either one or both of the first and second targeting arms.

In some embodiments, the hybridization Tms of the first and second targeting arms are similar. In some embodiments, the hybridization Tms of the first and second targeting arms are within 2-5 °C of each other. In some embodiments, the hybridization Tms of the first and second targeting arms are identical. In some embodiments, the hybridization Tms of the first and second targeting arms are close to empirically-determined optima but not necessarily identical.

In some embodiments, the first and second targeting arms of a molecular inversion probe have different Tms. For example, the Tm of the first targeting arm (at the 5' end of the molecular inversion probe) may be higher than the Tm of the second targeting arm (at the 3' end of the molecular inversion probe). According to aspects of the invention, and without wishing to be bound by theory, a relatively high Tm for the first targeting arm may help avoid or prevent the

Petitioner Illumina, Inc., EX1004, p. 389

first targeting arm from being displaced after hybridization by the extension product of the 3' end of the second targeting arm. It should be appreciated that a reference to the Tm of a targeting arm as used herein relates to the Tm of hybridization of the targeting arm to a nucleic acid having the complementary sequence (e.g., the region of the target nucleic acid that has a sequence that is complementary to the sequence of the targeting arm). It also should be appreciated that the Tms of the targeting arms described herein may be calculated using any appropriate method. For example, in some embodiments an experimental method (e.g., a gel shift assay, a hybridization assay, a melting curve analysis, for example in a PCR machine with a SYBR dye by stepping through a temperature ramp while monitoring signal level from an intercalating dye, for example, bound to a double-stranded DNA, etc.) may be used to determine one or more Tms empirically. In some embodiments, an optimal Tm may be determined by evaluating the number of products formed (e.g., for each of a plurality of MIP probes), and determining the optimal Tm as the center point in a histogram of Tm for all targeting arms. In some embodiments, a predictive algorithm may be used to determine a Tm theoretically. In some embodiments, a relatively simple predictive algorithm may be used based on the number of G/C and A/T base pairs when the sequence is hybridized to its target and/or the length of the hybridized product (e.g., for example, 64.9 + 41 * ([G+C] - 16.4) / (A + T + G + C), see for example, Wallace,R.B., Shaffer,J., Murphy,R.F., Bonner,J., Hirose,T., and Itakura,K. (1979) Nucleic Acids Res 6:3543-3557). In some embodiments, a more complex algorithm may be used to account for the effects of base stacking entropy and enthalpy, ion concentration, and primer concentration (see, for example, SantaLucia J (1998), Proc Natl Acad Sci USA, 95:1460-5). In some embodiments an algorithm may use modified parameters (e.g., nearest-neighbor parameters for basepair entropy/enthalpy values). It should be appreciated that any suitable algorithm may be used as aspects of the invention are not limited in this respect. However, it also should be appreciated that different methodologies may results in different calculated or predicted Tms for the same sequences. Accordingly, in some embodiments, the same empirical and/or theoretical method is used to determine the Tms of different sequences for a set of probes to avoid a negative impact of any systematic difference in the Tm determination or prediction when designing a set of probes with predetermined similarities or differences for different Tms.

In some embodiments, the Tm of the first targeting arm may be about 1 °C, about 2 °C, about 3 °C, about 4 °C, about 5 °C, or more than about 5 °C higher than the Tm of the second

Petitioner Illumina, Inc., EX1004, p. 390

targeting arm. In some embodiments, each probe in a plurality of probes (e.g., each probe in a set of 5-10, each probe in a set of at least 10, each probe in a set of 10-50, each probe in a set of 50-100, each probe in a set of 100-500, each probe in a set of 500-1,000, each probe in a set of 1,000-1,500, each probe in a set of 1,500-2,000, each probe in a set of 2,000-3,000, 3,000-5,000, 5,000-10,000 or each probe in a set of at least 5,000 different probes) has a unique first targeting arm (e.g., they all have different sequences) and a unique second targeting arm (e.g., they all have different sequences). In some embodiments, for at least 10% of the probes (e.g., at least 25%, 25%-50%, 50%- 75%, 75%-90%, 90%-95% or over 95%, or all of the probes) the first targeting arm has a Tm for its complementary sequence that is higher (e.g., about 1 °C, about 2 °C, about 3 °C, about 4 °C, about 5 °C, or more than about 5 °C higher) than the Tm of the second targeting arm for its complementary sequence. In some embodiments, each of the first targeting arms have similar or identical Tms for their respective complementary sequences and each of the second targeting arms have similar or identical Tms for their respective complementary sequences (and the first targeting arms have higher Tms than the second targeting arms). For example, in some embodiments, the Tm of the first arm(s) may be about 58 °C and the Tm of the second arm(s) may be about 56 °C. In some embodiments, the Tm of the first arm(s) may be about 68 °C, and the Tm of the second arm(s) may be about 65 °C. It should be appreciated that in some embodiments the similarity (e.g., within a range of 1 °C, 2 °C, 3 °C, 4 °C, 5 °C) or identity of the Tms for the different targeting arms should be based either on empirical data for each arm or based on the same predictive algorithm for each arm (e.g., Wallace,R.B., Shaffer,J., Murphy,R.F., Bonner,J., Hirose,T., and Itakura,K. (1979) Nucleic Acids Res 6:3543-3557, SantaLucia J (1998), Proc Natl Acad Sci USA, 95:1460-5, or other algorithm).

In some embodiments, the Tm of the first targeting arm of a molecular inversion probe (at the 5' end of the molecular inversion probe) is selected to be sufficiently stable to prevent displacement of the first targeting arm from its complementary sequence on a target nucleic acid. In some embodiments, the Tm of the first targeting arm is 50-55 °C, at least 55 °C, 55-60 °C, at least 60 °C, 60-65 °C, at least 65 °C, at least 70 °C, at least 75 °C, or at least 80 °C. As discussed above, it should be appreciated that the for a particular targeting arm may be determined empirically or theoretically. Different theoretical models may be used to determine a Tm and it should be appreciated that the predicted Tm for a particular sequence may be different

Petitioner Illumina, Inc., EX1004, p. 391

depending on the algorithm used for the prediction. In some embodiments, each probe in a plurality of probes (e.g., each probe in a set of 5-10, each probe in a set of at least 10, each probe in a set of 10-50, each probe in a set of 50-100, each probe in a set of 100-500, or each probe in a set of at least 500 different probes) has a different first targeting arm (e.g., different sequences) but each different first targeting arm has a similar or identical Tm for its complementary sequence on a target nucleic acid. It should be appreciated that in some embodiments the similarity (e.g., within a range of I C, 2 C, 3 C, 4 C, 5 C) or identity of the Tms for the different targeting arms should be based either on empirical data for each arm or based on the same predictive algorithm for each arm (e.g., Wallace,R.B., Shaffer,J., Murphy,R.F., Bonner,J., Hirose,T., and Itakura,K. (1979) Nucleic Acids Res 6:3543- 3557, SantaLucia J (1998), Proc Natl Acad Sci USA, 95:1460-5, or other algorithm).

In some embodiments, the sub-target nucleic acid contains a nucleic acid repeat. In some embodiments, the nucleic acid repeat is a dinucleotide or trinucleotide repeat. In some embodiments, the sub-target nucleic acid contains 10-100 copies of the nucleic acid repeat in the absence of an abnormal increase or decrease in nucleic acid repeats. In some embodiments, the sub-target nucleic acid is a region of the Fragile-X locus that contains a nucleic acid repeat. In some embodiments, one or both targeting arms hybridize to a region on the target nucleic acid that is immediately adjacent to a region of nucleic acid repeats. In some embodiments, one or both targeting arms hybridize to a region on the target nucleic acid that is separated from a region of nucleic acid repeats by a region that does not contain any nucleic acid repeats. In some embodiments, the molecular inversion probe further comprises a primer-binding region that can be used to sequence the captured sub-target nucleic acid and optionally the first and/or second targeting arm.

In some embodiments, aspects of the invention relate to evaluating the length of a plurality of different target nucleic acids in a biological sample. In some embodiments, the plurality of target nucleic acids are analyzed using a plurality of different molecular inversion probes. In some embodiments, each different molecular inversion probe comprises a different pair of first and second targeting arms at each of the 3' and 5' ends. In some embodiments, each different molecular inversion probe comprises the same primer-binding sequence.

In some embodiments, aspects of the invention relate to analyzing nucleic acid from a biological sample obtained from a subject. In some embodiments, the biological sample is a

Petitioner Illumina, Inc., EX1004, p. 392

blood sample. In some embodiments, the biological sample is a tissue sample, specific cell population, tumor sample, circulating tumor cells, or environmental sample. In some embodiments, the biological sample is a single cell. In some embodiments, nucleic acids are analyzed in biological samples obtained from a plurality of different subjects. In some embodiments, nucleic acids from a biological sample are analyzed in multiplex reactions. It should be appreciated that a biological sample contains a plurality of copies of a genome derived from a plurality of cells in the sample. Accordingly, a sample may contain a plurality of independent copies of a target nucleic acid region of interest, the capture efficiency of which can be used to evaluate its size as described herein.

In some embodiments, aspects of the invention relate to evaluating a nucleic acid capture efficiency by determining an amount of target nucleic acid that is captured (e.g., an amount of sub-target nucleic acid sequences that are captured). In some embodiments, the amount of target nucleic acid that is captured is determined by determining a number of independently captured target nucleic acid molecules (e.g., the amount of independently captured molecules that have the sequence of the sub-target region). In some embodiments, the amount of target nucleic acid that is captured is compared to a reference amount of captured nucleic acid. In some embodiments, the reference amount is determined by determining a number of independently captured molecules of a reference nucleic acid. In some embodiments, the reference nucleic acid is a nucleic acid of a different locus in the biological sample that is not suspected of containing a deletion or insertion. In some embodiments, the reference nucleic acid is a nucleic acid of known size and amount that is added to the capture reaction. As described herein, a number of independently captured nucleic acid sequences can be determined by contacting a nucleic acid sample with a preparation of a probe (e.g., a MIP probe as described herein). It should be appreciated that the preparation may comprise a plurality of copies of the same probe and accordingly a plurality of independent copies of the target region may be captured by different probe molecules. The number of probe molecules that actually capture a sequence can be evaluated by determining an amount or number of captured molecules using any suitable technique. This number is a reflection of both the number of target molecules in the sample and the efficiency of capture of those target molecules, which in turn is related to the size of the target molecules as described herein. Accordingly, the capture efficiency can be evaluated by controlling for the abundance of the target nucleic acid, for example by comparing the number or

Petitioner Illumina, Inc., EX1004, p. 393

amount of captured target molecules to an appropriate control (e.g., a known size and amount of control nucleic acid, or a different locus that should be present in the same amount in the biological sample and is not expected to contain any insertions or deletions). It should be appreciated that other factors may affect the capture efficiency of a particular target nucleic acid region (e.g., the sequence of the region, the GC content, the presence of secondary structures, etc.). However, these factors also can be accounted for by using appropriate controls (e.g., known sequences having similar properties, the same sequences, other genomic sequences expected to be present in the biological sample at the same frequency, etc., or any combination thereof).

In some embodiments, aspects of the invention relate to identifying a subject as having an insertion or deletion in one or more alleles of a genetic locus if the capture efficiency for that genetic locus is statistically significantly different than a reference capture efficiency.

It should be appreciated that hybridization conditions used for any of the capture techniques described herein (e.g., MIP capture techniques) can be based on known hybridization buffers and conditions.

In some embodiments, the methods disclosed herein are useful for any application where the detection of deletions or insertions is important.

In some embodiments, aspects of the invention relate to basing a nucleic acid sequence analysis on results from two or more different nucleic acid preparatory techniques that have different systematic biases in the types of nucleic acids that they sample. According to the invention, different techniques have different sequence biases that are systematic and not simply due to stochastic effects during nucleic acid capture or amplification. Accordingly, the degree of oversampling required to overcome variations in nucleic acid preparation needs to be sufficient to overcome the biases (e.g., an oversampling of 2-5 fold, 5-10 fold, 5-15 fold, 15-20 fold, 20-30 fold, 30-50 fold, or intermediate to higher fold).

According to some embodiments, different techniques have different characteristic or systematic biases. For example, one technique may bias a sample analysis towards one particular allele at a genetic locus of interest, whereas a different technique would bias the sample analysis towards a different allele at the same locus. Accordingly, the same sample may be identified as being different depending on the type of technique that is used to prepare nucleic

Petitioner Illumina, Inc., EX1004, p. 394

acid for sequence analysis. This effectively represents a sensitivity limitation, because each technique has different relative sensitivities for polymorphic sequences of interest.

According to aspects of the invention, the sensitivity of a nucleic acid analysis can be increased by combining the sequences from different nucleic acid preparative steps and using the combined sequence information for a diagnostic assay (e.g., for a making a call as to whether a subject is homozygous or heterozygous at a genetic locus of interest).

In some embodiments, the invention provides a method of increasing the sensitivity of a nucleic acid detection assay by obtaining a first preparation of a target nucleic acid using a first preparative method on a biological sample, obtaining a second preparation of a target nucleic acid using a second preparative method on the biological sample, assaying the sequences obtained in both first and second nucleic acid preparations, and using the sequence information from both first and second nucleic acid preparations to determine the genotype of the target nucleic acid in the biological sample, wherein the first and second preparative methods have different systematic sequence biases. In some embodiments, the first and second nucleic acid preparations are combined prior to performing a sequence assay. In some embodiments, separate sequence assays are performed on the first and second nucleic acid preparations and the sequence information from both assays are combined to determine the genotype of the target nucleic acid in the biological sample. In some embodiments, the first preparative method is an amplification-based, a hybridization-based, or a circular probe-based preparative method. In some embodiments, the second method is an amplification-based, a hybridization-based, or a circular probe-based preparative method. In some embodiments, the first and second methods are of different types (e.g., only one of them is an amplification-based, a hybridization-based, or a circular probe-based preparative method, and the other one is one or the other two types of method). Accordingly, in some embodiments the second preparative method is an amplification-based, a hybridization-based, or a circular probe-based preparative method, provided that the second method is different from the first method. However, in some embodiments, both methods may be of the same type, provided they are different methods (e.g., both are amplification based or hybridization-based, but are different types of amplification or hybridization methods, e.g., with different relative biases).

In amplification-based (e.g., PCR-based or LCR-based, etc.) preparative methods, genomic loci (target nucleic acids) are isolated directly by means of a polymerase chain reaction

Petitioner Illumina, Inc., EX1004, p. 395

or ligase chain reaction (or other amplification method) that selectively amplifies each locus using a pair of oligonucleotide primers. It is to be understood that primers will be sufficiently complementary to the target sequence to hybridize with and prime amplification of the target nucleic acid. Any one of a variety of art known methods may be utilized for primer design and synthesis. One or both of the primers may be perfectly complementary to the target sequence. Degenerate primers may also be used. Primers may also include additional nucleic acids that are not complementary to target sequences but that facilitate downstream applications, including for example restriction sites and identifier sequences (e.g., source sequences). PCR based methods may include amplification of a single target nucleic acid and multiplex amplification (amplification of multiple target nucleic acids in parallel).

Hybridization-based preparative may methods involve selectively immobilizing target nucleic acids for further manipulation. It is to be understood that one or more oligonucleotides (immobilization oligonucleotides), which in some embodiments may be from 10 to 200 nucleotides in length, are used which hybridize along the length of a target region of a genetic locus to immobilize it. In some embodiments, immobilization oligonucleotides are either immobilized before hybridization is performed (e.g., Roche/Nimblegen 'sequence capture'), or are prepared such that they include a moiety (e.g., biotin) which can be used to selectively immobilize the target nucleic acid after hybridization by binding to e.g., streptavidin-coated microbeads (e.g., Agilent 'SureSelect').

Circularization selection-based preparative methods selectively convert each region of interest into a covalently-closed circular molecule which is then isolated by removal (usually enzymatic, e.g., with exonuclease) of any non-circularized linear nucleic acid. Oligonucleotide probes are designed which have ends that flank the region of interest. The probes are allowed to hybridize to the genomic target, and enzymes are used to first (optionally) fill in any gap between probe ends and second ligate the probe closed. In some embodiments, following circularization, any remaining (non-target) linear nucleic acid can be removed, resulting in isolation (capture) of target nucleic acid. Circularization selection-based preparative methods include molecular inversion probe capture reactions and 'selector' capture reactions. However, other techniques may be used as aspects of the invention are not limited in this respect. In some embodiments, molecular inversion probe capture of a target nucleic acid is indicative of the presence of a polymorphism in the target nucleic acid.

Petitioner Illumina, Inc., EX1004, p. 396

A variety of methods may be used to evaluate and compare bias profiles of each preparative technique. Next-generation sequencing may be used to quantitatively measure the abundance of each isolated target nucleic acid obtained from a certain preparative method. This abundance may be compared to a control abundance value (e.g., a known starting abundance of the target nucleic acid) and/or with an abundance determined through the use of an alternative preparative method. For example, a set of target nucleic acids may be isolated by one or more of the three preparative methods; the target nucleic acid may be observed x times using the amplification technique, y times using the hybridization enrichment technique, and z times using the circularization selection technique. A pairwise correlation coefficient may be computed between each abundance value (e.g., x and y, x and z, and y and z) to assess bias in nucleic acid isolation between pairs of preparative methods. Since the mechanisms of isolation are different in each approach, the abundances will usually be different and largely uncorrelated with each other.

In some embodiments, the invention provides a method of obtaining a nucleic acid preparation that is representative of a target nucleic acid in a biological sample by obtaining a first preparation of a target nucleic acid using a first preparative method on a biological sample, obtaining a second preparation of a target nucleic acid using a second preparative method on the biological sample, and combining the first and second nucleic acid preparations to obtain a combined preparation that is representative of the target nucleic acid in the biological sample.

In some embodiments of any of the methods described herein, a third preparation of the target nucleic acid is obtained using a third preparative method that is different from the first and second preparative methods, wherein the first, second, and third preparative methods all have different systematic sequence biases. In some embodiments of any of the methods described herein, the different preparative methods are used for a plurality of different loci in the biological sample to increase the sensitivity of a multiplex nucleic acid analysis. In some embodiments, the target nucleic acid has a sequence of a gene selected from Table 1.

However, it should be appreciated that a genotyping method of the invention may include several steps, each of which independently may involve one or more different preparative techniques described herein. In some embodiments, a nucleic acid preparation may be obtained using one or more (e.g., 2, 3, 4, 5, or more) different techniques described herein (e.g., amplification, hybridization capture, circular probe capture, etc., or any combination thereof) and

Petitioner Illumina, Inc., EX1004, p. 397

the nucleic acid preparation may be analyzed using one or more different techniques (e.g., amplification, hybridization capture, circular probe capture, etc., or any combination thereof) that are selected independently of the techniques used for the initial preparation.

In some embodiments, aspects of the invention also provide compositions, kits, devices, and analytical methods for increasing the sensitivity of nucleic acid assays. Aspects of the invention are particularly useful for increasing the confidence level of genotyping analyses. However, aspects of the invention may be used in the context of any suitable nucleic acid analysis, for example, but not limited to, a nucleic acid analysis that is designed to determine whether more than one sequence variant is present in a sample.

In some embodiments, aspects of the invention relate to a plurality of nucleic acid probes (e.g., 10-50, 50-100, 100-250, 250-500, 500-1,000, 1,000-2,000, 2,000-5,000, 5,000-7,500, 7,500-10,000, or lower, higher, or intermediate number of different probes).

In some embodiments, each probe or each of a subset of probes (e.g., 10-25%, 25-50%, 50-75%, 75-90%, or 90-99%) has a different first targeting arm. In some embodiments, each probe or each probe of a subset of probes (e.g., 10-25%, 25-50%, 50-75%, 75-90%, or 90-99%) has a different second targeting arm. In some embodiments, the first and second targeting arms are separated by the same intervening sequence. In some embodiments, the first and second targeting arms are complementary to target nucleic acid sequences that are separated by the same or a similar length (e.g., number of nucleic acids, for example, 0-25, 25-50, 50-100, 100-250, 250-500, 500-1,000, 1,000-2,500 or longer or intermediate number of nucleotides) on their respective target nucleic acids (e.g., genomic loci). In some embodiments, each probe or a subset of probes (e.g., 10-30 25%, 25-50%, 50-75%, 75-90%, or 90-99%) includes a first primer binding sequence. In some embodiments, the primer binding sequence is the same (e.g., it can be used to prime sequencing or other extension reaction). In some embodiments, each probe or a subset of probes (e.g., 10-25%, 25-50%, 50-75%, 75-90%, or 90-99%) includes a unique identifier sequence tag (e.g., that is predetermined and can be used to distinguish each probe).

In some embodiments, the methods disclosed herein are useful for any application where sensitivity is important. For example, detection of cancer mutations in a heterogenous tissue sample, detection of mutations in maternally-circulating fetal DNA, and detection of mutations in cells isolated during a preimplantation genetic diagnostic procedure.

Petitioner Illumina, Inc., EX1004, p. 398

According to some aspects of the invention, methods of detecting a polymorphism in a nucleic acid in a biological sample are provided. In some embodiments, the methods comprise obtaining a nucleic acid preparation using a preparative method (e.g., any of the preparative methods disclosed herein) on a biological sample, and performing a molecular inversion probe capture reaction on the nucleic acid preparation, wherein a molecular inversion probe capture (e.g., using a mutation-detection MIP) of a target nucleic acid of the nucleic acid preparation is indicative of the presence of a mutation (polymorphism) in the target nucleic acid, optionally wherein the polymorphism is selected from Table 2.

According to some aspects of the invention, methods of genotyping a nucleic acid in a biological sample are provided. In some embodiments, the methods comprise obtaining a nucleic acid preparation using a preparative method on a biological sample, sequencing a target nucleic acid of the nucleic acid preparation, and performing a molecular inversion probe capture reaction on the biological sample, wherein a molecular inversion probe capture of the target nucleic acid in the biological sample is indicative of the presence of a polymorphism in the target nucleic acid, genotyping the target nucleic acid based on the results of the sequencing and the capture reaction.

In some embodiments of the methods disclosed herein, the target nucleic acid has a sequence of a gene selected from Table 1.

It should be appreciated that any one or more embodiments described herein may be used for evaluating multiple genetic markers in parallel. Accordingly, in some embodiments, aspects of the invention relate to determining the presence of one or more markers (e.g., one or more alleles) at multiple different genetic loci in parallel.

Accordingly, the risk or presence of multiple heritable disorders may be evaluated in parallel. In some embodiments, the risk of having offspring with one or more heritable disorders may be evaluated. In some embodiments, an evaluation may be performed on a biological sample of a parent or a child (e.g., at a pre-implantation, prenatal, perinatal, or postnatal stage). In some embodiments, the disclosure provides methods for analyzing multiple genetic loci (e.g., a plurality of target nucleic acids selected from Table 1 or 2) from a patient sample, such as a blood, pre-implantation embryo, chorionic villus or amniotic fluid sample. A patient or subject may be a human. However, aspects of the invention are not limited to humans and may be applied to other species (e.g., mammals, birds, reptiles, other vertebrates or invertebrates) as

Petitioner Illumina, Inc., EX1004, p. 399

aspects of the invention are not limited in this respect. A subject or patient may be male or female. In some embodiments, in connection with reproductive genetic counseling, samples from a male and female member of a couple may be analyzed. In some embodiments, for example, in connection with an animal breeding program, samples from a plurality of male and female subjects may be analyzed to determine compatible or optimal breeding partners or strategies for particular traits or to avoid one or more diseases or conditions. Accordingly, reproductive risks may be determined and/or reproductive recommendations may be provided based on information derived from one or more embodiments of the invention.

However, it should be appreciated that aspects of the invention may be used in connection with any medical evaluation where the presence of one or more alleles at a genetic locus of interest is relevant to a medical determination (e.g., risk or detection of disease, disease prognosis, therapy selection, therapy monitoring, etc.). Further aspects of the invention may be used in connection with detection, in tumor tissue or circulating tumor cells, of mutations in cellular pathways that cause cancer or predict efficacy of treatment regimens, or with detection and identification of pathogenic organisms in the environment or a sample obtained from a subject, e.g., a human subject.

These and other aspects of the invention are described in more detail in the following description and non-limiting examples and drawings.

BRIEF DESCRIPTION OF DRAWINGS

FIG. 1 illustrates a non-limiting embodiment of a tiled probe layout;

FIG. 2 illustrates a non-limiting embodiment of a staggered probe layout;

FIG. 3 illustrates a non-limiting embodiment of an alternating staggered probe layout;

~~FIGs. 4A, B, and C~~ FIGS. 4A-C depict various non-limiting methods for combining differentiator tag sequence and target sequences (NNNN depicts a differentiator tag sequence);

FIG. 5 depicts a non-limiting method for genotyping based on target and 5 differentiator tag sequences;

FIG. 6 depicts non-limiting results of a simulation of a MIP capture reaction;

FIG. 7 depicts a non-limiting graph of sequencing coverage;

FIG. 8 illustrates that shorter sequences are captured with higher efficiency that longer sequences using MIPs;

Petitioner Illumina, Inc., EX1004, p. 400

FIG. 9 illustrates a non-limiting scheme of padlock (MIP) capture of a region that includes both repetitive regions (thick wavy line) and the adjacent unique sequence (thick strait line);

FIG. 10 illustrates a non-limiting hypothetical relationship between target gap size and the relative number of reads of the repetitive region;

FIG. 11A depicts MIP capture of FMR1 repeat regions from a diploid genome;

FIG. 11B depicts preparative methods for biallelic resolution of FMR I repeat region lengths in a diploid genome using MIP capture probes and unique differentiator tags;

FIG. 11C depicts an analysis of FMR1 repeat region lengths in a diploid genome;

FIG. 12 is a schematic of an embodiment of an algorithm of the invention;

FIG. 13 illustrates a non-limiting example of a graph of per-target abundance with MIP capture; and,

FIG. 14 shows a non-limiting a graph of correlation between two MIP capture reactions.

DETAILED DESCRIPTION

Aspects of the invention relate to preparative and analytical methods and compositions for evaluating genotypes, and in particular, for determining the allelic identity (or identities in a diploid organism) of one or more genetic loci in a subject.

Aspects of the invention are based, in part, on the identification of different sources of ambiguity and error in genetic analyses, and, in part, on the identification of one or more approaches to avoid, reduce, recognize, and/or resolve these errors and ambiguities at different stages in a genetic analysis. Aspects of the invention relate to methods and compositions for addressing bias and/or stochastic variation associated with one or more preparative and/or analytical steps of a nucleic acid evaluation technology. In some embodiments, preparative methods can be adapted to avoid or reduce the risk of bias skewing the results of a genetic analysis. In some embodiments, analytical methods can be adapted to recognize and correct for data variations that may give rise to misinterpretation (e.g., incorrect calls such as homozygous when the subject is actually heterozygous or heterozygous when the subject is actually homozygous). Methods of the invention may be used for any type of mutation, for example a single base change (e.g., insertion, deletion, transversion or transition, etc.), a multiple base insertion, deletion, duplication, inversion, and/or any other change or combination thereof.

Petitioner Illumina, Inc., EX1004, p. 401

In some embodiments, additional or alternative techniques may be used to address loci characterized by multiple repeats of a core sequence where the length of the repeat is longer than a typical sequencing read thereby making it difficult to determine whether a deletion or duplication of one or more core sequence units has occurred based solely on a sequence read.

In some embodiments, increased confidence in an assay result may be obtained by i) selecting two or more different preparative and/or analytical techniques that have different biases (e.g., known to have different biases), ii) evaluating a patient sample using the two or more different techniques, iii) comparing the results from the two or more different techniques, and/or iv) determining whether the results are consistent for the two or more different techniques. In some embodiments, if determining in step (iv) indicates that the results are consistent (e.g., the same) then increased confidence in the assay result is obtained. In other embodiments, if determining in step (iv) indicates that the results are inconsistent (e.g., that the results are ambiguous) then one or more additional preparative and/or analytical techniques, which have a different bias (e.g., known to have a different bias) compared with the two or more different preparative and/or analytical techniques selected in step (i), are used to evaluate the patient sample, and the results of the one or more additional preparative and/or analytical techniques are compared with the results from step (ii) to resolve the inconsistency.

In some embodiments, two or more independent samples may be obtained from a subject and independently analyzed. In some embodiments, two or more independent samples are obtained at approximately the same time point. In some embodiments, two or more independent samples are obtained at multiple different time points. In some embodiments, the use of two or more independent sample facilitates the elimination, normalization, and/or quantification of stochastic measurement noise. It is to be appreciated that two or more independent samples may be obtained in connection with any of the methods disclosed herein, including, for example, methods for pathogen profiling in a human or other animal subjects, monitoring tumor progression/regression, analyzing circulating tumor cells, analyzing fetal cells in maternal circulation, and analyzing/monitoring/profiling of environmental pathogens.

In some embodiments, one or more of the techniques described herein may be combined in a single assay protocol for evaluating multiple patient samples in parallel.

It should be appreciated that aspects of the invention may be useful for high throughput, cost-effective, yet reliable, genotyping of multiple patient samples (e.g., in parallel, for example

Petitioner Illumina, Inc., EX1004, p. 402

in multiplex reactions). In some embodiments, aspects of the invention are useful to reduce the error frequency in a multiplex analysis. Certain embodiments may be particularly useful where multiple reactions (e.g., multiple loci and/or multiple patient samples) are being processed. For example, 10-25, 25-50, 50-75, 75-100 or more loci may be evaluated for each subject out of any number of subject samples that may be processed in parallel (e.g., 1-25, 25-50, 50-100, 100-500, 500-1,000, 1,000-2,500, 2,500-5,000 or more or intermediate numbers of patient samples). It should be appreciated that different embodiments of the invention may involve conducting two or more target capture reactions and/or two or more patient sample analyses in parallel in a single multiplex reaction. For example, in some embodiments a plurality of capture reactions (e.g., using different capture probes for different target loci) may be performed in a single multiplex reaction on a single patient sample. In some embodiments, a plurality of captured nucleic acids from each one of a plurality of patient samples may be combined in a single multiplex analysis reaction. In some embodiments, samples from different subjects are tagged with subject-specific (e.g., patient-specific) tags (e.g., unique sequence tags) so that the information from each product can be assigned to an identified subject. In some embodiments, each of the different capture probes used for each patient sample have a common patient-specific tag. In some embodiments, the capture probes do not have patient-specific tags, but the captured products from each subject may be amplified using one or a pair of amplification primers that are labeled with a patient-specific tag. Other techniques for associating a patient-specific tag with the captured product from a single patient sample may be used as aspects of the invention are not limited in this respect. It should be appreciated that patient-specific tags as used herein may refer to unique tags that are assigned to identified patients in a particular assay. The same tags may be used in a separate multiplex analysis with a different set of patient samples (e.g., from different patients) each of which is assigned one of the tags. In some embodiments, different sets of unique tags may be used in sequential (e.g., alternating) multiplex reactions in order to reduce the risk of contamination from one assay to the next and allow contamination to be detected on the basis of the presence of tags that are not expected to be present in a particular assay.

Embodiments of the invention may be used for any of a number of different settings: reproductive settings, disease screening, identifying subjects having cancer, identifying subjects having increased risk for a disease, stratifying a population of subjects according to one or more of a number of factors, for example responsiveness to a particular drug, lack or not of an adverse

Petitioner Illumina, Inc., EX1004, p. 403

reaction (or risk therefore) to a particular drug, and/or providing information for medical records (e.g., homozygosity, heterozygosity at one or more loci). It should be appreciated that the invention is not limited to genomic analysis of patient samples. For example, aspects of the invention may be useful for high throughput genetic analysis of environment samples to detect pathogens.

In some embodiments, the methods disclosed herein are useful for diagnosis of one or more heritable disorders. In some embodiments, a heritable disorder that may be diagnosed with the methods disclosed herein is a genetic disorder that is prevalent in the Ashkenazi Jewish population. In some embodiments, the heritable disorders are selected from: 21-Hydroxylase-Defiocient Congenital Adrenal Hyperplasia; ABCC8-Related Hyperinsulinism; Alpha-Thalassemia, includes Constant Spring, & MR associated; Arylsulfatase A Deficiency-Metyachromatic Leukodystrophy; Biotinidase Deficiency Holocarboxylase Synthetase Deficiency; Bloom's Syndrome; Canavan Disease; CFTR-Related Disorders-cystic fibrosis; Citrullinemia Type 1; Combined MMA & Homocystinuria-dblC; Dystrophinopathies (DMD & BMD); Familial Dysautonomia; Fanconi Anemia-FANCC; Galactosemia-Classical: Galactokinase Defiency & Galactose Epimerase Deficiency; Gaucher Disease; GJB2-Related DFNB 1 Nonsyndromic Hearing Loss and Deafness; Glutaric acidemia Type 1; Hemoglobinopathies beta-chain disorders; Glycogen Storage Disease Type 1A; Maple Syrup Urine Disease; Types 1A, 1B, 2, 3; Medium Chain Acyl-Coenzyme A; Dehydrogenase Deficiency-MCADD; Methylmalonic Acidemia; Mucolipidosis IV; Nemaline Myopathy; Nieman-Pick Type A-Acid Sphingomyelinase Deficiency; Non-Ketotic Hyperglycinemia-Glycine Encephalopathy; Ornithine Transcarbamylase Deficiency; PKU Phenylalanine Hydroxylase Deficiency; Propionic Acidemia; Short Chain Acyl-CoA Dehydrogenase Deficiency-SCADD; Smith-Lemli-Opitz Syndrome; Spinal Muscular Atrophy (SMNI)-SMA; Tay Sachs-HexA Deficiency; Usher Synbdrome-Type I (Type IB, Type IC, Type ID, Type IF, Type IG) ; X-Linked Mental Retardation ARX-Related Disorders; X-Linked Mental Retardation with Cerebellar Cypoplasia and sistinctive Facial Appearance; X-Linked Mental Retardation; includes 9, 21, 30, 46, 58, 63, 88, 89; X linked mental retardation: FM1-Related Disorders-FRXA, Fragile X MR; X-linked SMR: Renpenning Syndrome 1; Zellweger Spectrum disorders - Peroxisomal Bifunctional Enzyme Deficiencies including Zellweger, NALD, and/or infantile

Petitioner Illumina, Inc., EX1004, p. 404

Refsums. However, all of these, subsets of these, other genes, or combinations thereof may be used.

According to some aspects, the disclosure relates to multiplex diagnostic methods. In some embodiments, multiplex diagnostic methods comprise capturing a plurality of genetic loci in parallel (e.g., a genetic locus of Table 1). In some embodiments, genetic loci possess one or more polymorphisms (e.g., a polymorphism of Table 2) the genotypes of which correspond to disease causing alleles. Accordingly, in some embodiments, the disclosure provides methods for assessing multiple heritable disorders in parallel.

In some embodiments, methods are provided for diagnosing multiple heritable disorders in parallel at a pre-implantation, prenatal, perinatal, or postnatal stage. In some embodiments, the disclosure provides methods for analyzing multiple genetic loci (e.g., a plurality of target nucleic acids selected from Table 1) from a patient sample, such as a blood, pre-implantation embryo, chorionic villus or amniotic fluid sample. A patient or subject may be a human. However, aspects of the invention are not limited to humans and may be applied to other species (e.g., mammals, birds, reptiles, other vertebrates or invertebrates) as aspects of the invention are not limited in this respect. A subject or patient may be male or female. In some embodiments, in connection with reproductive genetic counseling, samples from a male and female member of a couple may be analyzed. In some embodiments, for example, in connection with an animal breeding program, samples from a plurality of male and female subjects may be analyzed to determine compatible or optimal breeding partners or strategies for particular traits or to avoid one or more diseases or conditions.

However, it should be appreciated that any other diseases may be studied and/or risk factors for diseases or disorders including, but not limited to allergies, responsiveness to treatment, cancer tumor profiling for treatment and prognosis, monitoring and identification of patient infections, and monitoring of environmental pathogens.

1. Reducing representational bias in multiplex amplification reactions:

In some embodiments, aspects of the invention relate to methods that reduce bias and increase reproducibility in multiplex detection of genetic loci, e.g., for diagnostic purposes.

Molecular inversion probe technology is used to detect or amplify particular nucleic acid sequences in potentially complex mixtures. Use of molecular inversion probes has been demonstrated for detection of single nucleotide polymorphisms (Hardenbol et al. 2005 Genome

Petitioner Illumina, Inc., EX1004, p. 405

Res 15:269-75) and for preparative amplification of large sets of exons (Porreca et al. 2007 Nat Methods 4:931-6, Krishnakumar et al. 2008 Proc Natl Acad Sci USA 105:9296-301). One of the main benefits of the method is in its capacity for a high degree of multiplexing, because generally thousands of targets may be captured in a single reaction containing thousands of probes. However, challenges associated with, for example, amplification efficiency (See, e.g., Turner EH, et al., Nat Methods. 2009 Apr 6:1-2.) have limited the practical utility of the method in research and diagnostic settings.

Aspects of the disclosure are based, in part, on the discovery of effective methods for overcoming challenges associated with systematic errors (bias) in multiplex genomic capture and sequencing methods, namely high variability in target nucleic acid representation and unequal sampling of heterozygous alleles in pools of captured target nucleic acids (e.g., isolated from a biological sample). Accordingly, in some embodiments, the disclosure provides methods that reduce variability in the detection of target nucleic acids in multiplex capture methods. In other embodiments, methods improve allelic representation in a capture pool and, thus, improve variant detection outcomes. In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve the use of different sets of multiple probes (e.g., molecular inversion probes MIPs) that capture overlapping regions of a target nucleic acid to achieve a more uniform representation of the target nucleic acids in a capture pool compared with methods of the prior art. In other embodiments, methods reduce bias, or the risk of bias, associated with large scale parallel capture of genetic loci, e.g., for diagnostic purposes. In other embodiments, methods are provided for increasing reproducibility (e.g., by reducing the effect of polymorphisms on target nucleic acid capture) in the detection of a plurality of genetic loci in parallel. In further embodiments, methods are provided for reducing the effect of probe synthesis and/or probe amplification variability on the analysis of a plurality of genetic loci in parallel.

In some aspects, the disclosure provides probe sets that comprise a plurality of different probes. As used herein, a 'probe' is a nucleic acid having a central region flanked by a 5' region and a 3' region that are complementary to nucleic acids flanking the same strand of a target nucleic acid or subregion thereof. An exemplary probe is a molecular inversion probe (MIP). A 'target nucleic acid' may be a genetic locus.

Exemplary genetic loci are disclosed herein in Table 1 (RefSeqGene Column).

Petitioner Illumina, Inc., EX1004, p. 406

While probes have been typically designed to meet certain constraints (e.g. melting temperature, G/C content, etc.) known to partially affect capture/amplification efficiency (Ball et al (2009) Nat Biotech 27:361-8 AND Deng et al (2009) Nat Biotech 27:353-60), a set of constraints which is sufficient to ensure either largely uniform or highly reproducible capture/amplification efficiency has not previously been achieved.

As disclosed herein, uniformity and reproducibility can be increased by designing multiple probes per target, such that each base in the target is captured by more than one probe. In some embodiments, the disclosure provides multiple MIPs per target to be captured, where each MIP in a set designed for a given target nucleic acid has a central region and a 5' region and 3' region ('targeting arms') which hybridize to (at least partially) different nucleic acids in the target nucleic acid (immediately flanking a subregion of the target nucleic acid). Thus, differences in efficiency between different targeting arms and fill-in sequences may be averaged across multiple MIPs for a single target, which results in more uniform and reproducible capture efficiency.

In some embodiments, the methods involve designing a single probe for each target (a target can be as small as a single base or as large as a kilobase or more of contiguous sequence).

It may be preferable, in some cases, to design probes to capture molecules (e.g., target nucleic acids or subregions thereof) having lengths in the range of 1 - 200 bp (as used herein, a bp refers to a base pair on a double-stranded nucleic acid - however, where lengths are indicated in bps, it should be appreciated that single-stranded nucleic acids having the same number of bases, as opposed to base pairs, in length also are contemplated by the invention). However, probe design is not so limited. For example, probes can be designed to capture targets having lengths in the range of up to 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 1000, or more bps, in some cases.

It is to be appreciated that the length of a capture molecule (e.g., a target nucleic acid or subregion thereof) is selected based upon multiple considerations. For example, where analysis of a target involves sequencing, e.g., with a next-generation sequencer, the target length should typically match the sequencing read-length so that shotgun library construction is not necessary. However, it should be appreciated that captured nucleic acids may be sequenced using any suitable sequencing technique as aspects of the invention are not limited in this respect.

Petitioner Illumina, Inc., EX1004, p. 407

It is also to be appreciated that some target nucleic acids are too large to be captured with one probe. Consequently, it may be necessary to capture multiple subregions of a target nucleic acid in order to analyze the full target.

In some embodiments, a subregion of a target nucleic acid is at least 1 bp. In other embodiments, a subregion of a target nucleic acid is at least 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 bp or more. In other embodiments, a subregion of a target nucleic acid has a length that is up to 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or more percent of a target nucleic acid length.

The skilled artisan will also appreciate that consideration is made, in the design of MIPs, for the relationship between probe length and target length. In some embodiments, MIPs are designed such that they are several hundred basepairs (e.g., up to 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 bp or more) longer than corresponding target (e.g., subregion of a target nucleic acid, target nucleic acid).

In some embodiments, lengths of subregions of a target nucleic acid may differ.

For example, if a target nucleic acid contains regions for which probe hybridization is not possible or inefficient, it may be necessary to use probes that capture subregions of one or more different lengths in order to avoid hybridization with problematic nucleic acids and capture nucleic acids that encompass a complete target nucleic acid.

Aspects of the invention involve using multiple probes, e.g., MIPs, to amplify each target nucleic acid. In some embodiments, the set of probes for a given target can be designed to 'tile' across the target, capturing the target as a series of shorter sub targets. In some embodiments, where a set of probes for a given target is designed to 'tile' across the target, some probes in the set capture flanking non-target sequence). Alternately, the set can be designed to 'stagger' the exact positions of the hybridization regions flanking the target, capturing the full target (and in some cases capturing flanking non-target sequence) with multiple probes having different targeting arms, obviating the need for tiling. The particular approach chosen will depend on the nature of the target set. For example, if small regions are to be captured, a staggered-end approach might be appropriate, whereas if longer regions are desired, tiling might be chosen. In all cases, the amount of bias-tolerance for probes targeting pathological loci can be adjusted ('dialed in') by changing the number of different MIPs used to capture a given molecule.

Petitioner Illumina, Inc., EX1004, p. 408

In some embodiments, the 'coverage factor', or number of probes used to capture a basepair in a molecule, is an important parameter to specify. Different numbers of probes per target are indicated depending on whether one is using the tiling approach (see, e.g., FIG. 1) or one of the staggered approaches (see, e.g., FIGs. 2 or 3).

FIG. 1 illustrates a non-limiting embodiment of a tiled probe layout showing ten captured sub-targets tiled across a single target. Each position in the target is covered by three sub-targets such that MIP performance per base pair is averaged across three probes.

FIG. 2 illustrates a non-limiting embodiment of a staggered probe layout showing the targets captured by a set of three MIPs. Each MIP captures the full target, shown in black, plus (in some cases) additional extra-target sequence, shown in gray, such that the targeting arms of each MIP fall on different sequence. Each position in the target is covered by three sub-targets such that MIP performance per basepair is averaged across three probes. Targeting arms land immediately adjacent to the black or gray regions shown. It should be appreciated that in some embodiments, the targeting arms (not shown) can be designed so that they do not overlap with each other.

FIG. 3 illustrates a non-limiting embodiment of an alternating staggered probe layout showing the targets captured by a set of three MIPs. Each MIP captures the full target, shown in black, plus (in some cases) additional extra-target sequence, shown in gray, such that the targeting arms of each MIP fall on different sequence. Each position in the target is covered by three sub-targets such that MIP performance per basepair is averaged across three probes. Targeting arms land immediately adjacent to the black or gray regions shown.

It should be appreciated that for any of the layouts, the targeting arms on adjacent tiled or staggered probes may be designed to either overlap, not overlap, or overlap for only a subset of the probes.

In certain embodiments for any of the layouts, a coverage factor of about 3 to about 10 is used. However, the methods are not so limited and coverage factors of up to 2, 3, 4, 5, 6, 7, 8, 9, 10, 20 or more may be used. It is to be appreciated that the coverage factor selected may depend the probe layout being employed. For example, in the tiling approach, for a desired coverage factor, the number of probes per target is typically a function of target length, sub-target length, and spacing between adjacent sub-target start locations (step size). For example, for a desired coverage factor of 3, a 200 bp target with a start-site separation of 20 bp and sub-target length of

Petitioner Illumina, Inc., EX1004, p. 409

60 bp may be encompassed with 12 MIPs (FIG. 1). Thus, a specific coverage factor may be achieved by varying the number of probes per target nucleic acid and the length of the molecules captured. In the staggered approach, a fixed-length target nucleic acid is captured as several subregions or as 'super-targets', which are molecules comprising the target nucleic acid and additional flanking nucleic acids, which may be of varying lengths. For example, a target of 50 bp can be captured at a coverage factor of 3 with 3 probes in either a 'staggered' (FIG. 2) or 'alternating staggered' configuration (FIG. 3).

The coverage factor will be driven by the extent to which detection bias is tolerable. In some cases, where the bias tolerance is small, it may be desirable to target more subregions of target nucleic acid with, perhaps, higher coverage factors. In some embodiments, the coverage factor is up to 2, 3, 4, 5, 6, 7, 8, 9, 10 or more.

In some embodiments, when a tiled probe layout is used, when the target length is greater than 1 bp and when a step size (distance between the 5'-end of a target and the 5' end of its adjacent target) is less than the length of a target or subregion thereof, it is possible to compute probe number for a particular target based on target length (T), sub target length (S), and coverage factor (C), such that probe number = T/(S/C) + (C-1).

In some aspects, the disclosure provides methods to increase the uniformity of amplification efficiency when multiple molecules are amplified in parallel; methods to increase the reproducibility of amplification efficiency; methods to reduce the contribution of targeting probe variability to amplification efficiency; methods to reduce the effect on a given target nucleic acid of polymorphisms in probe hybridization regions; and/or methods to simplify downstream workflows when multiplex amplification by MIPs is used as a preparative step for analysis by nucleic acid sequencing.

Polymorphisms in the target nucleic acid under the regions flanking a target can interfere with hybridization, polymerase fill-in, and/or ligation. Furthermore, this may occur for only one allele, resulting in allelic drop-out, which ultimately decreases downstream sequencing accuracy. In some embodiments, using a set of MIPs having multiple hybridization sites for the capture of any given target, the probability of loss from polymorphism is substantially decreased because not all targeting arms in the set of MIPs will cover the location of the mutation.

Probes for MIP capture reactions may be synthesized on programmable microarrays because of the large number of sequences required. Because of the low synthesis yields of these

Petitioner Illumina, Inc., EX1004, p. 410

methods, a subsequent amplification step is required to produce sufficient probe for the MIP amplification reaction. The combination of multiplex oligonucleotide synthesis and pooled amplification results in uneven synthesis error rates and representational biases. By synthesizing multiple probes for each target, variation from these sources may be averaged out because not all probes for a given target will have the same error rates and biases.

Multiplex amplification strategies disclosed herein may be used analytically, as in detection of SNPs, or preparatively, often for next-generation sequencing or other sequencing techniques. In the preparative setting, the output of an amplification reaction is generally the input to a shotgun library protocol, which then becomes the input to the sequencing platform. The shotgun library is necessary in part because next-generation sequencing yields reads significantly shorter than amplicons such as exons. In addition to the bias-reduction afforded by the multi-tiled approach described here, tiling also obviates the need for shotgun library preparation. Since the length of the capture molecule can be specified when the probes, e.g., MIPs, are designed, it can be chosen to match the readlength of the sequencer. In this way, reads can 'walk' across an exon by virtue of the start position of each capture molecule in the probe set for that exon. Reducing analytical errors associated with bias in nucleic acid preparations:

In some embodiments, aspects of the invention relate to preparative steps in DNA sequencing-related technologies that reduce bias and increase the reliability and accuracy of downstream quantitative applications.

There are currently many genomics assays that utilize next-generation (polony-based) sequencing to generate data, including genome resequencing, RNA-seq for gene expression, bisulphite sequencing for methylation, and Immune-seq, among others. In order to make quantitative measurements (including genotype calling), these methods utilize the counts of sequencing reads of a given genomic locus as a proxy for the representation of that sequence in the original sample of nucleic acids. The majority of these techniques require a preparative step to construct a high-complexity library of DNA molecules that is representative of a sample of interest. This may include chemical or biochemical treatment of the DNA (e.g., bisulphite treatment), capture of a specific subset of the genome (e.g., padlock probe capture, solution hybridization), and a variety of amplification techniques (e.g., polymerase chain reaction, whole genome amplification, rolling circle amplification).

Petitioner Illumina, Inc., EX1004, p. 411

Systematic and random errors are common problems associated with genome amplification and sequencing library construction techniques. For example, genomic sequencing library may contain an over- or under-representation of particular sequences from a source genome as a result of errors (bias) in the library construction process. Such bias can be particularly problematic when it results in target sequences from a genome being absent or undetectable in the sequencing libraries. For example, an under representation of particular allelic sequences (e.g., heterozygotic alleles) from a genome in a sequencing library can result in an apparent homozygous representation in a sequencing library. As most downstream sequencing library quantification techniques depend on stochastic counting processes, these problems have typically been addressed by sampling enough (over-sampling) to obtain a minimum number of observations necessary to make statistically significant decisions. However, the strategy of oversampling is generally limited to elimination of low-count Poisson noise, and the approach wastes resources and increases the expense required to perform such experiments. Moreover, oversampling can result in a reduced statistical confidence in certain conclusions (e.g., diagnostic calls) based on the data. Accordingly, new approaches are needed for overcoming bias in sequencing library preparatory methods.

Aspects of the disclosure are based, in part, on the discovery of methods for overcoming problems associated with systematic and random errors (bias) in genome capture, amplification and sequencing methods, namely high variability in the capture and amplification of nucleic acids and disproportionate representation of heterozygous alleles in sequencing libraries. Accordingly, in some embodiments, the disclosure provides methods that reduce variability in the capture and amplification of nucleic acids. In other embodiments, the methods improve allelic representation in sequencing libraries and, thus, improve variant detection outcomes. In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve the use of differentiator tag sequences to uniquely tag individual nucleic acid molecules. In some embodiments, the differentiator tag sequence permits the detection of bias based on the frequency with which pairs of differentiator tag and target sequences are observed in a sequencing reaction. In other embodiments, the methods reduce errors caused by bias, or the risk of bias, associated with the capture, amplification and sequencing of genetic loci, e.g., for diagnostic purposes.

Petitioner Illumina, Inc., EX1004, p. 412

Aspects of the invention relate to associating unique sequence tags (referred to as differentiator tag sequences) with individual target molecules that are independently captured and/or analyzed (e.g., prior to amplification or other process that may introduce bias). These tags are useful to distinguish independent target molecules from each other thereby allowing an analysis to be based on a known number of individual target molecules. For example, if each of a plurality of target molecule sequences obtained in an assay is associated with a different differentiator tag, then the target sequences can be considered to be independent of each other and a genotype likelihood can be determined based on this information. In contrast, if each of the plurality of target molecule sequences obtained in the assay is associated with the same differentiator tag, then they probably all originated from the same target molecule due to over-representation (e.g., due to biased amplification) of this target molecule in the assay. This provides less information than the situation where each nucleic acid was associated with a different differentiator tag. In some embodiments, a threshold number of independently isolated molecules (e.g., unique combinations of differentiator tag and target sequences) is analyzed to determine the genotype of a subject.

In some embodiments, the invention relates to compositions comprising pools (libraries) of preparative nucleic acids that each comprise "differentiator tag sequences" for detecting and reducing the effects of bias, and for genotyping target nucleic acid sequences. As used herein, a "differentiator tag sequence" is a sequence of a nucleic acid (a preparative nucleic acid), which in the context of a plurality of different isolated nucleic acids, identifies a unique, independently isolated nucleic acid. Typically, differentiator tag sequences are used to identify the origin of a target nucleic acid at one or more stages of a nucleic acid preparative method. For example, in the context of a multiplex nucleic acid capture reaction, differentiator tag sequences provide a basis for differentiating between multiple independent, target nucleic acid capture events. Also, in the context of a multiplex nucleic acid amplification reaction, differentiator tag sequences provide a basis for differentiating between multiple independent, primary amplicons of a target nucleic acid, for example. Thus, combinations of target nucleic acid and differentiator tag sequence (target:differentiator tag sequences) of an isolated nucleic acid of a preparative method provide a basis for identifying unique, independently isolated target nucleic acids. FIG. 4A-C depict various non-limiting examples of methods for combining differentiator tag sequence and target sequences.

Petitioner Illumina, Inc., EX1004, p. 413

It will be apparent to the skilled artisan that differentiator tags may be synthesized using any one of a number of different methods known in the art. For example, differentiator tags may be synthesized by random nucleotide addition.

Differentiator tag sequences are typically of a predefined length, which is selected to control the likelihood of producing unique target:differentiator tag sequences in a preparative reaction (e.g., amplification-based reaction, a circularization selection-based reaction, e.g., a MIP reaction). Differentiator tag sequences may be, up to 5, up to 6, up to 7 up to 8, up to 9, up to 10, up to 11, up to 12, up to 13, up to 14, up to 15, up to 16, up to 17, up to 18, up to 19, up to 20, up to 21, up to 22, up to 23, up to 24, up to 25, or more nucleotides in length. For purposes of genotyping, isolated nucleic acids are identified as independently isolated if they comprise unique combinations of target nucleic acid and differentiator tag sequences, and observance of threshold numbers of unique combinations of target nucleic acid and differentiator tag sequences provide a certain statistical confidence in the genotype.

During a library preparation process, each nucleic acid molecule may be tagged with a unique differentiator tag sequence in a configuration that permits the differentiator tag sequence to be sequenced along with the target nucleic acid sequence of interest (the nucleic acid sequence for which the library is being prepared, e.g., a polymorphic sequence). The incorporation of the nucleic acid comprising a differentiator tag sequence at a particular step allows the detection and correction of biases in subsequent steps of the protocol.

A large library of unique differentiator tag sequences may be created by using degenerate, random-sequence polynucleotides of defined length. The differentiator tag sequences of the polynucleotides may be read at the final stage of the sequencing. The observations of the differentiator tag sequences may be used to detect and correct biases in the final sequencing read-out of the library. For example, the total possible number of differentiator tag sequences, which may be produced, e.g., randomly, is 4^N, where N is the length of the differentiator tag sequence. Thus, it is to be understood that the length of the differentiator tag sequence may be adjusted such that the size of the population of MIPs having unique differentiator tag sequences is sufficient to produce a library of MIP capture products in which identical independent combinations of target nucleic acid and differentiator tag sequence are rare. As used herein combinations of target nucleic acid and differentiator tag sequences, may also be referred to as "target:differentiator tag sequences".

Petitioner Illumina, Inc., EX1004, p. 414

In the final readout of a sequencing process, each read may have an additional unique differentiator tag sequence. In some embodiments, when differentiator tag sequences are distributed randomly in a library, all the unique differentiator tag sequences will be observed about an equal number of times. Accordingly, the number of occurrences of a differentiator tag sequence may follow a Poisson distribution.

In some embodiments, overrepresentation of target:differentiator tag sequences in a pool of preparative nucleic acids (e.g., amplified MIP capture products) is indicative of bias in the preparative process (e.g., bias in the amplification process). For example, target:differentiator tag sequence combinations that are statistically overrepresented are indicative of bias in the protocol at one or more steps between the incorporation of the differentiator tag sequences into MIPs and the actual sequencing of the MIP capture products.

The number of reads of a given target:differentiator tag sequence may be indicative (may serve as a proxy) of the amount of that target sequence present in the originating sample. In some embodiments, the numbers of occurrence of sequences in the originating sample is the quantity of interest. For example, using the methods disclosed herein, the occurrence of differentiator tag sequences in a pool of MIPs may be predetermined (e.g., may be the same for all differentiator tag sequences). Accordingly, changes in the occurrence of differentiator tag sequences after amplification and sequencing may be indicative of bias in the protocol. Bias may be corrected to provide an accurate representation of the composition of the original MIP pool, e.g., for diagnostic purposes.

According to some aspects, a library of preparative nucleic acid molecules (e.g., MIPs, each nucleic acid in the library having a unique differentiator tag sequence, may be constructed such that the number of nucleic acid molecules in the library is significantly larger than the number prospective target nucleic acid molecules to be captured using the library. This ensures that products of the preparative methods include only unique target:differentiator tag sequence; e.g., in a MIP reaction the capture step would undersample the total population of unique differentiator tag sequences in the MIP library. For example, an experiment utilizing 1 ug of genomic DNA will contain about ~150,000 copies of a diploid genome. For a MIP library, each MIP in the library comprising a randomly produced 12-mer differentiator tag sequence (~1.6 million possible unique differentiator tag sequences), there would be more than 100 unique

Petitioner Illumina, Inc., EX1004, p. 415

differentiator tag sequences per genomic copy. For a MIP library, each MIP in the library comprising a randomly produced 15-mer differentiator tag sequence (~1 billion possible unique differentiator tag sequences), there would be more than 7000 unique differentiator tag sequences per genomic copy. Therefore, the probability of the same differentiator tag sequence being incorporated multiple times is incredibly small. Thus, it is to be appreciated that the length of the differentiator tag sequence is to be selected based on the amount of target sequence in a MIP capture reaction and the desired probability for having multiple, independent occurrences of target:differentiator tag sequence combinations.

FIG. 5 depicts a non-limiting method for genotyping based on target and differentiator tag sequences. Sequencing reads of target and differentiator tags sequences are collapsed to make diploid genotype calls. FIG. 6 depicts non-limiting results of a simulation of a MIP capture reaction in which MIP probes, each having a differentiator tag sequence of 15 nucleotides, are combined with 10000 target sequence copies (e.g., genome equivalents). In this simulated reaction, the probability of capturing one or more copies of a target sequence having the same differentiator tag sequence is 0.05. The Y axis reflects the number of observations. The X axis reflects the number of independent occurrences of target:differentiator tag combinations. FIG. 7 depicts a non-limiting graph of sequencing coverage, which can help ensure that alleles are sampled to sufficient depth (e.g., either 10x or 20x minimum sampling per allele, assuming 1000 targets). In this non-limiting example, the X axis is total per-target coverage required, and the Y axis is the probability that a given total coverage will result in at least 10x or 20x coverage for each allele.

The skilled artisan will appreciate that as part of a MIP library preparation process, adapters may be ligated onto the ends of the molecules of interest. Adapters often contain PCR primer sites (for amplification or emulsion PCR) and/or sequencing primer sites. In addition, barcodes may be included, for example, to uniquely identify individual samples (e.g., patient samples) that may be mixed together. (See, e.g., USPTO Publication Number US 2007/0020640 Al (McCloskey et al.)

The actual incorporation of the random differentiator tag sequences can be performed through various methods known in the art. For example, nucleic acids comprising differentiator tag sequences may be incorporated by ligation. This is a flexible method, because molecules having differentiator tag sequence can be ligated to any blunt-ended nucleic acids. The

Petitioner Illumina, Inc., EX1004, p. 416

sequencing primers must be incorporated subsequently such that they sequence both the differentiator tag sequence and the target sequence. Alternatively, the sequencing adaptors can be synthesized with the random differentiator tag sequences at their 3' end (as degenerate bases), so that only one ligation must be performed. Another method is to incorporate the differentiator tag sequence into a PCR primer, such that the primer structure is arranged with the common adaptor sequence followed by the random differentiator tag sequence followed by the PCR priming sequence (in 5' to 3' order). A differentiator tag sequence and adaptor sequence (which may contain the sequencing primer site) are incorporated as tags. Another method to incorporate the differentiator tag sequences is to synthesize them into a padlock probe prior to performing a gene capture reaction. The differentiator tag sequence is incorporated 3' to the targeting arm but 5' to the amplification primer that will be used downstream in the protocol. Another method to incorporate the differentiator tag sequences is as a tag on a gene-specific or poly-dT reverse-transcription primer. This allows the differentiator tag sequence to be incorporated directly at the cDNA level.

In some embodiments, at the incorporation step, the distribution of differentiator tag sequences can be assumed to be uniform. In this case, bias in any part of the protocol would change the uniformity of this distribution, which can be observed after sequencing. This allows the differentiator tag sequence to be used in any preparative process where the ultimate output is sequencing of many molecules in parallel.

Differentiator tag sequences may be incorporated into probes (e.g., MIPs) of a plurality when they are synthesized on-chip in parallel, such that degeneracy of the incorporated nucleotides is sufficient to ensure near-uniform distribution in the plurality of probes. It is to be appreciated that amplification of a pool of unique differentiator tag sequences may itself introduce bias in the initial pool. However, in most practical cases, the scale of synthesis (e.g., by column synthesis, chip based synthesis, etc.) is large enough that amplification of an initial pool of differentiator tag sequences is not necessary. By avoiding amplification or selection steps on the pool of unique differentiator tag sequences, potential bias may be minimized.

One example of the use of the differentiator tag sequences is in genome re-sequencing. Considering that the raw accuracy of most next-generation sequencing instruments is relatively low, it is crucial to oversample the genomic loci of interest.

Petitioner Illumina, Inc., EX1004, p. 417

Furthermore, since there are two alleles at every locus, it is important to sample enough to ensure that both alleles have been observed a sufficient number of times to determine with a sufficient degree of statistical confidence whether the sample is homozygous or heterozygous. Indeed, the sequencing is performed to sample the composition of molecules in the originating sample. However, after multiple reads have been collected for a given locus, it is possible that due to bias (e.g., caused by PCR amplification steps), a large fraction of the reads are derived from a single originating molecule. This would skew the population of target sequences observed, and would affect the outcome of the genotype call. For example, it is possible that a locus that is heterozygous is called as homozygous, because there are only a few observations of the second allele out of many observations of that locus. However, if information is available on differentiator tag sequences, this situation could be averted, because the over-represented allele would be seen to also have an over-represented differentiator tag sequence (i.e., the sequences with the overrepresented differentiator tag sequence all originated from the same single molecule). Therefore, the sequences and corresponding distribution of differentiator tag sequences can be used as an additional input to the genotype-calling algorithm to significantly improve the accuracy and confidence of the genotype calls.

In some aspects, the disclosure provides methods for analyzing a plurality of target sequences which are genetic loci or portions of genetic loci (e.g., a genetic locus of Table 1). The genetic loci may be analyzed by sequencing to obtain a genotype at one or more polymorphisms (e.g., SNPs). Exemplary polymorphisms are disclosed in Table 2. The skilled artisan will appreciate that other polymorphisms are known in the art and may

Petitioner Illumina, Inc., EX1004, p. 418

be identified, for example, by querying the Entrez Single Nucleotide Polymorphism database, for example, by searching with a GeneID from Table 1.

Table 1: Target Nucleic Acids

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
CYP21A2	1589	cytochrome P450, family 21, subfamily A, polypeptide 2	CAH1; CPS1; CA21H; CYP21; CYP21B; P450c21B; MGC150536; MGC150537; CYP21A2	201910	NG_008337.1	6p21.3
ABCC8	6833	ATP-binding cassette, sub-family C (CFTR/MRP), member 8	HI; SUR; HHF1; MRP8; PHHI; SUR1; ABC36; HRINS; TNDM2; ABCC8	600509	NG_008867.1	11p15.1
ATRX	546	alpha thalassemia/mental retardation syndrome X-linked (RAD54 homolog, S. cerevisiae)	SHS; XH2; XNP; ATR2; SFM1; RAD54; MRXHF1; RAD54L; ZNF-HX; MGC2094; ATRX	300032	NG_008838.1	Xq13.1-q21.1
ARSA	410	arylsulfatase A	MLD; ARSA	607574	NG_009260.1	22q13.31-qter; 22q13.33
PSAP	5660	Prosaposin	GLBA; SAP1; FLJ00245; MGC110993; PSAP	176801	NG_008835.1	10q21-q22
BTD	686	Biotinidase	BTD	609019	NG_008819.1	3p25

5

Petitioner Illumina, Inc., EX1004, p. 419

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
HLCS	3141	holocarboxylase synthetase (biotin-(proprionyl-Coenzyme A-carboxylase (ATP-hydrolysing)) ligase)	HCS; HLCS	609018	NC_000021.7	21q22.1; 21q22.13
BLM	641	Bloom syndrome, RecQ helicase-like	BS; RECQ2; RECQL2; RECQL3; MGC126616; MGC131618; MGC131620; BLM	604610	NG_007272.1	15q26.1
ASPA	443	aspartoacylase (Canavan disease)	ASP; ACY2; ASPA	608034	NG_008399.1	17pter-p13
CFTR	1080	cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7)	CF; MRP7; ABC35; ABCC7; CFTR/MRP; TNR-CFTR; dJ760C5.1; CFTR	602421	NC_000007.12	7q31.2
ASS1	445	argininosuccinate synthetase 1	ASS; CTLN1; ASS1	603470	NG_011542.1	9q34.1
MMACHC	25974	methylmalonic aciduria (cobalamin deficiency) cblC type, with homocystinuria	cblC; FLJ25671; DKFZp564I122; RP11-291L19.3; MMACHC	609831	NC_000001.9	1p34.1
IKBKAP	8518	inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase complex-associated protein	FD; DYS; ELP1; IKAP; IKI3; TOT1; FLJ12497; DKFZp781H1425; IKBKAP	603722	NG_008788.1	9q31
FANCC	2176	Fanconi anemia, complementation group C	FA3; FAC; FACC; FLJ14675; FANCC	227645	NG_011707.1	9q22.3
GALK1	2584	galactokinase 1	GK1; GALK; GALK1	604313	NG_008079.1	17q24
GALT	2592	galactose-1-phosphate uridylyltransferase	GALT	606999	NC_000009.10	9p13
GALE	2582	UDP-galactose-4-epimerase	SDR1E1; FLJ95174; FLJ97302; GALE	606953	NG_007068.1	1p36-p35
GBA	2629	glucosidase, beta; acid (includes glucosylceramidase)	GCB; GBA1; GLUC; GBA	606463	NG_009783.1	1q21
GJB2	2706	gap junction protein, beta 2, 26kDa	HID; KID; PPK; CX26; DFNA3; DFNB1; NSRD1; DFNA3A; DFNB1A; GJB2	121011	NG_008358.1	13q11-q12
GCDH	2639	glutaryl-Coenzyme A dehydrogenase	GCD; ACAD5; GCDH	608801	NG_009292.1	19p13.2
G6PC	2538	glucose-6-phosphatase, catalytic subunit	G6PT; GSD1; GSD1a; MGC163350; G6PC	232200	NG_011808.1	17q21
HBB	3043	hemoglobin, beta	CD113t-C; beta-globin; HBB	141900	NG_000007.3	11p15.5

Petitioner Illumina, Inc., EX1004, p. 420

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
BCKDHA	593	branched chain keto acid dehydrogenase E1, alpha polypeptide	MSU; MSUD1; OVD1A; BCKDE1A; FLJ45695; BCKDHA	608348	NC_000019.8	19q13.1-q13.2
BCKDHB	594	branched chain keto acid dehydrogenase E1, beta polypeptide	E1B; FLJ17880; dJ279A18.1; BCKDHB	248611	NG_009775.1	6q13-q15
DBT	1629	dihydrolipoamide branched chain transacylase E2	E2; E2B; BCATE2; MGC9061; DBT	248610	NG_011852.1	1p31
DLD	1738	dihydrolipoamide dehydrogenase	E3; LAD; DLDH; GCSL; PHE3; DLD	238331	NG_008045.1	7q31-q32
ACADM	34	acyl-Coenzyme A dehydrogenase, C-4 to C-12 straight chain	MCAD; ACAD1; MCADH; FLJ18227; FLJ93013; FLJ99884; ACADM	607008	NG_007045.1	1p31
MMAA	166785	methylmalonic aciduria (cobalamin deficiency) cblA type	cblA; MGC120010; MGC120011; MGC120012; MGC120013; MMAA	607481	NG_007536.1	4q31.22
MMAB	326625	methylmalonic aciduria (cobalamin deficiency) cblB type	ATR; cblB; MGC20496; MMAB	607568	NG_007096.1	12q24
MUT	4594	methylmalonyl Coenzyme A mutase	MCM; MUT	609058	NG_007100.1	6p12.3
MCOLN1	57192	mucolipin 1	ML4; MLIV; MST080; TRPML1; MSTP080; TRP-ML1; TRPM-L1; MCOLN1	605248	NC_000019.8	19p13.3-p13.2
ACTA1	58	actin, alpha 1, skeletal muscle	ACTA; ASMA; CFTD; MPFD; NEM1; NEM2; NEM3; CFTD1; CFTDM; ACTA1	102610	NG_006672.1	1q42.13
TPM3	7170	tropomyosin 3	TM3; TRK; NEM1; TM-5; TM30; TM30nm; TPMsk3; hscp30; MGC3261; FLJ41118; MGC14582; MGC72094; OK/SW-cl.5; TPM3	191030	NG_008621.1	1q21.2
TNNT1	7138	troponin T type 1 (skeletal, slow)	ANM; TNT; STNT; TNTS; FLJ98147; MGC104241; TNNT1	191041	NG_011829.1	19q13.4
NEB	4703	nebulin	NEM2; NEB177D; FLJ11305; FLJ36536; FLJ39568; FLJ39584; DKFZp686C1456; NEB	161650	NG_009382.1	2q22
SMPD1	6609	sphingomyelin phosphodiesterase 1, acid lysosomal	ASM; NPD; SMPD1	607608	NG_011780.1	11p15.4-p15.1
GLDC	2731	glycine dehydrogenase (decarboxylating)	GCE; NKH; GCSP; HYGN1; MGC138198; MGC138200; GLDC	238300	NC_000009.10	9p22

Petitioner Illumina, Inc., EX1004, p. 421

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
GCSH	2653	glycine cleavage system protein H (aminomethyl carrier)	GCE; NKH; GCSH	238330	NC_000016.8	16q23.2
AMT	275	aminomethyltransferase	GCE; NKH; GCST; AMT	238310	NC_000003.10	3p21.2-p21.1
OTC	5009	ornithine carbamoyltransferase	OCTD; MGC129967; MGC129968; MGC138856; OTC	300461	NG_008471.1	Xp21.1
PAH	5053	phenylalanine hydroxylase	PH; PKU; PKU1; PAH	612349	NG_008690.1	12q22-q24.2
DHPR	5860	quinoid dihydropteridine reductase	DHPR; PKU2; SDR33C1; FLJ42391; QDPR	612676	NG_008763.1	4p15.31
PTS	5805	6-pyruvoyltetrahydropterin synthase	PTPS; FLJ97081; PTS	261640	NG_008743.1	11q22.3-q23.3
PCCA	5095	propionyl Coenzyme A carboxylase, alpha polypeptide	PCCA	232000	NG_008768.1	13q32
PCCB	5096	propionyl Coenzyme A carboxylase, beta polypeptide	DKFZp451E113; PCCB	232050	NG_008939.1	3q21-q22
ACADS	35	acyl-Coenzyme A dehydrogenase, C-2 to C-3 short chain	SCAD; ACAD3; ACADS	606885	NG_007991.1	12q22-qter
DHCR7	1717	7-dehydrocholesterol reductase	SLOS; DHCR7	602858	NC_000011.8	11q13.2-q13.5
SMN1	6606	survival of motor neuron 1, telomeric	SMA; SMN; SMA1; SMA2; SMA3; SMA4; SMA@; SMNT; BCD541; T-BCD541; SMN1	600354	NG_008691.1	5q13
HEXA	3073	hexosaminidase A (alpha polypeptide)	TSD; MGC99608; HEXA	606869	NG_009017.1	15q23-q24
MYO7A	4647	myosin VIIA	DFNB2; MYU7A; NSRD2; USH1B; DFNA11; MYOVIIA; MYO7A	276903	NG_009086.1	11q13.5
USH1C	10083	Usher syndrome 1C (autosomal recessive, severe)	PDZ73; AIE-75; DFNB18; PDZ-45; PDZ-73; NY-CO-37; NY-CO-38; ush1cpst; PDZ-73/NY-CO-38; USH1C	605242	NC_000011.8	11p15.1-p14
CDH23	64072	cadherin-like 23	USH1D; DFNB12; FLJ00233; FLJ36499; KIAA1774; KIAA1812; MGC102761; DKFZp434P2350; CDH23	605516	NG_008835.1	10q21-q22
PCDH15	65217	protocadherin 15	USH1F; DFNB23; DKFZp667A1711; PCDH15	605514	NG_009191.1	10q21.1

Petitioner Illumina, Inc., EX1004, p. 422

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
SANS	124590	Usher syndrome 1G (autosomal recessive)	SANS; ANKS4A; FLJ33924; USH1G	607696	NG_007882.1	17q25.1
ARX	170302	aristaless related homeobox	ISSX; PRTS; MRX29; MRX32; MRX33; MRX36; MRX38; MRX43; MRX54; MRX76; MRX87; MRXS1; ARX	300382	NG_008281.1	Xp21
OPHN1	4983	oligophrenin 1	OPN1; MRX60; OPHN1	300127	NG_008960.1	Xq12
JARID1C	8242	lysine (K)-specific demethylase 5C	MRXJ; SMCX; MRXSJ; XE169; JARID1C; DXS1272E; KDM5C	314690	NG_008085.1	Xp11.22-p11.21
FTSJ1	24140	FtsJ homolog 1 (E. coli)	JM23; MRX9; SPB1; TRM7; CDLIV; MRX44; FTSJ1	300499	NG_008879.1	Xp11.23
SLC6A8	6535	solute carrier family 6 (neurotransmitter transporter, creatine), member 8	CRT; CT1; CRTR; MGC87396; SLC6A8	300036	NC_000023.9	Xq28
DLG3	1741	discs, large homolog 3 (Drosophila)	MRX; MRX90; NEDLG; NE-Dlg; SAP102; SAP-102; KIAA1232; DLG3	300189	NC_000023.9	Xq13.1
TM4SF2	7102	tetraspanin 7	AI5; MXS1; CD231; MRX58; CCG-B7; TM4SF2; TALLA-1; TM4SF2b; DXS1692E; TSPAN7	300096	NG_009160.1	Xp11.4
ZNF41	7592	zinc finger protein 41	MRX89; MGC8941; ZNF41	314995	NG_008238.1	Xp11.23
FACL4	2182	acyl-CoA synthetase long-chain family member 4	ACS4; FACL4; LACS4; MRX63; MRX68; ACSL4	300157	NG_008053.1	Xq22.3-q23
PQBP1	10084	polyglutamine binding protein 1	SHS; MRX55; MRXS3; MRXS8; NPW38; RENS1; PQBP1	300463	NC_000023.9	Xp11.23
PEX1	5189	peroxisomal biogenesis factor 1	ZWS1; PEX1	602136	NG_008341.1	7q21.2
PXMP3	5828	peroxisomal membrane protein 3, 35kDa	PAF1; PEX2; PMP3; PAF-1; PMP35; RNF72; PXMP3	170993	NG_008371.1	8q21.1
PEX6	5190	peroxisomal biogenesis factor 6	PAF2; PAF-2; PXAAA1; PEX6	601498	NG_008370.1	6p21.1
PEX10	5192	peroxisomal biogenesis factor 10	NALD; RNF69; MGC1998; PEX10	602859	NG_008342.1	1p36.32
PEX12	5193	peroxisomal biogenesis factor 12	PAF-3; PEX12	601758	NG_008447.1	17q12
PEX5	5830	peroxisomal biogenesis factor 5	PXR1; PTS1R; PTS1-BP; FLJ50634; FLJ50721; FLJ51948; PEX5	600414	NG_008448.1	12p13.31
PEX26	55670	peroxisomal biogenesis factor 26	FLJ20695; PEX26M1T; Pex26pM1T; PEX26	608666	NG_008339.1	22q11.21

Petitioner Illumina, Inc., EX1004, p. 423

The mutations listed in Table 2 are documented polymorphisms in several disease-associated genes (CFTR is mutated in cystic fibrosis, GBA is mutated in Gaucher disease, ASPA is mutated in Canavan disease, HEXA is mutated in Tay Sachs disease). The polymorphisms are of several types: insertion/deletion polymorphisms which will cause frameshifts (and thus generally interrupt protein function) unless the insertion/deletion length is a multiple of 3bp, and substitutions which can alter the amino acid sequence of the protein and in some cases cause complete inactivation by introduction of a stop codon.

Table 2: Non-limiting examples of polymorphisms

Gene name	GeneID	SNP ID	Mutation	SEQ ID NO:
CFTR	1080	rs63500661	TCACATCACCAAGTTAAAAAAAAAAA[A/G]G GGGCGGGGGGGCAGAATGAAAATT	1
CFTR	1080	rs63107760	AAACAAGGATGAATTAAGTTTTTTTT[-/T]AAAAAAGAAACATTTGGTAAGGGGA	2
CFTR	1080	rs62469443	ATCACCAAGTTAAAAAAAAAAAAGGG[A/G]C GGGGGGGCAGAATGAAAATTGCAT	3
CFTR	1080	rs62469442	CTATTGAACCAGAACCAAACAGGAAT[A/G]C CATAGCATTTTGTAAACTAAACTG	4
CFTR	1080	rs62469441	CAGGAGTTCAAGACCAGCCTACTAAA[A/C]C ACACACACACACACACACACACAC	5
CFTR	1080	rs62469439	GATTAAATAATAGTGTTTATGTACCC[C/G]GC TTATAGGAGAAGAGGGTGTGTGT	6
CFTR	1080	rs62469438	ATTGTTATCTTTTCATATAAGGTAAC[A/T]GA GGCCCAGAGAGATTAAATAACAT	7
CFTR	1080	rs62469437	TAATTTTAATTAAGTAAATTTAATTG[A/G]TA GATAAATAAGTAGATAAAAAATA	8
CFTR	1080	rs62469436	GTATAAAAAAAAAAAAAAAAAAAGTT[A/T]G AATGTTTTCTTGCATTCAGAGCCT	9
CFTR	1080	rs62469435	ATACTAAAAATTTAAAGTTCTCTTGC[A/G]AT ATATTTTCTTAATATCTTACATC	10
CFTR	1080	rs62469434	TGCTGGGATTACAGGCGTGAGCCACC[A/G]C GCCTGGCCTGATGGGACATATTTT	11
CFTR	1080	rs62469433	CTACAATATAAGTATAGTATTGCAAA[A/C]CC ATCAGGAAGGGTGTTAACTATTT	12
CFTR	1080	rs61763210	GTTGTCTCCAAACTTTTTTTCAGGTG[-/AGA]AGGTGGCCAACCGAGCTTCGGAAAG	13
CFTR	1080	rs61720488	TTTTTTCATAAAAGATTATATAAAGG[A/C]TA TTGCTTTTGAATCACAAACACTA	14
CFTR	1080	rs61481156	ATCTAGTGAGCAGTCAGGAAAGAGAA[C/T]T TCCAGATCCTGGAAATCAGGGTTA	15
CFTR	1080	rs61443875	TAGAGTATAAAAAAAAAAAAAAAAAA[-/A]GTTTGAATGTTTTCTTGCATTCAGA	16

Petitioner Illumina, Inc., EX1004, p. 424

CFTR	1080	rs61312222	TGCAAATGCCAACTATCAAAGATATT[C/G]GA GTATACTGTCAATAAACTTCATA	17
CFTR	1080	rs61159372	TCCTCAACAGTTAGAAACAATATTTT[C/G]AG TGATTTCCCATGCCAACTTTACT	18
CFTR	1080	rs61094145	TTTTTGGTATTGTTGTTAAATAAGTG[A/G]GA ATTCAATACAGTATAATGTCTGT	19
CFTR	1080	rs61086387	CTTGAAATCGGATATATATATATATATA[- /TGTATATATATATATATATATATATATAT ACATATATATATATA]GTATTATCCCTGTTTTC ACAGTTTT	20
CFTR	1080	rs60996744	AGAGGGGCTGTGAAGGACACCAAGGA[A/G]G AGACTAAGAGCCAGGAGGGAAAAC	21
CFTR	1080	rs60960860	TAGAGTTTATTAGCTTTTACTACTCT[A/G]CTT AGTTACTTTGTGTTACAGAATA	22
CFTR	1080	rs60923902	ACTAGTGATGATGAGCTTCTTTTCAT[- /AT]GTTTGTTGGCTGCATAAATGTCTTC	23
CFTR	1080	rs60912824	GCAGAGAAAAGAGGGGCTGTGAAGGA[C/G]A CCAAGGAGGAGACTAAGAGCCAGG	24
CFTR	1080	rs60887846	TTCAGAGGTCTACCACTGGTGCATAC[G/T]CT AATCACAGTGTCGAAAATTTTAC	25
CFTR	1080	rs60793174	AAGAAAGAGCAAAAGAGGGCAAACTT[C/T]T CATACATTTTTGATGTCGAAACCA	26
CFTR	1080	rs60788575	CCTAAAGTTTAAAAAGAAAAAAAAAA[- /A]GGAAGAAGGAATTAAAAATCCAAAG	27
CFTR	1080	rs60760741	GTGTGTGTGTGTATATATATATATATAT[A/T]TA TATATTTTTTTTTTCCTGAGCCA	28
CFTR	1080	rs60456599	AAACTGTTGATGTTTTCATTTATTTA[C/G]ATC ATTGGAAAACTTTAGATTCTAG	29
CFTR	1080	rs60363249	TTTATCCATTCTTAACCAGAACAGAC[A/G]TT TTTTCAGAGCTGGTCCAGGAAAA	30
CFTR	1080	rs60355115	TTGAAATCGGATATATATATATATATAT[A/G]TA TATATATATATATATATATATAT	31
CFTR	1080	rs60308689	TAGTTTTTTATTTCCTCATATTATTT[- /T]CAGTGGCTTTTTCTTCCACATCTTT	32
CFTR	1080	rs60271242	ACATAGTTCTCAGTGGTACAACTACA[A/G]GT GATTTCTCTTTTCTTATTTCTGG	33
CFTR	1080	rs60010318	AGAGCAATGGCATCCCTTGTCTTGTG[C/T]TA TACAGGATGCAGCAATTTATAGG	34
CFTR	1080	rs59961323	TTCTGTCTACATAAGATGTCATACTA[A/G]AT TATCTTTTCCAGCATGCATTCAG	35
CFTR	1080	rs59961270	CAGGGTGGCATGTTAGGCAGTGCTTA[A/G]A ATAAATGAGTTGGTTATACAAGTA	36
CFTR	1080	rs59837506	AGGACACACACACACACACACACACA[- /CA]TGCACACACATTTAAATAGATGCAT	37
CFTR	1080	rs59572090	TAAAAAATTGGTATAATGAAATTGCA[C/T]TT GTAGTCTTTGGACATTTAAATCC	38
CFTR	1080	rs59548252	TTTCAATACTTAAGAGGTACGCAGAG[A/G]A AAGAGGGGCTGTGAAGGACACCAA	39
CFTR	1080	rs59519859	CAGCAATGAATATTTTGAGGCTGAGG[C/T]GC TGAGGGGTAAAATTGCAGCCTGG	40
CFTR	1080	rs59509837	TTATGGTTTATATTTTTGTGTCTTCT[- /CTTT]AACACATCTTTTCTAGCAGAATTCA	41

Petitioner Illumina, Inc., EX1004, p. 425

CFTR	1080	rs59417037	GTATTTTAGTTTTTTTTTTTTGTTTG[-/T]TTTGTTTTGTTTTGTTTTGTTTTTG	42
CFTR	1080	rs59159458	TGGGTGACTCCATTTTTACTTTTAGT[C/T]TGG TCTGTTGAGGCCTCGTGAGAGA	43
CFTR	1080	rs59048119	TATTTTCATGTATTTTAGTTTTTTTTT[-/TTTT]GTTTGTTTTGTTTTGTTTTGTTTTG	44
CFTR	1080	rs58970500	GTGTGTGTGTATATATATATATATAT[A/T]TA TATTTTTTTTTTCCTGAGCCAAA	45
CFTR	1080	rs58942292	AACCTATTAGCATGTCTGGCAGAAAA[-/A]TAGATACTTAATAAATTTCTTAAAT	46
CFTR	1080	rs58917054	GAGGCTTAGACAGTTTAAGTAACTCA[A/G]G CATGGTTACACAACTAGCTAGGGC	47
CFTR	1080	rs58837484	GTGTGAGTATTATGAGACCATATGTT[A/G]GG AGATTTTATTTGGTATTGAGGAT	48
CFTR	1080	rs58829491	GAAACCCCACCCCTTCTATAGTTTTC[C/T]CTT TAATATTTACAATGGAACCATT	49
CFTR	1080	rs58805195	CATATATATATAGTGTGTGTGTGTGT[A/G]TA TATATATATATATATATATTTTT	50
GBA	2629	rs60866785	CGAGCGAGAGAGAGAGAGAGAGAGAG[-/AG]GAGCCGGCGCGAGAACTACGCATGC	51
GBA	2629	rs60239603	GGCAGGTAATATCTAGTACCTTACTT[A/T]TA TTTCCTGAGCACATTCTACATTT	52
GBA	2629	rs56310840	GGCCAGGAATGGGAGTGCTTAGGTGC[A/G]G AGGTGGCACTGTTCCCGCAGCTGC	53
GBA	2629	rs41264927	TGAAAACTCCATCCCCTCAGGGTCAT[C/T]AG ATGAAGAGAAGACCACAGGGGTT	54
GBA	2629	rs41264925	TGTAGGTAAGGGTCACATGTGGGAGA[C/G]G CAGCTGTGGGTAGGTCAGCCCTGT	55
GBA	2629	rs36024691	CCAAGAAGGCGCCATTACACTCCAGC[-/C]TGGGCGACAGGGCGAGACTCCCTCA	56
GBA	2629	rs36024092	TGCCACACCCAGCTAATTTGTGTGTG[-/G]TATGTGTGTGTATGTATGTGTGTGT	57
GBA	2629	rs35682967	GTTCCTCCAGTAATTTTTTTTTTTTT[-/T]CGTTTTGAGACAGAGTCTTGCCCTG	58
GBA	2629	rs35033592	ATCATGCCCAGATAATTTTTTTTTT[-/T]GTATTTTAGTAGACACAGGGTTTCA	59
GBA	2629	rs34732744	CGAGCGAGAGAGAGAGAGAGAGAGAG[-/AG]GAGCCGGCGCGAGAACTACGCATGC	60
GBA	2629	rs34620635	CCTGTGAGGGGCACATTCCTTAGTAG[-/C]TAAGGAGTTGGGGGTGTGAAGATCC	61
GBA	2629	rs34302637	ACAGGCTACTGGCTGGGCCCAGGCAA[-/A]GGGGGCCTTGGCAGGAAAAGTTCCT	62
GBA	2629	rs33949225	GCGAGAGAGAGAGAGAGAGAGAGAGG[-/AG]AGCCGGCGCGAGAACTACGCATGCG	63
GBA	2629	rs28678003	AAGAAGAAAAATAAAAAGAAAGTGGG[C/T]C AGACCGAGAGAACAGGAAGCCTGA	64
GBA	2629	rs28559737	AAGGACAAAGGCAAAGAGACAAAGGC[G/T]C AACACTGGGGGTCCCCAGAGAGTG	65
GBA	2629	rs28373017	TACCTAGTCACTTCCTGCCTCCATGG[C/T]GC AAAAGGGGATGGGTGTGCCTCTT	66
GBA	2629	rs12752133	CTCTTCCGAGGTTCCACCCTGAACAC[C/T]TT CCTGCTCCCTCGTGGTGTAGAGT	67

Petitioner Illumina, Inc., EX1004, p. 426

GBA	2629	rs12747811	TTCTGACTGGCAACCAGCCCCACTCT[C/T]TG GGAGCCCTCAGGAATGAACTTGC	68
GBA	2629	rs12743554	gctcagcctcccaggctggagtgcag[A/T]ggcgcgatctcggctcac cgcaacc	69
GBA	2629	rs12041778	CATGAACCACATCAAATGAGATTTAG[C/T]GG GAGTGGCACACACAGTCATGACC	70
GBA	2629	rs12034326	AAGCAGCCCTGGGGAGTCGGGGCGGG[A/G]C CTGGATTGGAAAAGAGACGGTCAC	71
GBA	2629	rs11558184	CTCCAAGTTCTGGGAGCAGAGTGTGC[A/G]G CTAGGCTCCTGGGATCGAGGGATG	72
GBA	2629	rs11430678	GTTCCTCCAGTAAttttttttttt[- /G/T]gtttgagacagagtcttgccctgt	73
GBA	2629	rs11264345	CTAGTACCTTACTTCCCTCAAGTTCA[A/T]TC ATCTCACAGATATTTCCTGAGCA	74
GBA	2629	rs10908459	aattagccgtgcggtggtggcgggtgc[C/T]gtaatcccacgtacttgg gaggct	75
GBA	2629	rs10796940	CCATGGCCAGCCGGGGAGGGGACGGG[A/C]A CACACAGACCCACACAGAGACTCA	76
GBA	2629	rs10668496	agcgagagagagagagagagagagag[- /AG]gagCCGGCGCGAGAACTACGCATGC	77
GBA	2629	rs7416991	CGTAGCAGTTAGCAGATGATAGGCGG[C/G/T] GAAATCTTATTTCACAGGGCATTAA	78
GBA	2629	rs4024049	CTGGCCCTGGTGACAGTGGGGCTGTG[C/T]GT GGGGCCAGAGCCTTCTCAGAGGT	79
GBA	2629	rs4024048	CAGATACTGGCCCTGGTGACAGTGGG[A/G]C TGTGCGTGGGGCCAGAGCCTTCTC	80
GBA	2629	rs4024047	GACAGATACTGGCCCTGGTGACAGTG[G/T]G GCTGTGCGTGGGGCCAGAGCCTTC	81
GBA	2629	rs3841430	GGCTCctctctctctctctctctc[- /TC]gctcgctctctcgctctctcgctct	82
GBA	2629	rs3754485	GTTTCAGACCAGCCTGGCCAACATAG[C/T]GA AACCCCATCTCTACTAAAAATAA	83
GBA	2629	rs3205619	AGTGGGCGATTGGATGGAGCTGAGTA[C/T]G GGGCCCATCCAGGCTAATCACACC	84
GBA	2629	rs2990227	CCGGGCTCCGTGAATGTTTGTCACAT[C/G]TC TGAAGAACGTATGAATTACATAA	85
GBA	2629	rs2990226	GAATCCCAACCCCGACGCTCGTCGCC[C/G]G GCTCCGTGAATGTTTGTCACATGT	86
GBA	2629	rs2990225	GCGAATCCCAACCCCGACGCTCGTCG[C/T]CG GGCTCCGTGAATGTTTGTCACAT	87
GBA	2629	rs2990224	TGGGCAGAAGTCAGGGTCCAAAGAAA[G/T]G GCAAAGAAAAGTGTcagtggctca	88
ASPA	443	rs63751297	TAAGAAAGACGTTTTTGATTTTTTTC[A/G]GA CTTCTCTGGCTCCACTACCCTGC	89
ASPA	443	rs62071301	CTGATTCCTGGCCAGGAGCGGTGGCT[C/T]AC GCCTGTAATCCCAGCGCTTTGGG	90
ASPA	443	rs62071300	TAAAAATGCTGATTCCTGGCCAGGAG[C/T]GG TGGCTCACGCCTGTAATCCCAGC	91
ASPA	443	rs62071299	TTTAAAAATGCTGATTCCTGGCCAGG[A/C]GC GGTGGCTCACGCCTGTAATCCCA	92
ASPA	443	rs62071297	CAAGACCTGTCAAAGATCTGAGAAAT[A/T]TT ACCCGACTTACAAGCTAACCATT	93

Petitioner Illumina, Inc., EX1004, p. 427

ASPA	443	rs61697033	ACTGTAATAAGTGCTGTAAAAGAAAT[A/G]C ACAAAATAATATAGCAGAGGGTAT	94
ASPA	443	rs60743592	CTTGAGGTCAGGAGTTCAAGACCAGT[C/T]TG GGCAACATGGGGAAAACCTTGTC	95
ASPA	443	rs60666840	AGGTTGCAGTGAGCCGAGATCATGCC[A/G]TT GCACTCCAGCCGGGGCAACAAAA	96
ASPA	443	rs60147514	ACAAGTGTCTTGAAATTATCTGTGAT[C/T]TG CTATAGAGCAATACTTTTGTAAA	97
ASPA	443	rs59930743	GTGGGTATATGCAGCTCTATGCACTA[C/T]CT GCTCATTTATTTGGTAAATCTAA	98
ASPA	443	rs59690349	TGTGTGTGTGTGCGTGTGTGTGTGTG[-/TGTGTGTG]ATCATAAGAGTGGCTGCAGCAA ACT	99
ASPA	443	rs59676360	AGTCTGGAGTGCAATGGTGCAATCTC[A/G]GC TCACTGCAGCCTCCACCTCCGGG	100
ASPA	443	rs59335404	CTCCTAATGGATATTTCCTAAATTTT[G/T]CTG AACAGAATTTAACTTGAGCTGG	101
ASPA	443	rs58879097	ATTTAAAAATGGATTTCTAGAAAAAC[A/G]AT CACATACTTGAATATTTTAGCAA	102
ASPA	443	rs58686774	CTATAAATGGGTAGCATGAGGGATTC[A/G]A GGAGGTGGCTGAAAGAAGCACGTA	103
ASPA	443	rs57511162	AAGAAACCAAGCATAGTAGAGTGTTA[A/G]A AAACCAAAGCAACTAAACAACTGT	104
ASPA	443	rs55859596	CGGGGCTCAGAACTTGTAACAGAAAA[A/T]T AAAATATACTCCACTCAAGGGAAT	105
ASPA	443	rs55742972	TACTACACTTCACGGATACTGTACTT[-/GTACTT]TTTTTCCAAATTGAAGGTTTTTGGC	106
ASPA	443	rs55640436	TTGTTTTTGTTTTTGTTTTTGTTTTT[-/GTTTTTGTTTTT]TGAGATGGAGTCTCGCTCT GTCGCC	107
ASPA	443	rs36225687	TTTGCCTTACTACACTTCACGGATAC[-/TGTACT]TGTACTTTTTTTCCAAATTGAAGGT	108
ASPA	443	rs36051310	GAGGTGGCTGAAAGAAGCACGTATCC[-/C]TGATGGCATGGTTGCGGGTTATATG	109
ASPA	443	rs36034906	GAGAAAAGCAGTTCCTGGAACACCCC[-/C]ACCCCTTAACCCCTTATCTCTGCTT	110
ASPA	443	rs36033666	TTACATATGTATACATGTGCCATGTT[-/T]GGTGTGCCGCACCCATTAACTCGTC	111
ASPA	443	rs35730123	CTTTTTCCAGATTTTTTTTTTTTTTT[-/T]GAGACAGAGTTTCACTCTTGTTGCC	112
ASPA	443	rs35629100	TTTGGAAATCTTAAGCTTTTATTTGG[-/G]TGTCACAGAGAAACAGGATCTGTAT	113
ASPA	443	rs35614631	TACTTTAAGTTTTAGGGTACATGTGC[-/A]CCATGTGCAGGTTTGTTACATATGT	114
ASPA	443	rs35225782	ATTCATGACCAGCCACATAAATGCAC[-/A]GTATTACTTCGCAAGCATGCCAATG	115
ASPA	443	rs35178659	GTGCACTAGAATTAGCTAAAGTGGGG[-/G]AAAAAAAGATGCATTTGATGGTCTA	116
ASPA	443	rs35095578	AACCTCCACCTCCCAGGTTCAAGAGA[-/A]TTCTCCTGCCTCAGCCTCCCAAGTA	117
ASPA	443	rs35002210	CCTCCCTGTGATCCGAAGTAGCAGAC[A/G]TA CTTAACTTCCATGGTGGATTGTT	118

Petitioner Illumina, Inc., EX1004, p. 428

ASPA	443	rs34744839	AAAACATTATTATATCTAGAAAAAAA[-/A]TGTATCTTAACCATTGTGGGAAGTG	119
ASPA	443	rs34680506	TTGAAGGTAAAATCATAGGGAGTTGG[-/G]AGCTGTCCTCTTGCGCTGAATCAGT	120
ASPA	443	rs34365618	ACTTGTGGCCTTTTTGGAGAGGTTAG[-/CA]ACTCTGAAAACTCTGTCCCTGGACC	121
ASPA	443	rs34275920	GAAGGAGAAAAAGAGAGGAAATAAGT[-/T]AAAATAATAAACACAATTAATAAAG	122
ASPA	443	rs34109510	TGTATACATGTGCCATGTTGGTGTGC[C/T]GC ACCCATTAACTCGTCATTTAGCA	123
ASPA	443	rs34054576	TCACCTGTCACCTCCTATAGAACTTT[-/C]CCCTGACCCTCCTCTATAGCATTAA	124
ASPA	443	rs34015272	ATAAATGATCATCATTCACAGTAGGG[-/G]TTTTGTTTTGTTTTTTTTCTGGAAA	125
ASPA	443	rs34002091	ACAGACATATCTACAAACACACTTTT[-/T]CACATATTTGTGTAAGTCATTTATG	126
ASPA	443	rs28940574	AAAGACAACTAAACTAACGCTCAATG[A/C]A AAAAGTATTCGCTGCTGTTTACAT	127
ASPA	443	rs28940279	TACCGTGTACCCCGTGTTTGTGAATG[A/C]GG CCGCATATTACGAAAAGAAAGAA	128
ASPA	443	rs17850703	CAGGGCTGGAGGTAAAACCATTTATT[A/G]CT AACCCCAGAGCAGTGAAGAAGTG	129
ASPA	443	rs17222495	TTCTTCATTGCCTATTGAAGAGAGAG[C/T]GG AATGCTTTGGTTGCCAGATATGG	130
ASPA	443	rs17175228	CACAAGATCTCATTACTCAGGAGCTG[C/T]CC AAGTGTCTAATGTACTTAGTTAA	131
ASPA	443	rs16953074	TTCTGTGTAACATTTCATTTAAGCAA[A/G]GG ATTCGGCAAATCAAAAATTGTCA	132
ASPA	443	rs16953070	TAAAACGTATTGAAGGTATTATTGAC[G/T]CT GTTGAAGCAAAGAGAACAAAACA	133
HEXA	3073	rs62022858	ATCTGCTCTTCCAGTTGGATGACAAG[C/T]CT TGCTGTCTAACACCTGCTGCAGA	134
HEXA	3073	rs62022857	CCATTTTTTGTTGTATTTTTTTTTTC[C/T]TGAA TACTTTTTATCGCAGTTGGTT	135
HEXA	3073	rs62017872	CCCTGTCTCTAAAAGAAAAAAAAAAA[A/G]A AAAAAAAAAGAAAACAAAACCCAA	136
HEXA	3073	rs62017871	AGTGGCTCCAAAAAGGTCATGGAACC[C/T]CT TGAGGATGATGCAAATTGACTCT	137
HEXA	3073	rs61662730	TAAAGTTACTTTTCTTTTATTGACTT[C/T]CCC TTATTTTTTAACCTTATGCTTT	138
HEXA	3073	rs61329913	CAGAGTTAAAAAAAAAAAAAAAAAAA[-/A]GGAAGTAGCAGCAACAGCTTGGAAA	139
HEXA	3073	rs60920713	GTTGCCCAGGGTTGAGTGCAGAGGCA[C/T]AT CTTGGCTCACAGCAACCTCTGCC	140
HEXA	3073	rs60783213	AAGGCTTTTTTTTTTTTTTTTTTTT[-/TTTT]GAGACAGAGTCTTGCTGTGTCACCC	141
HEXA	3073	rs60644867	GCCTACATTCTGCAAAGAGGAGGGAA[C/G]A TTCACAGCTCCATACTTGAACCCT	142
HEXA	3073	rs60288568	CCAAAGGAGAATAGCTCTAGGGGAGG[C/G]A GGTGGATGAGTATGCATGGGGGAG	143
HEXA	3073	rs59888548	GACTCCATCTCAAAAAAAAAAAAAAA[-/A]TGCAGTCTAATGGCAGAATTAGACT	144

Petitioner Illumina, Inc., EX1004, p. 429

HEXA	3073	rs59733856	TTATTTATTTATTTATTTATTTTTGA[A/G]ACA GGGTCTCTGTTGTCCAGGCTGG	145
HEXA	3073	rs59427837	TTTTGAGGCAGGGTCTCACTCTGTTG[C/T]CC AGGGTTGAGTGCAGAGGCACATC	146
HEXA	3073	rs59171976	CGCCTTGCGAAGGCCCCACAGCTTGC[C/T]TG TGACAAACGTTCATAGGCAAATG	147
HEXA	3073	rs58706602	GGAGGTCTGTACAAAGCACCACCTAC[C/T]TC ATGGGTCAGTTTCCACAGCAGAA	148
HEXA	3073	rs58696963	GAATCTTATAATTCACTGTGTACCTC[-/CCTC]TGTTTCATATTTTCGCAATTGAACT	149
HEXA	3073	rs58610850	AACATAGTATCTAATATAGCTTTACA[C/T]CC AAAGCCAAAATATGAATACACTG	150
HEXA	3073	rs58016062	TTGTTTTGTTTTGTTTGGGGGGGGGG[-/G]TTGTTTTTCTGAGAGGGAGTCTTGC	151
HEXA	3073	rs57733983	CATACCAAAGGGCAGCTGGAGGGATA[C/T]A GACGGAAGTCATGTGGAGAGTGAA	152
HEXA	3073	rs57476645	CAGGTGTGAGCCACCACGACCACCAA[A/T]T TAGCTCTTTTTACTCCTTCCCTTC	153
HEXA	3073	rs56870003	AGTGGTAGCTGATTTTGCTTCTGGAT[A/C]CT TTGCCACCTTCCCACTCTTTAAT	154
HEXA	3073	rs56338339	AAAGACCTGTTTCTTAAAAAAAAAAA[-/AGAAAAAAAAAAA]GAAAGAAAAGAAAAG AAAAAAACAG	155
HEXA	3073	rs55995352	TAAAAAATCTTTCAATGAGGAGATGT[C/T]CC CAGAGCAAGACAGCTGTAGGATG	156
HEXA	3073	rs55860138	AAAAGAAAAAAAAAAAAAAAAAAAA[-/A]GAAAACAAAACCCAAACCCATAAAG	157
HEXA	3073	rs55743646	CCTGTCTCTAAAAGAAAAAAAAAAAA[A/G]A AAAAAAAAGAAAACAAAACCCAAA	158
HEXA	3073	rs55665666	GTTATCATAGAAAAATATCACACTCT[-/GT]CTGTATCCCCACTTCCAGAAACTGT	159
HEXA	3073	rs36106892	CAGGAGCTCATAGAATTACATACAAT[-/C]TTTTTTTTTTTTTTTTGAGACAGCG	160
HEXA	3073	rs36091525	TTGAGAATCTTATAATTCACTGTGTA[-/CCTC]CCTCTGTTTCATATTTTCGCAATTG	161
HEXA	3073	rs35949555	CCACTACCACAGTGCCTAGAGAACAA[C/T]A TGTGTTTAATAATATTTAAATAAT	162
HEXA	3073	rs35827424	CCCTGTCTCTAAAAGAAAAAAAAAAA[-/A]AAAAAAAAAAGAAAACAAAACCCAA	163
HEXA	3073	rs35729578	CCATTATATCATTCATTTCCCACTCA[-/T]TTTCTTCATTCCAACCAAGATATAT	164
HEXA	3073	rs35649102	TCCGTCTCAAAAAAAAAAAAAAAAAG[-/A]GAAAGGAATTATTCTCATGTATACA	165
HEXA	3073	rs35118677	CTGGGGCAGTTAAAAAGAAAAACAAA[-/C]CCCTGGTCCCTGCCCTTGAGGAGAT	166
HEXA	3073	rs35005352	CTCCAGGGTCCCATTCCAGGACCACA[-/C]GCCTGCTACCTCTGCAGCTCACTCA	167
HEXA	3073	rs34736306	GGATTGACATATACCAGTTAGACGGA[-/T]TTTTTTTTTCCATAAACCAGGCTCA	168
HEXA	3073	rs34607939	ACAAATAATTACTACATATCTACAAC[A/G]TT CCAGATACAGAAGAAATGGCCAA	169

Petitioner Illumina, Inc., EX1004, p. 430

HEXA	3073	rs34496117	TAAACACACTTGAAACATCATATAAA[-/ATG]ATATTACTACAAGACTTAACCGTAA	170
HEXA	3073	rs34300017	ACACAGGTAATCCATGTTTATTATAG[-/A]AAAATGCCACATTACTCTTTATTGA	171
HEXA	3073	rs34206496	AGTTATCATAGAAAAATATCACACTC[-/TG]TCTGTATCCCCACTTCCAGAAACTG	172
HEXA	3073	rs34110830	AATGAACTTACAGGAAGGTAATATAT[-/G]GGAAATAAACATCTTATTGAATTTA	173
HEXA	3073	rs34093438	GGACCCCTGAAAGGCACAAGACACCC[-/T]TTCAGGTTCACACTTCCTGAAAGCT	174
HEXA	3073	rs34085965	CCACCAATCACCAGAGCCTTCTGCTC[A/G]GG GGTACCTGAGGGAAAACAAGCAA	175
HEXA	3073	rs34004907	AAAGACTGAAAAAACATTCATAACTA[-/T]TTTTCTTGTTATCCTCGGAAATGTC	176
HEXA	3073	rs28942072	TATCTTCATCTTGGAGGAGATGAGGT[C/T]GA TTTCACCTGCTGGAAGTCCAACC	177
HEXA	3073	rs28942071	TTGCCTATGAACGTTTGTCACACTTC[C/T]GCT GTGAGTTGCTGAGGCGAGGTGT	178
HEXA	3073	rs28941771	GCTTGCTGTTGGATACATCTCGCCAT[C/T]AC CTGCCACTCTCTAGCATCCTGGA	179
HEXA	3073	rs28941770	CCGGGGCTTGCTGTTGGATACATCTC[G/T]CC ATTACCTGCCACTCTCTAGCATC	180

3. Nucleic acid target length evaluation:

In some embodiments, aspects of the invention relate to methods for detecting nucleic acid deletions or insertions in regions containing nucleic acid sequence repeats.

Genomic regions that contain nucleic acid sequence repeats are often the site of genetic instability due to the amplification or contraction of the number of sequence repeats (e.g., the insertion or deletion of one or more units of the repeated sequence). Instability in the length of genomic regions that contain high numbers of repeat sequences has been associated with a number of hereditary and non hereditary diseases and conditions.

For example, "Fragile X syndrome, or Martin-Bell syndrome, is a genetic syndrome which results in a spectrum of characteristic physical, intellectual, emotional and behavioral features which range from severe to mild in manifestation. The syndrome is associated with the expansion of a single trinucleotide gene sequence (CGG) on the X chromosome, and results in a failure to express the FMR-1 protein which is required for normal neural development. There are four generally accepted forms of Fragile X syndrome which relate to the length of the repeated CGG sequence; Normal (29-31 CGG repeats), Premutation (55-200 CGG repeats), Full

Petitioner Illumina, Inc., EX1004, p. 431

Mutation (more than 200 CGG repeats), and Intermediate or Gray Zone Alleles (40 - 60 repeats)."

Other examples include cancer, which has been associated with microsatellite instability (MSI) involving an increase or decrease in the genomic copy number of nucleic acid repeats at one or more microsatellite loci (e.g., BAT-25 and/or BAT-26). The are currently many sequencing-based assays for determining the number of nucleic acid sequence repeats at a particular locus and identifying the presence of nucleic acid insertions or deletions. However, such techniques are not useful in a high throughput multiplex analysis where the entire length of a region may not be sequenced.

In contrast, in some embodiments, aspects of the invention relate to detecting the presence of an insertion or deletion at a genomic locus without requiring the locus to be sequenced (or without requiring the entire locus to be sequenced). Aspects of the invention are particularly useful for detecting an insertion or deletion in a nucleic acid region that contains high levels of sequence repeats. The presence of sequence repeats at a genetic locus is often associated with relatively high levels of polymorphism in a population due to insertions or deletions of one or more of the sequence repeats at the locus. The polymorphisms can be associated with diseases or predisposition to diseases (e.g., certain polymorphic alleles are recessive alleles associated with a disease or condition). However, the presence of sequence repeats often complicates the analysis of a genetic locus and increases the risk of errors when using sequencing techniques to determine the precise sequence and number of repeats at that locus.

In some embodiments, aspects of the invention relate to determining the size of a genetic locus by evaluating the capture frequency of a portion of that locus suspected of containing an insertion or deletion (e.g., due to the presence of sequence repeats) using a nucleic acid capture technique (e.g., a nucleic acid sequence capture technique based on molecular inversion probe technology). According to aspects of the invention, a statistically significant difference in capture efficiency for a genetic locus of interest in different biological samples (e.g., from different subjects) is indicative of different relative lengths in those samples. It should be appreciated that the length differences may be at one or both alleles of the genetic locus. Accordingly, aspects of the invention may be used to identify polymorphisms regardless of whether biological samples being interrogated at heterozygous or homozygous for the

Petitioner Illumina, Inc., EX1004, p. 432

polymorphisms. According to aspects of the invention, subjects that contain one or more loci with an insertion or deletion can be identified by analyzing capture efficiencies for nucleic acids obtained from one or more biological samples using appropriate controls (e.g., capture efficiencies for known nucleic acid sizes, capture efficiencies for other regions that are not suspected of containing an insertion or deletion in the biological sample(s), or predetermined reference capture efficiencies, or any combination thereof. However, it should be appreciated that aspects of the invention are not limited by the nature or presence of the control. In some embodiments, if a statistically significant variation in capture efficiency is detected, a subject may be identified as being at risk for a disease or condition associated with insertions or deletions at that genetic locus. In some embodiments, the subject may be analyzed in greater detail in order to determine the precise nature of the insertion or deletion and whether the subject is heterozygous or homozygous for one or more insertions or deletions. For example, gel electrophoresis of an amplification (e.g., PCR) product of the locus, or Southern blotting, or any combination thereof can be used as an orthogonal approach to verify the length of the locus. In some embodiments, a more exhaustive and detailed sequence analysis of the locus can be performed to identify the number and types of insertions and deletions. However, other techniques may be used to further analyze a locus identified as having an abnormal length according to aspects of the invention.

Accordingly, aspects of the invention relate to detecting abnormal nucleic acid lengths in genomic regions of interest. In some embodiments, the invention aims to estimate the size of genomic regions that are hard to be accessed, such as repetitive elements. However, it should be appreciated that methods of the invention do not require that the precise length be estimated. In some embodiments, it is sufficient to determine that one or more alleles with abnormal lengths are present at a locus of interest (e.g., based on the detection of abnormal capture efficiencies).

In a non-limiting example, fragile X can be used to illustrate aspects of the invention where the size of trinucleotide repeats (genotype) is linked to a symptom (phenotype). However, it should be appreciated that fragile X is a non-limiting example and similar analyses may be performed for other genetic loci (e.g., independently or simultaneously in multiplex analyses).

Use of molecular inversion probes (MIPs) has been demonstrated for detection of single nucleotide polymorphisms (Hardenbol et al. 2005 Genome Res 15:269-75) and for preparative amplification of large sets of exons (Porreca et al. 2007 Nat Methods 4:931-6, Krishnakumar et

Petitioner Illumina, Inc., EX1004, p. 433

al. 2008 Proc Natl Acad Sci USA 105:9296-301). In both cases, oligonucleotide probes are designed which have ends ('targeting arms') that hybridize up-stream and down-stream of the locus that is to be amplified.

In some embodiments, aspects of the invention are based on the recognition that the effect of length on probe capturing efficiency can be used in the context of an assay (e.g., a high throughput and/or multiplex assay) to allow the length of sequences to be determined without requiring sequencing of the entire region being evaluated. This is particularly useful for repeat regions that are prone to changes in size. As illustrated in FIG. 8, which is reproduced from Deng et al., Nature Biotech. 27:353-60, (see Supplemental Fig. 1G of Deng et al.,) illustrates that shorter sequences are captured with higher efficiency that longer sequences using MIPs. The statistical package R and its effects module were used for this analysis. A linear model was used, and each individual factor was assumed to be independent. The dashed lines represent a 95% confidence interval. Shorter target sequences were captured with higher efficiency than long target sequences ($p<2x10^{-16}$). However, the use of this differential capture efficiency for systematic sequence length analysis was not previously recognized.

In some embodiments, following probe hybridization, polymerase fill-in and ligation reactions are performed to convert the hybridized probe to a covalently-closed, circular molecule containing the desired target. PCR or rolling circle amplification plus exonuclease digestion of non-circularized material is performed to isolate and amplify the circular targets from the starting nucleic acid pool. Since one of the main benefits of the method is the potential for a high degree of multiplexing, generally thousands of targets are captured in a single reaction containing thousands of probes.

According to aspects of the invention, repetitive regions are surrounded by non repetitive unique sequences, which can be used to amplify the repeat-containing regions using, for example, PCR or padlock (MIP)-based method.

In addition to the repetitive regions, a probe (e.g., a MIP or padlock probe) can be designed to include at least a sequence that is sufficient to be uniquely identified in the genome (or target pool). After the probe is circularized and amplified, the amplicon can be end-sequenced so that the unique sequence can be identified and served as the "representative" of the repetitive region as illustrated in FIG. 9. FIG. 9 illustrates a non-limiting scheme of padlock (MIP) capture of a region that includes both repetitive regions (thick wavy line) and the adjacent

Petitioner Illumina, Inc., EX1004, p. 434

unique sequence (thick strait line). The regions of the probe are indicated with the targeting arms shown as regions "1" and "3." An intervening region that may be, or include, a sequencing primer binding site is shown as "2." After the padlock is circularized and amplified, it can be end-sequenced to obtain the sequence of the unique sequence, which represents the repetitive region of interest.

Although capturing efficiency is overall negatively correlated with target length, different probe sequences may have unique features. Therefore, multiple probes could be designed and tested so that an optimal one is chosen to be sensitive enough to differentiate repetitive sizes of roughly 0-150bp, 150-600bp, and beyond, which represent normal, premutation and full mutation of fragile X syndrome, respectively. However, it should be appreciated that other probe sizes and sequences can be designed, and optionally optimized, to distinguish a range of repeat region size differences (e.g., length differences of about 3-30 bases, about 30-60 bases, about 60-90 bases, about 90-120 bases, about 120-150 bases, about 150-300 bases, about 300-600 bases, about 600-900 bases, or any intermediate or longer length difference). It should be appreciated that a length difference may be an increase in size or a decrease in size.

In some embodiments, an initial determination of an unexpected capture frequency is indicative of the presence of size difference. In some embodiments, an increase in capture frequency is indicative of a deletion. In some embodiments, a decrease in capture frequency is indicative of an insertion. However, it should be appreciated that depending on specific sequence parameters and the relative sizes of the capture probes, the target region, and the deletions or insertions, a change in capture frequency can be associated with either an increase or decrease in target region length. In some embodiments, the precise nature of the change can be determined using one or more additional techniques as described herein.

Accordingly, in some aspects a MIP probe includes a linear nucleic acid strand that contains two hybridization sequences or targeting arms, one at each end of the linear probe, wherein each of the hybridization sequences is complementary to a separate sequence on a the same strand of a target nucleic acid, and wherein these sequences on the target nucleic acid flank the two ends of the target nucleic acid sequence of interest. It should be appreciated that upon hybridization, the two ends of the probe are inverted with respect to each other in the sense that both 5' and 3' ends of the probe hybridize to the same strand to separate regions flanking the target region (as illustrated in FIG. 9 for example).

Petitioner Illumina, Inc., EX1004, p. 435

In some embodiments, the hybridization sequences are between about 10-100 nucleotides long, for example between about 10-30, about 30-60, about 60-90, or about 20, about 30, about 40, or about 50 nucleotides long. However, other lengths may be used depending on the application. In some embodiments, the hybridization Tms of both targeting arms of a probe are designed or selected to be similar. In some embodiments, the hybridization Tms of the targeting arms of a plurality of probes designed to capture different target regions are selected or designed to be similar so that they can be used together in a multiplex reaction. Accordingly, a typical size of a MIP probe prior to fill in is about 60-80 nucleotides long. However, other sizes can be used depending on the sizes of the targeting arms and any other sequences (e.g., primer binding or tag sequences) that are present in the MIP probe. In some embodiments, MIP probes are designed to avoid sequence-dependent secondary structures. In some embodiments, MIP probes are designed such that the targeting arms do not overlap with known polymorphic regions. In some embodiments, targeting arms that can be used for capturing the repeat region of the Fragile X locus can have the following sequences or complementary to these sequences depending on the strand that is captured.

left: CTCCUITYCGGTITCACTTC (SEQ ID NO: 181)

right: ATCTTCTCTTCAGCCCTGCT (SEQ ID NO: 182)

The typical captured size using these targeting arms is about 100 nucleotides in length (e.g., about 30 repeats of a tri-nucleotide repeat).

In some embodiments, the number of reads obtained for the "representative" of the repetitive region is not informative to estimate the target length because it is dependent on the total number of reads obtained. To overcome this, it is useful to include one or more probes that target other "control" regions where no or minimal polymorphism exists among populations. Because of the systematic consistency of capturing efficiency (see, e.g., FIG. 9), the ratio of reads obtained for the repetitive "representative" to reads obtained for the control region(s) will be tuned using DNA with defined numbers of repeats. Ultimately, the ratio can serve as a measure of the repeat length as illustrated in FIG. 10. FIG. 10 illustrates a non-limiting hypothetical relationship between target gap size and the relative number of reads of the repetitive region, which is measured by the ratio of the repeat "representative" reads vs. the "control" region reads. The unit of y-axis is arbitrary.

Petitioner Illumina, Inc., EX1004, p. 436

In some embodiments, to better tell targets with similar size range apart, the whole repetitive region can be sequenced by making a shotgun library (e.g., by making a shotgun library from a captured sequence, for example a sequence captured using a MIP probe). The longer the repeat is, the more short reads of repeats will be obtained. Therefore, the target length will contribute twice to the relative number of "repetitive" reads, which will gain better resolution of differentiating targets. In some embodiments, the expectation is that the number of reads from any given repeat will be a direct function of the number of repeats present. However, in some embodiments, a Poisson sampling induced spread may need to be considered and in some embodiments may be sufficiently large to limit the resolution.

When a precise measurement of the length of both alleles from a diploid sample is desired, further manipulations may be required. This is because the capture efficiency measured will actually be the average efficiency of the two alleles. To effectively achieve separate measurements for each allele, barcodes (e.g., sequence tags) can be used that allow the efficiency of individual capture events (from individual genomic loci) to be followed. FIG. 11A-C shows the approach. For a given locus, MIPs are synthesized to contain one of a large number differentiator tags in their backbone such that the probability of any two MIPs in a reaction having the same differentiator tag sequence is low. MIP capture is performed on the sample; the reaction will be biased for shorter target lengths, and therefore the reaction product will be comprised of more 'short' circles than 'long' circles. Each circle should bear a unique differentiator tag sequence. Then, linear RCA (1RCA) is performed on the circles. In the 1RCA reaction, circles are converted into long, linear concatemers of themselves. The 1RCA reaction for a given circle stops when the concatemer has reached a 'fixed' length (based on the processivity/error rate of the polymerase). Concatemers derived from smaller circles will therefore contain more copies of the differentiator tag, and concatemers derived from larger circles will contain fewer copies of the differentiator tag. The number of each differentiator tag sequence is counted, for example, by next-generation sequencing.

When number of occurrences is plotted against differentiator tag ID, the data will naturally cluster into two groups reflecting the lengths of the two alleles in the diploid sample. The allele lengths can therefore be read directly off this graph, after absolute length calibration using known standards. In some embodiments, a sequencing technique (e.g., a next-generation sequencing technique) is used to sequence part of one or more captured targets (e.g., or

Petitioner Illumina, Inc., EX1004, p. 437

amplicons thereof) and the sequences are used to count the number of different barcodes that are present. Accordingly, in some embodiments, aspects of the invention relate to a highly-multiplexed qPCR reaction.

Other non-limiting examples of loci at which insertions or deletions or repeat sequences may be associated with a disease or condition are provided in Tables 3 and 4. It should be appreciated that the presence of an abnormal length at any one or more of these loci may be evaluated according to aspects of the invention. In some embodiments, two or more of these loci or other loci may be evaluated in a single multiplex reaction using different probes designed to hybridize under the same reaction conditions to different target nucleic acid in a biological sample.

Table 3.

Polyglutamine (PolyQ) Diseases

Type	Gene	Normal/wildtype	Pathogenic
DRPLA (Dentatorubropallidoluysian atrophy)	ATN1 or DRPLA	6 - 35	49 - 88
HD (Huntington's disease)	HTT (Huntingtin)	10 - 35	35+
SBMA (Spinobulbar muscular atrophy or Kennedy disease)	Androgen receptor on the X chromosome.	9 - 36	38 - 62
SCA1 (Spinocerebellar ataxia Type 1)	ATXN1	6 - 35	49 - 88
SCA2 (Spinocerebellar ataxia Type 2)	ATXN2	14 - 32	33 - 77
SCA3 (Spinocerebellar ataxia Type 3 or Machado-Joseph disease)	ATXN3	12 - 40	55 - 86
SCA6 (Spinocerebellar ataxia Type 6)	CACNA1A	4 - 18	21 - 30
SCA7 (Spinocerebellar ataxia Type 7)	ATXN7	7 - 17	38 - 120
SCA17 (Spinocerebellar ataxia Type 17)	TBP	25 - 42	47 - 63

Table 4.

Petitioner Illumina, Inc., EX1004, p. 438

Non-Polyglutamine Diseases

Type	Gene	Codon	Normal/wildtype	Pathogenic
FRAXA (Fragile X syndrome)	*FMR1*, on the X-chromosome	CGG	6 - 53	230+
FXTAS (Fragile X-associated tremor/ ataxia syndrome)	*FMR1*, on the X-chromosome	CGG	6 - 53	55-200
FRAXE (Fragile XE mental retardation)	AFF2 or FMR2, on the X-chromosome	GCC	6 - 35	200+
FRDA (Friedreich's ataxia)	FXN or X25, (frataxin)	GAA	7 - 34	100+
DM (Myotonic dystrophy)	DMPK	CTG	5 - 37	50+
SCA8 (Spinocerebellar ataxia Type 8)	OSCA or SCA8	CTG	16 - 37	110 – 250
SCA12 (Spinocerebellar ataxia Type 12)	PPP2R2B or SCA12	CAG On 5' end	7 - 28	66 - 78

The following examples illustrate aspects and embodiments of the invention and are not intended to be limiting or restrictive. Many variations of the invention will become apparent to those skilled in the art upon review of this specification. The full scope of the invention should be determined by reference to the claims, along with their full scope of equivalents, and the specification, along with such variations.

4. Increasing detection sensitivity:

In some embodiments, aspects of the invention relate to methods for increasing the sensitivity of nucleic acid detection assays.

There are currently many genomic assays that utilize next-generation (e.g., polony-based) sequencing to generate data, including genome resequencing, RNA-seq for gene expression, bisulphite sequencing for methylation, and Immune-seq, among others. In order to make

Petitioner Illumina, Inc., EX1004, p. 439

quantitative measurements (including genotype calling), these methods utilize the counts of sequencing reads of a given genomic locus as a proxy for the representation of that sequence in the original sample of nucleic acids. The majority of these techniques require a preparative step to construct a high-complexity library of DNA molecules that is representative of a sample of interest. Current assays use one of several alternative nucleic acid preparative techniques (e.g., amplification, for example PCR-based amplification; sequence-specific capture, for example, using immobilized capture probes; or target capture into a circularized probe followed by a sequence analysis step. In order to reduce errors associated with the unpredictability (stochastic nature) of nucleic acid isolation and sequence analysis techniques, current methods involve oversampling a target nucleic acid preparation in order to increase the likelihood that all sequences that are present in the original nucleic acid sample will be represented in the final sequence data. For example, a genomic sequencing library may contain an over- or under-representation of particular sequences from a source nucleic acid sample (e.g., genome preparation) as a result of stochastic variations in the library construction process. Such variations can be particularly problematic when they result in target sequences from a genome being absent or undetectable in a sequencing library. For example, an under-representation of particular allelic sequences (e.g., heterozygotic alleles) from a genome in a sequencing library can result in an apparent homozygous representation in a sequencing library.

In contrast, aspects of the invention relate to basing a nucleic acid sequence analysis on results from two or more different nucleic acid preparatory techniques that have different systematic biases in the types of nucleic acids that they sample rather than simply oversampling the target nucleic acid. According to some embodiments, different techniques have different sequence biases that are systematic and not simply due to stochastic effects during nucleic acid capture or amplification. Accordingly, in some embodiments, the degree of oversampling required to overcome variations in nucleic acid preparation needs to be sufficient to overcome the biases. In some embodiments, the invention provides methods that reduce the need for oversampling by combining nucleic acid and/or sequence results obtained from two or more different nucleic acid preparative techniques that have different biases.

According to the invention, different techniques have different characteristic or systematic biases. For example, one technique may bias a sample analysis towards one particular allele at a genetic locus of interest, whereas a different technique would bias the

Petitioner Illumina, Inc., EX1004, p. 440

sample analysis towards a different allele at the same locus. Accordingly, the same sample may be identified as being different depending on the type of technique that is used to prepare nucleic acid for sequence analysis. This effectively represents a sensitivity issue, because each technique has a different relative sensitivities for polymorphic sequences of interest.

According to aspects of the invention, the sensitivity of a nucleic acid analysis can be increased by combining the sequences from different nucleic acid preparative steps and using the combined sequence information for a diagnostic assay (e.g., for a making a call as to whether a subject is homozygous or heterozygous at a genetic locus of interest).

Currently, the ability of DNA sequencing to detect mutations is limited by the ability of the upstream sample isolation (e.g., by amplification, immobilization enrichment, circularization capture, etc.) methods to reliably isolate the locus of interest.

If one wishes to make heterozygote base-calls for a diploid genome (e.g. a human sample presented for molecular diagnostic sequencing), it is important in some embodiments that the isolation method produces near- or perfectly- uniform amounts of the two alleles to be sequenced (at least sufficiently uniform to be "called" unambiguously as a heterozygote or a homozygote for a locus of interest).

Sample preparative methods may fall into three classes: 1) single- or several target amplification (e.g., uniplex PCR, 'multiplex' PCR), 2) multi-target hybridization enrichment (e.g., Agilent SureSelect 'hybrid capture' [Gnirke et al 2009, *Nature methods* 27:182-9], Roche/Nimblegen 'sequence capture' [Hodges et al 2007, *Nature genetics* 39:1522-7], and 3) multi-target circularization selection (e.g. molecular inversion probes or padlock probes, [Porreca et al 2007, *Nature methods* 4:931-6, Turner et al 2009, *Nature methods* 6:315-6], 'selectors' [Dahl et al 2005, *Nucleic acids research* 33:e71]). Each of these methods can result in a pool of isolated product that does not adequately represent the input abundance distribution. For example, the two alleles at a heterozygous position can become skewed far from their input 50:50 ratio to something that results in a missed basecall during downstream sequencing. For example, if the ratio was skewed from 50:50 to 10:90, and the sample was sequenced to 10X average coverage, there is a high probability that one of the two alleles would not be observed once in the ten sequencing reads. This would reduce the sensitivity of the sequencing method by converting a heterozygous position to homozygous (where potentially the 'mutant' allele was the one not observed). In some embodiments, a skewed ratio is a particular issue that decreases the

Petitioner Illumina, Inc., EX1004, p. 441

sensitivity of detecting mutations present in a heterogeneous tumor tissue. For example, if only 10% of the cells analyzed in a heterogeneous sample harbored a heterozygous mutation, the mutation would be expected to be present in 5% of sequence reads, not 50%. In this scenario, the need for robust, sensitive detection may be even more acute.

The methods disclosed herein are based, in part, on the discovery that certain classes of isolation methods have different modes of bias. The disclosure provide methods for increasing the sensitivity of the downstream sequencing by using a combination of multiple isolation methods (e.g., one or more from at least two of the classes disclosed herein) for a sample. This is particularly important in molecular diagnostics where high sensitivity is required to minimize the chances of 'missing' a disease-associated mutation. For example, given a nominal false-negative error rate of $1x10^{-3}$ for sequencing following circularization selection, and a false-negative error rate of $1x10^{-3}$ for sequencing following hybridization enrichment, one can achieve a final false-negative rate of $1x10^{-6}$ by performing both techniques on the sample (assuming failures in each method are fully independent). For a recessive disease with carrier frequency of 0.1, caused by a single fully-penetrant mutant allele, the number of missed carrier diagnoses would decrease from 1000 per million patients tested to 1 per million patients tested. Furthermore, if the testing was used in the context of prenatal carrier screening, the number of affected children born as a result of missing the carrier call in one parent would decrease from 25 per million to 25 per billion born.

Additionally, the disclosure provides combinations of preparative methods to effectively increase sequencing coverage in regions containing disease-associated alleles. Since heterozygote error rate is largely tied to both deviations from 50:50 allele representation, and in the case of next-generation DNA sequencing deviations from average abundance (such that less abundant isolated targets are more likely to be undersampled at one or both alleles), selectively increasing coverage in these regions will also selectively increase sensitivity. Furthermore, MIPs that detect presence or absence of specific known disease-associated mutations can be used to increase sensitivity selectively. In some embodiments, these MIPs would have a targeting arm whose 3' -most region is complementary to the expected mutation, and has a fill-in length of 0 or more bp. Thus, the MIP will form only if the mutation is present, and its presence will be detected by sequencing.

Petitioner Illumina, Inc., EX1004, p. 442

Additionally, algorithms disclosed herein may be used to determine base identity with varying levels of stringency depending on whether the given position has any known disease-associated alleles. Stringency can be reduced in such positions by decreasing the minimum number of observed mutant reads necessary to make a consensus base-call. This will effectively increase sensitivity for mutant allele detection at the cost of decreased specificity.

An embodiment of the invention combines MIPs plus hybridization enrichment, plus optionally extra MIPs targeted to specific known, common disease-associated loci, e.g., to detect the presence of a polymorphism in a target nucleic acid. A non-limiting example is illustrated in FIG. 12 that illustrates a schematic using MIPs plus hybridization enrichment, plus optionally extra MIPs targeted to specific known, common disease-associated loci, e.g., to detect the presence of a polymorphism in a target nucleic acid.

FIGs. 13 and 14 illustrate different capture efficiencies for MIP-based captures. FIG. 13 shows a graph of per-target abundance with MIP capture. In this graph, bias largely drives the heterozygote error rate, since targets which are less abundant here are less likely to be covered in sufficient depth during sequencing to adequately sample both alleles. This is from Turner et al 2009, Nature methods 6:315-6. Hybridization enrichment results in a qualitatively similar abundance distribution, but the abundance of a given target is likely not correlated between the two methods. FIG. 14 shows a graph of correlation between two MIP capture reactions from Ball et al 2009, Nature biotechnology 27:361-8. Each point represents the target abundance in replicate 1 and replicate 2. Pearson correlation r=0.956. This indicates that MIP capture reproducibly biases targets to specific abundances. Hybridization enrichment is similarly correlated from one capture to the next.

According to aspects of the invention, such biases can be detected or overcome by systematically combining different capture and/or analytical techniques in an assay that interrogates a plurality of loci in a plurality of subject samples.

Accordingly, it should be appreciated that in any of the embodiments described herein (e.g., tiling/staggering, tagging, size-detection, sensitivity enhancing algorithms, or any combination thereof), aspects of the invention involve preparing genomic nucleic acid and/or contacting them with one or more different probes (e.g., capture probes, hybridization probes, MIPs, others etc.). In some embodiments, the amount of genomic nucleic acid used per subject ranges from 1 ng to 10 micrograms (e.g., 500 ng to 5 micrograms). However, higher or lower

Petitioner Illumina, Inc., EX1004, p. 443

amounts (e.g., less than 1 ng, more than 10 micrograms, 10-50 micrograms, 50-100 micrograms or more) may be used. In some embodiments, for each locus of interest, the amount of probe used per assay may be optimized for a particular application. In some embodiments, the ratio (molar ratio, for example measured as a concentration ratio) of probe to genome equivalent (e.g., haploid or diploid genome equivalent, for example for each allele or for both alleles of a nucleic acid target or locus of interest) ranges from 1/100, 1/10, 1/1, 10/1, 100/1, 1000/1. However, lower, higher, or intermediate ratios may be used.

In some embodiments, the amount of target nucleic acid and probe used for each reaction is normalized to avoid any observed differences being caused by differences in concentrations or ratios. In some embodiments, in order to normalize genomic DNA and probe, the genomic DNA concentration is read using a standard spectrophotometer or by fluorescence (e.g., using a fluorescent intercalating dye). The probe concentration may be determined experimentally or using information specified by the probe manufacturer.

Similarly, once a locus has been captured (e.g., on a MIP or other probe or in another form), it may be amplified and/or sequenced in a reaction involving one or more primers. The amount of primer added for each reaction can range from 0.1 pmol to 1 nmol, 0.15 pmol to 1.5 nmol (for example around 1.5 pmol). However, other amounts (e.g., lower, higher, or intermediate amounts) may be used.

In some embodiments, it should be appreciated that one or more intervening sequences (e.g., sequence between the first and second targeting arms on a MIP capture probe), identifier or tag sequences, or other probe sequences that are not designed to hybridize to a target sequence (e.g., a genomic target sequence) should be designed to avoid excessive complementarity (to avoid cross-hybridization) to target sequences or other sequences (e.g., other genomic sequences) that may be in a biological sample. For example, these sequences may be designed have a sufficient number of mismatches with any genomic sequence (e.g., at least 5, 10, 15, or more mismatches out of 30 bases) or as having a Tm (e.g., a mismatch Tm) that is lower (e.g., at least 5, 10, 15, 20, or more degrees C lower) than the hybridization reaction temperature.

It should be appreciated that a targeting arm as used herein may be designed to hybridize (e.g., be complementary) to either strand of a genetic locus of interest if the nucleic acid being analyzed is DNA (e.g., genomic DNA). However, in the context of MIP probes, whichever strand is selected for one targeting arm will be used for the other one. However, in the context of

Petitioner Illumina, Inc., EX1004, p. 444

RNA analysis, it should be appreciated that a targeting arm should be designed to hybridize to the transcribed RNA. It also should be appreciated that MIP probes referred to herein as "capturing" a target sequence are actually capturing it by template-based synthesis rather than by capturing the actual target molecule (other than for example in the initial stage when the arms hybridize to it or in the sense that the target molecule can remain bound to the extended MIP product until it is denatured or otherwise removed).

It should be appreciated that in some embodiments a targeting arm may include a sequence that is complementary to one allele or mutation (e.g., a SNP or other polymorphism, a mutation, etc.) so that the probe will preferentially hybridize (and capture) target nucleic acids having that allele or mutation. However, in many embodiments, each targeting arm is designed to hybridize (e.g., be complementary) to a sequence that is not polymorphic in the subjects of a population that is being evaluated. This allows target sequences to be captured and/or sequenced for all alleles and then the differences between subjects (e.g., calls of heterozygous or homozygous for one or more loci) can be based on the sequence information and/or the frequency as described herein.

It should be appreciated that sequence tags (also referred to as barcodes) may be designed to be unique in that they do not appear at other positions within a probe or a family of probes and they also do not appear within the sequences being targeted. Thus they can be used to uniquely identify (e.g., by sequencing or hybridization properties) particular probes having other characteristics (e.g., for particular subjects and/or for particular loci).

It also should be appreciated that in some embodiments probes or regions of probes or other nucleic acids are described herein as comprising or including certain sequences or sequence characteristics (e.g., length, other properties, etc.). However, it should be appreciated that in some embodiments, any of the probes or regions of probes or other nucleic acids consist of those regions (e.g., arms, central regions, tags, primer sites, etc., or any combination thereof) of consist of those sequences or have sequences with characteristics that consist of one or more characteristics (e.g., length, or other properties, etc.) as described herein in the context of any of the embodiments (e.g., for tiled or staggered probes, tagged probes, length detection, sensitivity enhancing algorithms or any combination thereof).

It should be appreciated that probes, primers, and other nucleic acids designed or used herein may be synthetic, natural, or a combination thereof. Accordingly, as used herein, the term

Petitioner Illumina, Inc., EX1004, p. 445

"nucleic acid" refers to multiple linked nucleotides (i.e., molecules comprising a sugar (e.g., ribose or deoxyribose) linked to an exchangeable organic base, which is either a pyrimidine (e.g., cytosine (C), thymidine (T) or uracil (U)) or a purine (e.g., adenine (A) or guanine (G)). "Nucleic acid" and "nucleic acid molecule" may be used interchangeably and refer to oligoribonucleotides as well as oligodeoxyribonucleotides. The terms shall also include polynucleosides (i.e., a polynucleotide minus a phosphate) and any other organic base containing nucleic acid.

The organic bases include adenine, uracil, guanine, thymine, cytosine and inosine. Unless otherwise stated, nucleic acids may be single or double stranded. The nucleic acid may be naturally or non-naturally occurring. Nucleic acids can be obtained from natural sources, or can be synthesized using a nucleic acid synthesizer (i.e., synthetic).

Harvest and isolation of nucleic acids are routinely performed in the art and suitable methods can be found in standard molecular biology textbooks. (See, for example, Maniatis' Handbook of Molecular Biology.) The nucleic acid may be DNA or RNA, such as genomic DNA, mitochondrial DNA, mRNA, cDNA, rRNA, miRNA, or a combination thereof. Non-naturally occurring nucleic acids such as bacterial artificial chromosomes (BACs) and yeast artificial chromosomes (YACs) can also be used.

The invention also contemplates the use of nucleic acid derivatives. As will be described herein, the use of certain nucleic acid derivatives may increase the stability of the nucleic acids of the invention by preventing their digestion, particularly when they are exposed to biological samples that may contain nucleases. As used herein, a nucleic acid derivative is a non-naturally occurring nucleic acid or a unit thereof. Nucleic acid derivatives may contain non-naturally occurring elements such as non-naturally occurring nucleotides and non-naturally occurring backbone linkages.

Nucleic acid derivatives may contain backbone modifications such as but not limited to phosphorothioate linkages, phosphodiester modified nucleic acids, phosphorothiolate modifications, combinations of phosphodiester and phosphorothioate nucleic acid, methylphosphonate, alkylphosphonates, phosphate esters, alkylphosphonothioates, phosphoramidates, carbamates, carbonates, phosphate triesters, acetamidates, carboxymethyl esters, methylphosphorothioate, phosphorodithioate, p-ethoxy, and combinations thereof. The backbone composition of the nucleic acids may be homogeneous or heterogeneous.

Petitioner Illumina, Inc., EX1004, p. 446

Nucleic acid derivatives may contain substitutions or modifications in the sugars and/or bases. For example, they include nucleic acids having backbone sugars which are covalently attached to low molecular weight organic groups other than a hydroxyl group at the 3' position and other than a phosphate group at the 5' position (e.g., an 2'-0-alkylated ribose group). Nucleic acid derivatives may include non-ribose sugars such as arabinose. Nucleic acid derivatives may contain substituted purines and pyrimidines such as C-5 propyne modified bases, 5-methylcytosine, 2-aminopurine, 2-amino-6-chloropurine, 2,6-diaminopurine, hypoxanthine, 2-thiouracil and pseudoisocytosine. In some embodiments, substitution(s) may include one or more substitutions/modifications in the sugars/bases, groups attached to the base, including biotin, fluorescent groups (fluorescein, cyanine, rhodamine, etc), chemically-reactive groups including carboxyl, NHS, thiol, etc., or any combination thereof.

A nucleic acid may be a peptide nucleic acid (PNA), locked nucleic acid (LNA), DNA, RNA, or co-nucleic acids of the same such as DNA-LNA co-nucleic acids. PNA are DNA analogs having their phosphate backbone replaced with 2-aminoethyl glycine residues linked to nucleotide bases through glycine amino nitrogen and methylenecarbonyl linkers. PNA can bind to both DNA and RNA targets by Watson-Crick base pairing, and in so doing form stronger hybrids than would be possible with DNA or RNA based oligonucleotides in some cases.

PNA are synthesized from monomers connected by a peptide bond (Nielsen, P.E. et al. Peptide Nucleic Acids, Protocols and Applications, Norfolk: Horizon Scientific Press, p. 1-19 (1999)). They can be built with standard solid phase peptide synthesis technology. PNA chemistry and synthesis allows for inclusion of amino acids and polypeptide sequences in the PNA design. For example, lysine residues can be used to introduce positive charges in the PNA backbone. All chemical approaches available for the modifications of amino acid side chains are directly applicable to PNA. Several types of PNA designs exist, and these include single strand PNA (ssPNA), bisPNA and pseudocomplementary PNA (pcPNA).

The structure of PNA/DNA complex depends on the particular PNA and its sequence. ssPNA binds to single stranded DNA (ssDNA) preferably in antiparallel orientation (i.e., with the N-terminus of the ssPNA aligned with the 3' terminus of the ssDNA) and with a Watson-Crick pairing. PNA also can bind to DNA with a Hoogsteen base pairing, and thereby forms triplexes with double stranded DNA (dsDNA) (Wittung, P. et al., *Biochemistry* 36:7973 (1997)).

Petitioner Illumina, Inc., EX1004, p. 447

A locked nucleic acid (LNA) is a modified RNA nucleotide. An LNA form hybrids with DNA, which are at least as stable as PNA/DNA hybrids (Braasch, D.A. et al., *Chem & Biol.* 8(1):1-7(2001)). Therefore, LNA can be used just as PNA molecules would be. LNA binding efficiency can be increased in some embodiments by adding positive charges to it. LNAs have been reported to have increased binding affinity inherently.

Commercial nucleic acid synthesizers and standard phosphoramidite chemistry are used to make LNAs. Therefore, production of mixed LNA/DNA sequences is as simple as that of mixed PNA/peptide sequences. The stabilization effect of LNA monomers is not an additive effect. The monomer influences conformation of sugar rings of neighboring deoxynucleotides shifting them to more stable configurations (Nielsen, P.E. et al. Peptide Nucleic Acids, Protocols and Applications, Norfolk: Horizon Scientific Press, p. 1-19 (1999)). Also, lesser number of LNA residues in the sequence dramatically improves accuracy of the synthesis. Most of biochemical approaches for nucleic acid conjugations are applicable to LNA/DNA constructs.

These and other aspects of the invention are illustrated by the following non-limiting examples.

EXAMPLES

The following examples illustrate non-limiting embodiments of the invention.

Example 1: Design a set of capture probes for a human target exon

All targets are captured as a set of partially-overlapping subtargets. For example, in the tiling approach, a 200 bp target exon might be captured as a set of 12 subtargets, each 60bp in length (FIG. 1). Each subtarget is chosen such that it partially overlaps two or three other targets.

In some embodiments, all probes are composed of three regions: 1) a 20 bp 'targeting arm' comprised of sequence which hybridizes immediately upstream from the sub-target, 2) a 30 bp 'constant region' comprised of sequence used as a pair of amplification priming sites, and 3) a second 20 bp 'targeting arm' comprised of sequence which hybridizes immediately downstream from the sub-target. Targeting arm sequences will be different for each capture probe in a set, while constant region sequence will be the same for all probes in the set, allowing all captured targets to be amplified with a single set of primers. Targeting arm sequences should be designed

Petitioner Illumina, Inc., EX1004, p. 448

such that any given pair of 20 bp sequences is unique in the target genome (to prevent spurious capture of undesired sites). Additionally, melting temperatures should be matched for all probes in the set such that hybridization efficiency is uniform for all probes at a constant temperature (e.g., 60 C). Targeting arm sequences should be computationally screened to ensure they do not form strong secondary structure that would impair their ability to basepair with the genomic target.

Hybridize capture probes to human genomic sample

Assemble hybridization reaction:

- 1.0 ul capture probe mix (~2.5 pmol)
- 2.0 ul 10X Ampligase buffer (Epicentre)
- 6.0 ul 500 ng/ul human genomic DNA (~16.7 fmol)
- 11 ul dH20

In a thermal cycler, heat reaction to 95C for 5 min to denature genomic DNA, then cool to 60C. Allow to incubate at 60C for 40 hours.

Convert hybridized probes into covalently-closed circular products containing subtargets

Prepare fill-in / ligation reaction mixture:

- 0.25 ul 2 mM dNTP mix (Invitrogen)
- 2.5 ul 10X Ampligase buffer (Epicentre)
- 5.0 ul 5 U/ul Taq Stoffel fragment (Applied Biosystems)
- 12.5 ul 5 U/ul Ampligase (Epicentre)
- 4.75 ul dH20

Add 1.0 ul of this mix to the hybridized probe reaction, and incubate at 60C for 10 hours.

Purify circularized probe / subtarget products from un-reacted probes and genomic DNA

Prepare exonuclease reaction mixture:

- 21 ul fill-in / ligation reaction product
- 2.0 ul 10X exonuclease I buffer (New England Biolabs)

Petitioner Illumina, Inc., EX1004, p. 449

- 2.0 ul 20 U/ul exonuclease I (New England Biolabs)
- 2.0 ul 100 U/ul exonuclease III (New England Biolabs)

Incubate at 37C for 60 min, then heat-inactivate by incubating at 80C for 15 min. Immediately cool to 4C for storage.

Amplify circular material by PCR using primers specific to the 'constant region' of the probes

Prepare PCR mixture:

- 5.0 ul 10X Accuprime reaction buffer (Invitrogen)
- 1.5 ul 10 uM CP-2-FA (5'-GCACGATCCGACGGTAGTGT-3') (SEQ ID NO:183)
- 1.5 ul 10 uM CP-2-RA (5'-CCGTAATCGGGAAGCTGAAG-3') (SEQ ID NO:184)
-
- 0.4 ul 25 mM dNTP mix (Invitrogen)
- 2.0 ul heat-inactivated exonuclease reaction mix
- 1.5 ul 10X SybrGreen (Invitrogen)
- 0.4 ul 2.5 U/ul Accuprime Pfx polymerase (Invitrogen)
- 37.7 ul dH20

Thermal cycle in real-time thermal cycler according to the following protocol, but stop cycling before amplification yield plateaus (generally 8-12 cycles):

1. 95C for 5 min
2. 95C for 30 sec
3. 58C for 60 sec
4. 72C for 60 sec
5. goto 2, N more times

Prepare a shotgun next-generation sequencing library for analysis

- Purify desired amplicon population from non-specific amplification products by gel extraction.

- Concatemerize amplicons into high-molecular weight products suitable for shearing
- Mechanically shear, using either a nebulizer, BioRuptor, Hydroshear, Covaris, or similar instrument. DNA should be sheared into fragments several hundred basepairs in length.
- Ligate adapters required for amplification by the sequencing platform used. If necessary, purify ligated product from unligated product and adapters.

Example 2: Use of differentiator tag sequences to detect and correct bias in a MIP-capture reaction of a set of exon targets

The first step in performing the detection/correction is to determine how many differentiator tag sequences are necessary for the given sample. In this example, 1000 genomic targets corresponding to 1000 exons were captured. Since the differentiator tag sequence is part of the probe, it will measure/report biases that occur from the earliest protocol steps. Also, being located in the backbone, the differentiator tag sequence can easily be sequenced from a separate priming site, and therefore not impact the total achievable read-length for the target sequence. MIP probes are synthesized using standard column-based oligonucleotide synthesis by any number of vendors (e.g. IDT), and differentiator tag sequences are introduced as 'degenerate' positions in the backbone. Each degenerate position increases the total number of differentiator tag sequences synthesized by a factor of 4, so a 10 nt degenerate region implies a differentiator tag sequence complexity of ~1e6 species.

Hybridize capture probes to human genomic sample

Assemble hybridization reaction:

- 1.0 ul capture probe mix (~2.5 pmol)
- 2.0 ul 10X Ampligase buffer (Epicentre)
- 6.0 ul 500 ng/ul human genomic DNA (~16.7 fmol)
- 11 ul dH20

In a thermal cycler, heat reaction to 95C for 5 min to denature genomic DNA, then cool to 60C. Allow to incubate at 60C for 40 hours.

Convert hybridized probes into covalently-closed circular products containing subtargets

Petitioner Illumina, Inc., EX1004, p. 451

Prepare fill-in / ligation reaction mixture:

- 0.25 ul 2 mM dNTP mix (Invitrogen)
- 2.5 ul 10X Ampligase buffer (Epicentre)
- 5.0 ul 5 U/ul Taq Stoffel fragment (Applied Biosystems)
- 12.5 ul 5 U/ul Ampligase (Epicentre)
- 4.75 ul dH20

Add 1.0 ul of this mix to the hybridized probe reaction, and incubate at 60C for 10 hours.

Purify circularized probe / subtarget products from un-reacted probes and genomic DNA

Prepare exonuclease reaction mixture:

- 21 ul fill-in / ligation reaction product
- 2.0 ul 10X exonuclease I buffer (New England Biolabs)
- 2.0 ul 20 U/ul exonuclease I (New England Biolabs)
- 2.0 ul 100 U/ul exonuclease III (New England Biolabs)

Incubate at 37C for 60 min, then heat-inactivate by incubating at 80C for 15 min. Immediately cool to 4C for storage.

Amplify circular material by PCR using primers specific to the 'constant region' of the probes

Prepare PCR mixture:

- 5.0 ul 10X Accuprime reaction buffer (Invitrogen)
- 1.5 ul 10 uM CP-2-FA (5'-GCACGATCCGACGGTAGTGT-3') (SEQ ID NO: 183)
-
- 1.5 ul 10 uM CP-2-RA (5'-CCGTAATCGGGAAGCTGAAG-3') (SEQ ID NO: 184)
-
- 0.4 ul 25 mM dNTP mix (Invitrogen)
- 2.0 ul heat-inactivated exonuclease reaction mix
- 1.5 ul 10X SybrGreen (Invitrogen)
- 0.4 ul 2.5 U/ul Accuprime Pfx polymerase (Invitrogen)

Petitioner Illumina, Inc., EX1004, p. 452

- 37.7 ul dH20

Thermal cycle in real-time thermal cycler according to the following protocol, but stop cycling before amplification yield plateaus (generally 8-12 cycles):

6. 95C for 5 min
7. 95C for 30 sec
8. 58C for 60 sec
9. 72C for 60 sec
10. goto 2, N more times

Prepare a shotgun next-generation sequencing library for analysis

- Purify desired amplicon population from non-specific amplification products by gel extraction.
- Concatemerize amplicons into high-molecular weight products suitable for shearing
- Mechanically shear, using either a nebulizer, BioRuptor, Hydroshear,
- Covaris, or similar instrument. DNA should be sheared into fragments several hundred basepairs in length.
- Ligate adapters required for amplification by the sequencing platform used. If necessary, purify ligated product from unligated product and adapters.

Perform sequencing of library according to manufacturer's directions (e.g. Illumina,ABI, etc), reading both the target sequence and the differentiator tag sequence.

Analyze data by correcting for any biases detected by quantitation of differentiator tag sequence abundance

Construct a table of target: differentiator tag abundances from the read data, e.g.:

Target ID	Differentiator Tag sequence ID	Count
1	3547	1
2	4762	1
1	9637	1
1	1078	5
3	4762	1
1	2984	1

Petitioner Illumina, Inc., EX1004, p. 453

All 'count' entries should be '1', since any particular target:differentiator tag mapping will not occur more than once by chance, and therefore will only be observed if bias was present somewhere in the sample preparation process. For any target:differentiator tag combination observed more than once, all such reads are 'collapsed' into a single read before consensus basecalls are determined. This will cancel the effect of bias on consensus basecall accuracy. FIG. 5 depicts a method for making diploid genotype calls in which repeat target:differentiator tag combination are collapsed.

Example 3: Differentiator tag sequence design for MIP capture reactions

For a set of targets, the number of differentiator tag sequences necessary to be confident (within some statistical bounds) that a certain differentiator tag sequence will not be observed more than once by chance in combination with a certain target sequence was determined. The total number of unique differentiator tag sequences for a certain differentiator tag sequence length is determined as $4^{(\text{Length in nucleotides of the differentiator tag sequence})}$. For a molecular inversion probe capture reaction that uses MIP probes having differentiator tag sequences, the probability of performing the capture reaction and capturing one or more copies of a target sequence having the same differentiator tag sequence is calculated as: $p = 1 - [N! / (N-M)!] / [NAM]$, wherein N is the total number of possible unique differentiator tag sequences and M is the number of target sequence copies in the capture reaction. Thus, by varying the differentiator tag sequence length it is possible to perform a MIP capture reaction in which the probability of capturing one or more copies of a target sequence having the same differentiator tag sequence is set at a predetermined probability value.

For example, for a differentiator tag sequence of 15 nucleotides in length, there are 1,073,741,824 possible differentiator tag sequences. A MIP capture reaction in which MIP probes, each having a differentiator tag sequence of 15 nucleotides, are combined with 10000 target sequence copies (e.g., genome equivalents), the probability of capturing one or more copies of a target sequence having the same differentiator tag sequence is 0.05. In this example, the MIP reaction will produce very few (usually 0, but occasionally 1 or more) targets where multiple copies are tagged with the same differentiator tag sequence. FIG. 6 depicts results of a simulation for 100000 capture reactions having 15 nucleotide differentiator tag sequences and 10000 target sequences.

Petitioner Illumina, Inc., EX1004, p. 454

Example 4: Assessment of the probability for obtaining enough sequencing reads to make accurate base-calls at multiple independent loci, as a function of sequencing coverage.

Monte Carlo simulations were performed to determine sequencing coverage requirements. The simulations assume 10000 genomic copies of a given locus (target) half mom alleles and half dad alleles. The simulations further assume 1% efficiency of capture for the MIP reaction. The simulation samples from a capture mix 100 times without replacement to create a set of 100 capture products. The simulation then samples from the set of 100 capture products with replacement (assuming unbiased amplification) to generate 'reads' from either mom or dad. The number of reads sampled depends on the coverage. The number of independent reads from both mom and dad necessary to make a high-quality base-call (assumed to be 10 or 20 reads) were then determined. The process was repeated 1000 times for each coverage level, and the fraction of times that enough reads from both parents were successfully obtained was determined. This fraction was raised to the power 1000, assuming we have 1000 independent loci that must obtain successful base-calls, plotted (See FIG. 7). Result show that roughly 50x coverage is required to capture each allele >= 10x with > 0.95 probability.

Example 5: MIP capture of 'target' locus and 'control' loci

In some embodiments, to accurately quantify the efficiency of target locus capture, at least three sets of control loci are captured in parallel that have a priori been shown to serve as proxies for various lengths of target locus. For example, if the target locus is expected to have a length between 50 and 1000 bp, then sets of control loci having lengths of 50, 250, and 1000 bp could be captured (e.g. 20 loci per set should provide adequate protection from outliers), and their abundance digitally measured by sequencing. These loci should be chosen such that minimal variation in efficiency between samples and on multiple runs of the same sample is observed (and are therefore 'efficiency invariant'). These will serve as 'reference' points that define the shape of the curve of abundance-vs-length. Determining the length of the target is then simply a matter of 'reading' the length from the appropriate point on the calibration curve.

In some embodiments, the statistical confidence one has in the estimate of target length from this method is driven largely by three factors: 1) reproducibility / variation of the abundance data used to generate the calibration curve; 2) goodness of fit of the regression to the 'control' datapoints; 3) reproducibility of abundance data for the target locus being measured. Statistical bounds on 1) and 2) will be known in advance, having been measured during

Petitioner Illumina, Inc., EX1004, p. 455

development of the assay. Additionally, statistical bounds on 3) will be known in general in advance, since assay development should include adequate population sampling and measure of technical reproducibility. Standard statistical methods should be used to combine these three measures into a single P value for any given experimental measure of target abundance.

In some embodiments, given the set of calibration observations, and a linear regression fit to that data, the regression can be used to predict the length value for n observations of the target locus whose length is unknown. First, choose an acceptable range for the confidence interval of the length estimate. For example, in the case of distinguishing "normal" (87-93bp) from "premutation" (165-600bp) potential cases of Fragile X, the goal is to measure length to sufficient precision to distinguish 93 bp from 165 bp. The predicted response value, computed when n observations is substituted into the equation for the regressed line, will have arbitrary precision. However, if for example a 95% confidence level is desired, that 95% confidence interval must be sufficiently short that it does not overlap both the "normal" and "premutation" length ranges. Continuing the example, if one calculates a length of 190 from n=400 MIP observations, and based on the regression from calibration data, the 95% confidence interval is 190 +/- 20 bp, one can conclude the sample represents a "premutation" length with 95% certainty. Conversely, if the calibration data were less robust, error estimates of the regression would be higher, leading to larger confidence intervals on the predicted response value. In some embodiments, if the 95% CI were calculated as 190 +/- 100 bp from n=400, one could not determine whether the predicted response value corresponds to a "normal" or "premutation" length.

In some embodiments, the confidence interval for a predicted response is calculated as:

The estimate for the response $\hat{y}$ *is identical to the estimate for the mean of the response:* $\hat{y} = b_0 + b_1 x^*$. *The confidence interval for the predicted value is given by* $\hat{y} \pm t^* s_{\hat{y}}$, *where* $\hat{y}$ *is the fitted value corresponding to* x^*. *The value* t^* *is the upper (1 - C)/2 critical value for the t(n - 2) distribution.*

In some embodiments, a technique for analyzing a locus of interest can involve the following steps.

Petitioner Illumina, Inc., EX1004, p. 456

Convert hybridized probes into covalently-closed circular products containing subtargets

Prepare fill-in / ligation reaction mixture:

- 0.25 ul 2 mM dNTP mix (Invitrogen)
- 2.5 ul 10X Ampligase buffer (Epicentre)
- 5.0 ul 5 U/ul Taq Stoffel fragment (Applied Biosystems)
- 12.5 ul 5 U/ul Ampligase (Epicentre)
- 4.75 ul dH20

Add 1.0 ul of this mix to the hybridized probe reaction, and incubate at 60C for 10 hours.

Purify circularized probe / subtarget products from un-reacted probes and 30 genomic DNA

Prepare exonuclease reaction mixture:

- 21 ul fill-in / ligation reaction product
- 2.0 ul 10X exonuclease I buffer (New England Biolabs)
- 2.0 ul 20 U/ul exonuclease I (New England Biolabs)
- 2.0 ul 100 U/ul exonuclease III (New England Biolabs)

Incubate at 37C for 60 min, then heat-inactivate by incubating at 80C for 15 min. Immediately cool to 4C for storage.

Amplify circular material by PCR using primers specific to the 'constant region' of the probes

Prepare PCR mixture:

- 5.0 ul 10X Accuprime reaction buffer (Invitrogen)
- 1.5 ul 10 uM CP-2-FA-Ilmn (platform-specific amplification sequence plus 'circle constant region'-specific sequence)
- 1.5 ul 10 uM CP-2-RA-Ilmn (platform-specific amplification sequence plus 'circle constant region'-specific sequence)
- 0.4 ul 25 mM dNTP mix (Invitrogen)
- 2.0 ul heat-inactivated exonuclease reaction mix
- 1.5 ul 10X SybrGreen (Invitrogen)
- 0.4 ul 2.5 U/ul Accuprime Pfx polymerase (Invitrogen)

Petitioner Illumina, Inc., EX1004, p. 457

- 37.7 ul dH20

Thermal cycle in real-time thermal cycler according to the following protocol, but stop cycling before amplification yield plateaus (generally 8-12 cycles):

11. 95C for 5 min
12. 95C for 30 sec
13. 58C for 60 sec
14. 72C for 60 sec
15. goto 2, N more times

Perform sequencing (e.g., next-generation sequencing) on sample for digital quantitation according to manufacturer's instructions (e.g., Illumina, ABI)

Example 6: MIP-capture reaction of a set of exon target nucleic acids

MIP probes are synthesized using standard column-based oligonucleotide synthesis by any number of vendors (e.g. IDT).

Hybridize capture probes to human genomic sample

Assemble hybridization reaction:

- 1.0 ul capture probe mix (-2.5 pmol)
- 2.0 ul 10X Ampligase buffer (Epicentre)
- 6.0 ul 500 ng/ul human genomic DNA (-16.7 fmol)
- 11 ul dH20

In a thermal cycler, heat reaction to 95C for 5 min to denature genomic DNA, then cool to 60C. Allow to incubate at 60C for 40 hours.

Convert hybridized probes into covalently-closed circular products containing target nucleic acids

Prepare fill-in / ligation reaction mixture:

- 0.25 ul 2 mM dNTP mix (Invitrogen)
- 2.5 ul 10X Ampligase buffer (Epicentre)
- 5.0 ul 5 U/ul Taq Stoffel fragment (Applied Biosystems)
- 12.5 ul 5 U/ul Ampligase (Epicentre)
- 4.75 ul dH20

Add 1.0 ul of this mix to the hybridized probe reaction, and incubate at 60C for 10 hours.

Purify, circularized probe/target nucleic acid products from un-reacted probes and genomic DNA

Prepare exonuclease reaction mixture:

- 21 ul fill-in / ligation reaction product
- 2.0 ul 10X exonuclease I buffer (New England Biolabs)
- 2.0 ul 20 U/ul exonuclease I (New England Biolabs)
- 2.0 ul 100 U/ul exonuclease III (New England Biolabs)

Incubate at 37C for 60 min, then heat-inactivate by incubating at 80C for 15 min. Immediately cool to 4C for storage.

Amplify circular material by PCR using primers specific to the 'constant region' of the probes

Prepare PCR mixture:

- 5.0 ul 10X Accuprime reaction buffer (Invitrogen)
- 1.5 ul 10 uM CP-2-FA (5'-GCACGATCCGACGGTAGTGT-3') (SEQ ID NO: 183)
-
- 1.5 ul 10 uM CP-2-RA (5'-CCGTAATCGGGAAGCTGAAG-3') (SEQ ID NO: 184)
- 0.4 ul 25 mM dNTP mix (Invitrogen)
- 2.0 ul heat-inactivated exonuclease reaction mix
- 1.5 ul 10X SybrGreen (Invitrogen)
- 0.4 ul 2.5 U/ul Accuprime Pfx polymerase (Invitrogen)
- 37.7 ul dH20

Thermal cycle in real-time thermal cycler according to the following protocol, but stop cycling before amplification yield plateaus (generally 8-12 cycles):

16. 95C for 5 min
17. 95C for 30 sec
18. 58C for 60 sec
19. 72C for 60 sec
20. goto 2, N more times

Petitioner Illumina, Inc., EX1004, p. 459

Prepare a shotgun next-generation sequencing library for analysis

- Purify desired amplicon population from non-specific amplification products by gel extraction.
- Concatemerize amplicons into high-molecular weight products suitable for shearing
- Mechanically shear, using either a nebulizer, BioRuptor, Hydroshear, Covaris, or similar instrument. DNA should be sheared into fragments several hundred basepairs in length.
- Ligate adapters required for amplification by the sequencing platform used. If necessary, purify ligated product from unligated product and adapters.

Perform sequencing of library according to manufacturer's directions (e.g. Illumina, ABI, etc), reading the target sequence to determine abundance of the target nucleic acid.

Example 7: Use of MIPs, hybridization, and mutation-detection MIPs to genotype a set of 1000 targets

MIPs, hybridization, and mutation-detection MIPs are used to genotype a set of 1000 targets. The protocol permits detection of any of 50 specific known point mutations

First, separate MIP, hybridization, and mutation-detection MIP reactions are performed on a biological sample. A MIP capture reaction is performed essentially as described in Turner et al 2009, Nature methods 6:315-6. A set of MIPs is designed such that each probe in the set flanks one of the 1000 targets. Separately, a hybridization enrichment reaction is performed using the Agilent SureSelect procedure. Prior to selection, the genomic DNA to be enriched is converted into a shotgun sequencing library using Illumina's 'Fragment Library' kit and protocol. Agilent's web interface is used to design a set of probes which will hybridize to the target nucleic acids. Separately, a set of probes are designed (mutation-detection MIPs) which will form MIPs only if mutations (e.g., specific polymorphisms) are present. Each mutation-detection MIP has a 3'-most base identity that is specific for a single known mutation. A reaction with this set of mutation-detection MIPs is performed to selectively detect the presence of any mutant alleles.

Once all three reactions have been performed, the two MIP reactions are combined (e.g., at potentially non-equimolar ratios to further increase sensitivity of mutation detection) into a

Petitioner Illumina, Inc., EX1004, p. 460

single tube, and run as one sample on the next-generation DNA sequencing instrument. The hybridization-enriched reaction is run as a separate sample on the next-generation DNA sequencing instrument. Reads from each 'sample' are combined by a software algorithm which forms a consensus diploid genotype at each position in the target set by evaluating the total coverage at each position, the origin of each read in that total coverage, the quality score of each individual read, and the presence (or absence) of any reads derived from mutation-specific MIPs overlapping the region.

It should be appreciated that the preceding examples are non-limiting and aspects of the invention may be implemented as described herein using alternative techniques and/or protocols that are available to one or ordinary skill in the art.

It will be clear that the methods may be practiced other than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present disclosure are possible in light of the above teachings and, therefore, are within the scope of the claims. Preferred features of each aspect of the Disclosure are as for each of the other aspects mutatis mutandis. The documents including patents, patent applications, journal articles, or other disclosures mentioned herein are hereby incorporated by reference in their entirety. In the event of conflict, the disclosure of present application controls, other than in the event of clear error.

Petitioner Illumina, Inc., EX1004, p. 461

METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS

RELATED APPLICATIONS

This application is a continuation of U.S. Application No. 15/231,687, filed August 8, 2016, which is a continuation of U.S. Application No. 13/266,862, filed March 13, 2012, which is a national phase application that claims the benefit of, and priority to, international PCT Application No. PCT/US2010/001293, filed April 30, 2010, which claims the benefit under 35 U.S.C. §119(e) of U.S. Provisional Application No. 61/174,470, filed April 30, 2009, U.S. Provisional Application No. 61/178,923, filed May 15, 2009, U.S. Provisional Application No. 61/179,358, filed May 18, 2009, and U.S. Provisional Application No. 61/182,089, filed May 28, 2009, the entire contents of each of which are incorporated herein by reference.

FIELD OF INVENTION

The invention relates to methods and compositions for determining genotypes in patient samples.

BACKGROUND OF THE INVENTION

Information about the genotype of a subject is becoming more important and relevant for a range of healthcare decisions as the genetic basis for many diseases, disorders, and physiological characteristics is further elucidated. Medical advice is increasingly personalized, with individual decisions and recommendations being based on specific genetic information. Information about the type and number of alleles at one or more genetic loci impacts disease risk, prognosis, therapeutic options, and genetic counseling amongst other healthcare considerations.

For cost-effective and reliable medical and reproductive counseling on a large scale, it is important to be able correctly and unambiguously identify the allelic status for many different genetic loci in many subjects.

Numerous technologies have been developed for detecting and analyzing nucleic acid sequences from biological samples. These technologies can be used to genotype subjects and determine the allelic status of any locus of interest. However, they are not sufficiently robust and

Petitioner Illumina, Inc., EX1004, p. 462

cost-effective to be scaled up for reliable high throughput analysis of many genetic loci in large numbers of patients. The frequency of incorrect or ambiguous calls is too high for current technology to manage large numbers of patient samples without involving expensive and time-consuming steps to resolve uncertainties and provide confidence in the information output.

SUMMARY OF THE INVENTION

Aspects of the invention relate to preparative and analytical methods and compositions for evaluating genotypes, and in particular, for determining the allelic identity (or identities in a diploid organism) of one or more genetic loci in a subject.

Aspects of the invention are based, in part, on the identification of different sources of ambiguity and error in genetic analyses, and, in part, on the identification of one or more approaches to avoid, reduce, recognize, and/or resolve these errors and ambiguities at different stages in a genetic analysis.

According to aspects of the invention, certain types of genetic information can be under-represented or over-represented in a genetic analysis due to a combination of stochastic variation and systematic bias in any of the preparative stages (e.g., capture, amplification, etc.), determining stages (e.g., allele-specific detection, sequencing, etc.), data interpretation stages (e.g., determining whether the assay information is sufficient to identify a subject as homozygous or heterozygous), and/or other stages.

According to aspects of the invention, error or ambiguity may be apparent in a genetic analysis, but not readily resolved without running additional samples or more expensive assays (*e.g.*, array-based assays may report no-calls due to noisy/low signal). According to further aspects of the invention, error or ambiguity may not be accounted for in a genetic analysis and incorrect base calls may be made even when the evidence for them is limited and/or not statistically significant (*e.g.*, next-generation sequencing technologies may report base calls even if the evidence for them is not statistically significant). According to further aspects of the invention error or ambiguity may be problematic for a multi-step genetic analysis because it is apparent but not readily resolved in one or more steps of the analysis and not apparent or accounted for in other steps of the analysis.

In some embodiments, sources of error and ambiguity in one or more steps can be addressed by capturing and/or interrogating each target locus of interest with one or more sets of

Petitioner Illumina, Inc., EX1004, p. 463

overlapping probes that are designed to overcome any systematic bias or stochastic effects that may impact the complexity and/or fidelity of the genetic information that is generated.

In some embodiments, sources of error and ambiguity in one or more steps can be addressed by capturing and/or interrogating each target locus of interest with at least one set of probes, wherein different probes are labeled with different identifiers that can be used to track the assay reactions and determine whether certain types of genetic information are under-represented or over-represented in the information that is generated.

In some embodiments, errors and ambiguities associated with the analysis of regions containing large numbers of sequence repeats are addressed by systematically analyzing frequencies of certain nucleic acids at particular stages in an assay (e.g., at a capture, sequencing, or detection stage). It should be appreciated that such techniques may be particularly useful in the context of a standardized protocol that is designed to allow many different loci to be evaluated in parallel without requiring different assay procedures for each locus. In some embodiments, the use of a single detection modality (*e.g.*, sequencing) to assay multiple types of genetic lesions (*e.g.*, point mutations, insertions/deletions, length polymorphisms) is advantageous in the clinical setting. In some embodiments of the invention, methods are provided that facilitate the use of multiple sample preparation steps in parallel, coupled with multiple analytical processes following sequence detection. Thus, in some embodiments of the invention, an improved workflow is provided that reduces error and uncertainty when simultaneously assaying different types of genetic lesions across multiple loci in multiple patients.

In some embodiments, aspects of the invention provide methods for overcoming preparative and/or analytical bias by combining two or more techniques, each having a different bias (e.g., a known bias towards under-representation or over-representation of one or more types of sequences), and using the resulting data to determine a genetic call for a subject with greater confidence.

It should be appreciated that in some embodiments, aspects of the invention relate to multiplex diagnostic methods. In some embodiments, multiplex diagnostic methods comprise capturing a plurality of genetic loci in parallel (e.g., one or more genetic loci from Table 1). In some embodiments, the genetic loci possess one or more polymorphisms (e.g., one or more polymorphisms from Table 2) the genotypes of which correspond to disease causing alleles.

Petitioner Illumina, Inc., EX1004, p. 464

Accordingly, in some embodiments, the disclosure provides methods for assessing multiple heritable disorders in parallel. In some embodiments, methods are provided for diagnosing multiple heritable disorders in parallel at a pre-implantation, prenatal, perinatal, or postnatal stage. In some embodiments, the disclosure provides methods for analyzing multiple genetic loci (e.g., a plurality of target nucleic acids selected from Table 1) from a patient sample, such as a blood, pre-implantation embryo, chorionic villus or amniotic fluid sample, or other sample (e.g., other biological fluid or tissue sample such as a biopsy sample) as aspects of the invention are not limited in this respect.

Other samples may include tumor tissue or circulating tumor cells. In some embodiments, a patient sample (*e.g.*, a tumor tissue or cell sample) is mosaic for one or more mutations of interest, and thus, may require higher sensitivity than is needed for a germline mutation analysis. In some embodiments, a sample comprises cells from a non-host organism (*e.g.*, bacterial or viral infections in a human subject) or a sample for environmental monitoring (*e.g.*, bacterial, viral, fungal composition of a soil, water, or air sample).

Accordingly, in some embodiments, aspects of the methods disclosed herein relate to genotyping a polymorphism of a target nucleic acid. In some embodiments, the genotyping may comprise determining that one or more alleles of the target nucleic acid are heterozygous or homozygous. In further embodiments, the genotyping may comprise determining the sequence of a polymorphism and comparing that sequence to a control sequence that is indicative of a disease risk. In some embodiments, the polymorphism is selected from a locus in Table 1 or Table 2. However, it should be appreciated that any locus associated with a disease or condition of interest may be used.

In some embodiments, a diagnosis, prognosis, or disease risk assessment is provided to a subject based on a genotype determined for that subject at one or more genetic loci (e.g., based on the analysis of a biological sample obtained from that subject). In some embodiments, an assessment is provided to a couple, based on their respective genotypes at one or more genetic loci, of the risk of their having one or more children having a genotype associated with a disease or condition (e.g., a homozygous or heterozygous genotype associated with a disease or condition). In some embodiments, a subject or a couple may seek genetic or reproductive counseling in connection with a genotype determined according to embodiments of the

Petitioner Illumina, Inc., EX1004, p. 465

invention. In some embodiments, genetic information from a tumor or circulating tumor cells is used to determine prognosis and guide selection of appropriate drugs/treatments.

It should be appreciated that any of the methods or compositions described herein may be used in combination with any of the medical evaluations associated with one or more genetic loci as described herein.

In some embodiments, aspects of the invention provide effective methods for overcoming challenges associated with systematic errors (bias) and/or stochastic effects in multiplex genomic capture and/or analysis (including sequencing analysis). In some embodiments, aspects of the invention are useful to avoid, reduce and/or account for variability in one or more sampling and/or analytical steps. For example, in some embodiments, variability in target nucleic acid representation and unequal sampling of heterozygous alleles in pools of captured target nucleic acids can be overcome.

Accordingly, in some embodiments, the disclosure provides methods that reduce variability in the detection of target nucleic acids in multiplex capture methods. In other embodiments, methods improve allelic representation in a capture pool and, thus, improve variant detection outcomes. In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve the use of different sets of multiple probes (e.g., molecular inversion probes MIPs) that capture overlapping regions of a target nucleic acid to achieve a more uniform representation of the target nucleic acids in a capture pool compared with methods of the prior art. In other embodiments, methods reduce bias, or the risk of bias, associated with large scale parallel capture of genetic loci, e.g., for diagnostic purposes. In other embodiments, methods are provided for increasing reproducibility (e.g., by reducing the effect of polymorphisms on target nucleic acid capture) in the detection of a plurality of genetic loci in parallel. In further embodiments, methods are provided for reducing the effect of probe synthesis and/or probe amplification variability on the analysis of a plurality of genetic loci in parallel.

According to some aspects, methods of analyzing a plurality of genetic loci are provided. In some embodiments, the methods comprise contacting each of a plurality of target nucleic acids with a probe set, wherein each probe set comprises a plurality of different probes, each probe having a central region flanked by a 5' region and a 3' region that are complementary to nucleic acids flanking the same strand of one of a plurality of subregions of the target nucleic

Petitioner Illumina, Inc., EX1004, p. 466

acid, wherein the subregions of the target nucleic acid are different, and wherein each subregion overlaps with at least one other subregion, isolating a plurality of nucleic acids each having a nucleic acid sequence of a different subregion for each of the plurality of target nucleic acids, and analyzing the isolated nucleic acids.

In other embodiments, methods comprise contacting each of a plurality of target nucleic acids with a probe set, wherein each probe set comprises a plurality of different probes, each probe having a central region flanked by a 5' region and a 3' region that are complementary to nucleic acids flanking the same strand of one of a plurality of subregions of the target nucleic acid, wherein the subregions of the target nucleic acid are different, and wherein a portion of the 5' region and a portion of the 3' region of a probe have, respectively, the sequence of the 5' region and the sequence of the 3' region of a different probe, isolating a plurality of nucleic acids each having a nucleic acid sequence of a different subregion for each of the plurality of target nucleic acids, and analyzing the isolated nucleic acids.

Aspects of the disclosure are based, in part, on the discovery of methods for overcoming problems associated with systematic and random errors (bias) in genome capture, amplification and sequencing methods, namely high variability in the capture and amplification of nucleic acids and disproportionate representation of heterozygous alleles in sequencing libraries. Accordingly, in some embodiments, the disclosure provides methods that reduce errors associated with the variability in the capture and amplification of nucleic acids. In other embodiments, the methods improve allelic representation in sequencing libraries and, thus, improve variant detection outcomes. In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve the use of differentiator tag sequences to uniquely tag individual nucleic acid molecules. In some embodiments, the differentiator tag sequence permit the detection of bias based on the occurrence of combinations of differentiator tag and target sequences observed in a sequencing reaction. In other embodiments, the methods reduce errors caused by bias, or the risk of bias, associated with the capture, amplification and sequencing of genetic loci, e.g., for diagnostic purposes.

Aspects of the invention relate to providing sequence tags (referred to as differentiator tags) that are useful to determine whether target nucleic acid sequences identified in an assay are from independently isolated target nucleic acids or from multiple copies of the same target

Petitioner Illumina, Inc., EX1004, p. 467

nucleic acid molecule (e.g., due to bias in a preparative step, for example, amplification). This information can be used to help analyze a threshold number of independently isolated target nucleic acids from a biological sample in order to obtain sequence information that is reliable and can be used to make a genotype conclusion (e.g., call) with a desired degree of confidence. This information also can be used to detect bias in one or more nucleic acid preparative steps.

In some embodiments, the methods disclosed herein are useful for any application where reduction of bias, e.g., associated with genomic isolation, amplification, sequencing, is important. For example, detection of cancer mutations in a heterogeneous tissue sample, detection of mutations in maternally-circulating fetal DNA, and detection of mutations in cells isolated during a preimplantation genetic diagnostic procedure.

Accordingly, in some aspects, methods of genotyping a subject are provided. In some embodiments, the methods comprise determining the sequence of at least a threshold number of independently isolated nucleic acids, wherein the sequence of each isolated nucleic acid comprises a target nucleic acid sequence and a differentiator tag sequence, wherein the threshold number is a number of unique combinations of target nucleic acid and differentiator tag sequences, wherein the isolated nucleic acids are identified as independently isolated if they comprise unique combinations of target nucleic acid and differentiator tag sequences, and wherein the target nucleic acid sequence is the sequence of a genomic locus of a subject.

In some embodiments, the isolated nucleic acids are products of a circularization selection-based preparative method, e.g., molecular inversion probe capture products. In other embodiments, the isolated nucleic acids are products of an amplification-based preparative methods. In other embodiments, the isolated nucleic acids are products of hybridization-based preparative methods.

Circularization selection-based preparative methods selectively convert regions of interest (target nucleic acids) into a covalently-closed circular molecule which is then isolated typically by removal (usually enzymatic, e.g. with exonuclease) of any non-circularized linear nucleic acid. Oligonucleotide probes (e.g., molecular inversion probes) are designed which have ends that flank the region of interest (target nucleic acid) and, optionally, primer sites, e.g., sequencing primer sites. The probes are allowed to hybridize to the genomic target, and enzymes are used to first (optionally) fill in any gap between probe ends and second ligate the probe closed. Following circularization, any remaining (non-target) linear nucleic acid is

Petitioner Illumina, Inc., EX1004, p. 468

typically removed, resulting in isolation (capture) of target nucleic acid. Circularization selection-based preparative methods include molecular inversion probe capture reactions and 'selector' capture reactions. In some embodiments, molecular inversion probe capture of a target nucleic acid is indicative of the presence of a polymorphism in the target nucleic acid.

In amplification-based (e.g., PCR-based or LCR-based, etc.) preparative methods, genomic loci (target nucleic acids) are isolated directly by means of a polymerase chain reaction or ligase chain reaction (or other amplification method) that selectively amplifies each locus using one or more oligonucleotide primers. It is to be understood that primers will be sufficiently complementary to the target sequence to hybridize with and prime amplification of the target nucleic acid. Any one of a variety of art known methods may be utilized for primer design and synthesis. One or more of the primers may be perfectly complementary to the target sequence. Degenerate primers may also be used. Primers may also include additional nucleic acids that are not complementary to target sequences but that facilitate downstream applications, including for example restriction sites and differentiator tag sequences. Amplification-based methods include amplification of a single target nucleic acid and multiplex amplification (amplification of multiple target nucleic acids in parallel).

Hybridization-based preparative methods involve selectively immobilizing target nucleic acids for further manipulation. It is to be understood that one or more oligonucleotides (immobilization oligonucleotides), which comprise differentiator tag sequences, and which may be from 15 to 170 nucleotides in length, are used which hybridize along the length of a target region of a genetic locus to immobilize it. In some embodiments, immobilization oligonucleotides, are either immobilized before hybridization is performed (e.g., Roche/Nimblegen 'sequence capture'), or are prepared such that they include a moiety (e.g. biotin) which can be used to selectively immobilize the target nucleic acid after hybridization by binding to e.g., streptavidin-coated microbeads (e.g. Agilent 'SureSelect').

It should be appreciated that any of the circularization, amplification, and/or hybridization based methods described herein may be used in connection with one or more of the tiling/staggering, tagging, size-detection, and/or sensitivity enhancing algorithms described herein.

In some embodiments, the methods disclosed herein comprise determining the sequence of molecular inversion probe capture products, each comprising a molecular inversion probe and

Petitioner Illumina, Inc., EX1004, p. 469

a target nucleic acid, wherein the sequence of the molecular inversion probe comprises a differentiator tag sequence and, optionally, a primer sequence, and wherein the target nucleic acid is a captured genomic locus of a subject, and genotyping the subject at the captured genomic locus based on the sequence of at least a threshold number of unique combinations of target nucleic acid and differentiator tag sequences of molecular inversion probe capture products.

In some embodiments, the methods disclosed herein comprise obtaining molecular inversion probe capture products, each comprising a molecular inversion probe and a target nucleic acid, wherein the sequence of the molecular inversion probe comprises a differentiator tag sequence and, optionally, a primer sequence, wherein the target nucleic acid is a captured genomic locus of the subject, amplifying the molecular inversion probe capture products, and genotyping the subject by determining, for each target nucleic acid, the sequence of at least a threshold number of unique combinations of target nucleic acid and differentiator tag sequence of molecular inversion probe capture products. In certain embodiments, obtaining comprises capturing target nucleic acids from a genomic sample of the subject with molecular inversion probes, each comprising a unique differentiator tag sequence. In specific embodiments, capturing is performed under conditions wherein the likelihood of obtaining two or more molecular inversion probe capture products with identical combinations of target and differentiator tag sequences is equal to or less than a predetermined value, optionally wherein the predetermined value is about 0.05.

In one embodiment, the threshold number for a specific target nucleic acid sequence is selected based on a desired statistical confidence for the genotype. In some embodiments, the methods further comprising determining a statistical confidence for the genotype based on the number of unique combinations of target nucleic acid and differentiator tag sequences.

According to some aspects, methods of analyzing a plurality of genetic loci are provided. In some embodiments, the methods comprise obtaining a plurality of molecular inversion probe capture products each comprising a molecular inversion probe and a target nucleic acid, wherein the sequence of the molecular inversion probe comprises a differentiator tag sequence and, optionally, a primer sequence (e.g., a sequence that is complementary to the sequence of a nucleic acid that is used as a primer for sequencing or other extension reaction), amplifying the plurality of molecular inversion probe capture products, determining numbers of occurrence of combinations of target nucleic acid and differentiator tag sequence of molecular inversion probe

Petitioner Illumina, Inc., EX1004, p. 470

capture products in the amplified plurality, and if the number of occurrence of a specific combination of target nucleic acid sequence and differentiator tag sequence exceeds a predetermined value, detecting bias in the amplification of the molecular inversion probe comprising the specific combination. In some embodiments, the methods further comprise genotyping target sequences in the plurality, wherein the genotyping comprises correcting for bias, if detected.

In some embodiments, the target nucleic acid is a gene (or portion thereof) selected from Table 1. In some embodiments, the genotyping comprises determining the sequence of a target nucleic acid (e.g., a polymorphic sequence) at one or more (both) alleles of a genome (a diploid genome) of a subject. In certain embodiments, the genotyping comprises determining the sequence of a target nucleic acid at both alleles of a diploid genome of a subject, wherein in the target nucleic acid comprises, or consists of, a sequence of Table 1, Table 2, or other locus of interest.

In some embodiments, aspects of the invention provide methods and compositions for identifying nucleic acid insertions or deletions in genomic regions of interest without determining the nucleotide sequences of these regions. Aspects of the invention are particularly useful for detecting nucleic acid insertions or deletions in genomic regions containing nucleic acid sequence repeats (e.g., di- or tri-nucleotide repeats). However, the invention is not limited to analyzing nucleic acid repeats and may be used to detect insertions or deletions in any target nucleic acid of interest. Aspects of the invention are particularly useful for analyzing multiple loci in a multiplex assay.

In some embodiments, aspects of the invention relate to determining whether an amount of target nucleic acid that is captured in a genomic capture assay is higher or lower than expected. In some embodiments, a statistically significant deviation from an expected amount (e.g., higher or lower) is indicative of the presence of a nucleic acid insertion or deletion in the genomic region of interest. In some embodiments, the amount is a number of nucleic acid molecules that are captured. In some embodiments, the amount is a number of independently captured nucleic acid molecules in a sample. It should be appreciated that the captured nucleic acids may be literally captured from a sample, or their sequences may be captured without actually capturing the original nucleic acids in the sample. For example, nucleic acid sequences

Petitioner Illumina, Inc., EX1004, p. 471

may be captured in an assay that involves a template-based extension of nucleic acids having the region of interest, in the sample.

Aspects of the invention are based on the recognition that the efficiency of certain capture techniques is affected by the length of the nucleic acid being captured.

Accordingly, an increase or decrease in the length of a target nucleic acid (e.g., due to an insertion or deletion of a repeated sequence) can alter the capture efficiency of that nucleic acid. In some embodiments, a difference in the capture efficiency (e.g., a statistically significant difference in the capture efficiency) of a target nucleic acid is indicative of an insertion or deletion in the target nucleic acid. It should be appreciated that the capture efficiency for a target nucleic acid may be evaluated based on an amount of captured nucleic acid (e.g., number of captured nucleic acid molecules) relative to a control amount (e.g., based on an amount of control nucleic acid that is captured). However, the invention is not limited in this respect and other techniques for evaluating capture efficiency also may be used.

According to aspects of the invention, evaluating the capture efficiency as opposed to determining the sequence of the entire repeat region reduces errors associated with sequencing through repeat regions. Repeat sequences often give rise to stutters or skips in sequencing reactions that make it very difficult to accurately determine the number of repeats in a target region without running multiple sequencing reactions under different conditions and carefully analyzing the results. Such procedures are cumbersome and not readily scalable in a manner that is consistent with high throughput analyses of target nucleic acids. In some embodiments, repeat regions may be longer than the length of the individual sequence read, making length determination on the basis of a single read impossible. For example, when using next-generation sequencing the repeat regions may be longer than the length of the individual sequence read, making length determination on the basis of a single read impossible. Accordingly, aspects of the invention are useful to increase the sensitivity of detecting insertions or deletions in target regions, particularly target regions containing repeated sequences.

In some embodiments, aspects of the invention relate to capturing genomic nucleic acid sequences using a molecular inversion probe (e.g., MIP or Padlock probe) technique, and determining whether the amount (e.g., number) of captured sequences is higher or lower than expected. In some embodiments, the amount (e.g., number) of captured sequences is compared to an amount (e.g., number) of sequences captured in a control assay. The control assay may

Petitioner Illumina, Inc., EX1004, p. 472

involve analyzing a control sample that contains a nucleic acid from the same genetic locus having a known sequence length (e.g., a known number of nucleic acid repeats). However, a control may involve analyzing a second (e.g., different) genetic locus that is not expected to contain any insertions or deletions. The second genetic locus may be analyzed in the same sample as the locus being interrogated or in a different sample where its length has been previously determined. The second genetic locus may be a locus that is not characterized by the presence of nucleic acid repeats (and thus not expected to contain insertions or deletions of the repeat sequence).

In some embodiments, a target nucleic acid region that is being evaluated may be determined by the identity of the targeting arms of a probe that is designed to capture the target region (or sequence thereof). For example, the targeting arms of a MIP probe may be designed to be complementary (e.g., sufficiently complementary for selective hybridization and//or polymerase extension and/or ligation) to genomic regions flanking a target region suspected of containing an insertion or deletion. It should be appreciated that two targeting arms may be designed to be complementary (e.g., sufficiently complementary for selective hybridization and/or polymerase extension and/or ligation) to the two flanking regions that are immediately adjacent (e.g., immediately 5' and 3', respectively) to a region of a sequence repeat on one strand of a genomic nucleic acid. However, one or both targeting arms may be designed to hybridize several bases (e.g., 1-5, 5-10, 10-25, 25-50, or more) upstream or downstream from the repeat region in such a way that the captured sequence includes a region of unique genomic sequence that on one or both sides of the repeat region. This unique region can then be used to identify the captured target (e.g., based on sequence or hybridization information).

In some embodiments, two or more (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10 or more) different loci may be interrogated in parallel in a single assay (e.g., in a multiplex assay). In some embodiments, the ratio of captured nucleic acids for each locus may be used to determine whether a nucleic acid insertion or deletion is present in one locus relative to the other. For example, the ratio may be compared to a control ratio that is representative of the two loci when neither one has an insertion or deletion relative to control sequences (e.g., sequences that are normal or known to be associated with healthy phenotypes for those loci). However, the amount of captured nucleic acids may be compared to any suitable control as discussed herein.

Petitioner Illumina, Inc., EX1004, p. 473

The locus of a captured sequence may be identified by determining a portion of unique sequence 5' and/or 3' to the repeat region in the target nucleic acid suspected of containing a deletion or insertion. This does not require sequencing the captured repeat region itself. However, some or all of the repeat region also could be sequenced as aspects of the invention are not limited in this respect.

Aspects of the invention may be combined with one or more sequence-based assays (e.g., SNP detection assays), for example in a multiplex format, to determine the genotype of one or more regions of a subject.

In some embodiments, methods of detecting a polymorphism in a nucleic acid in a biological sample are provided. In some embodiments, the methods comprise evaluating the efficiency of capture at one or more loci and determining whether one or both alleles at that locus contain an insertion or deletion relative to a control locus (e.g., a locus indicative of a length of repeat sequence that is associated with a healthy phenotype).

Accordingly, aspects of the invention relate to methods for determining whether a target nucleic acid has an abnormal length by evaluating the capture efficiency of a target nucleic acid in a biological sample from a subject, wherein a capture efficiency that is different from a reference capture efficiency is indicative of the presence, in the biological sample, of a target nucleic acid having an abnormal length. It should be appreciated that the term "abnormal" is a relative term based on a comparison to a "normal" length. In some embodiments, a normal length is a length that is associated with a normal (e.g., healthy or non-carrier phenotype). Accordingly, an abnormal length is a length that is either shorter or longer than the normal length. In some embodiments, the presence of an abnormal length is indicative of an increased risk that the locus is associated with a disease or a disease carrier phenotype. In some embodiments, the abnormal length is indicative that the subject is either has a disease or condition or is a carrier of a disease or condition (e.g., associated with the locus). However, it should be appreciated that the description of embodiments relating to detecting the presence of an abnormal length also support detecting the presence of a length that is different from an expected or control length.

In some embodiments, aspects of the invention relate to estimating the length of a target nucleic acid (e.g., of a sub-target region within a target nucleic acid). In some embodiments, aspects of the invention relate to methods for estimating the length of a target nucleic acid by

Petitioner Illumina, Inc., EX1004, p. 474

contacting the target nucleic acid with a plurality of detection probes under conditions that permit hybridization of the detection probes to the target nucleic acid, wherein each detection probe is a polynucleotide that comprises a first arm that hybridizes to a first region of the target nucleic acid and a second arm that hybridizes to a second region of the target nucleic acid, wherein the first and second regions are on a common strand of the target nucleic acid, and wherein the nucleotide sequence of the target between the 5' end of the first region and the 3' end of the second region is the nucleotide sequence of a sub-target nucleic acid; and capturing a plurality of sub-target nucleic acids that are hybridized with the plurality of detection probes; and measuring the frequency of occurrence of a sub-target nucleic acid in the plurality of sub-target nucleic acids, wherein the frequency of occurrence of the sub-target nucleic acid in the plurality of sub-target nucleic acids is indicative of the length of the sub-target nucleic acid. It should be appreciated that methods for estimating a nucleic acid length may involve comparing a capture efficiency for a target nucleic acid region to two or more reference efficiencies for known nucleic acid lengths in order to determine whether the target nucleic acid region is smaller, intermediate, or larger in size than the known control lengths. In some embodiments, a series of nucleic acids of known different lengths may be used to provide a calibration curve for evaluating the length of a target nucleic acid region of interest.

In some embodiments, the capture efficiency of a target region suspected of having a deletion or insertion is determined by comparing the capture efficiency to a reference indicative of a normal capture efficiency. In some embodiments, the capture efficiency is lower than the reference capture efficiency. In some embodiments, the subject is identified as having an insertion in the target region. In some embodiments, the capture efficiency is higher than the reference capture efficiency. In some embodiments, the subject is identified as having a deletion in the target region. In some embodiments, the subject is identified as being heterozygous for the insertion. In some embodiments, the subject is identified as being heterozygous for the deletion.

In some embodiments of any of the methods described herein (e.g., tiling/staggering, tagging, size-detection, and/or sensitivity enhancement) aspects of the invention relate to capturing a sub-target nucleic acid (or a sequence of a sub-target nucleic acid). In some embodiments, a molecular inversion probe technique is used. In some embodiments, a molecular inversion probe is a single linear strand of nucleic acid that comprises a first targeting arm at its 5' end and a second targeting arm at its 3' end, wherein the first targeting arm is capable of

Petitioner Illumina, Inc., EX1004, p. 475

specifically hybridizing to a first region flanking one end of the sub-target nucleic acid, and wherein the second targeting arm is capable of specifically hybridizing to a second region flanking the other end of the sub target nucleic acid on the same strand of the target nucleic acid. In some embodiments, the first and second targeting arms are between about 10 and about 100 nucleotides long. In some embodiments, the first and second targeting arms are about 10-20, 20-30, 30-40, or 40-50 nucleotides long. In some embodiments, the first and second targeting arms are about 20 nucleotides long. In some embodiments, the first and second targeting arms have the same length. In some embodiments, the first and second targeting arms have different lengths. In some embodiments, each pair of first and second targeting arms in a set of probes has the same length. Accordingly, if one of the targeting arms is longer, the other one is correspondingly shorter. This allows for a quality control step in some embodiments to confirm that all captured probe/target sequence products have the same length after a multiplexed plurality of capture reactions. In some embodiments, a set of probes may be designed to have the same length if the intervening region is varied to accommodate any differences in the length of either one or both of the first and second targeting arms.

In some embodiments, the hybridization Tms of the first and second targeting arms are similar. In some embodiments, the hybridization Tms of the first and second targeting arms are within 2-5 °C of each other. In some embodiments, the hybridization Tms of the first and second targeting arms are identical. In some embodiments, the hybridization Tms of the first and second targeting arms are close to empirically-determined optima but not necessarily identical.

In some embodiments, the first and second targeting arms of a molecular inversion probe have different Tms. For example, the Tm of the first targeting arm (at the 5' end of the molecular inversion probe) may be higher than the Tm of the second targeting arm (at the 3' end of the molecular inversion probe). According to aspects of the invention, and without wishing to be bound by theory, a relatively high Tm for the first targeting arm may help avoid or prevent the first targeting arm from being displaced after hybridization by the extension product of the 3' end of the second targeting arm. It should be appreciated that a reference to the Tm of a targeting arm as used herein relates to the Tm of hybridization of the targeting arm to a nucleic acid having the complementary sequence (e.g., the region of the target nucleic acid that has a sequence that is complementary to the sequence of the targeting arm). It also should be appreciated that the Tms of the targeting arms described herein may be calculated using any

Petitioner Illumina, Inc., EX1004, p. 476

appropriate method. For example, in some embodiments an experimental method (e.g., a gel shift assay, a hybridization assay, a melting curve analysis, for example in a PCR machine with a SYBR dye by stepping through a temperature ramp while monitoring signal level from an intercalating dye, for example, bound to a double-stranded DNA, etc.) may be used to determine one or more Tms empirically. In some embodiments, an optimal Tm may be determined by evaluating the number of products formed (e.g., for each of a plurality of MIP probes), and determining the optimal Tm as the center point in a histogram of Tm for all targeting arms. In some embodiments, a predictive algorithm may be used to determine a Tm theoretically. In some embodiments, a relatively simple predictive algorithm may be used based on the number of G/C and A/T base pairs when the sequence is hybridized to its target and/or the length of the hybridized product (e.g., for example, 64.9 + 41 * ([G+C] - 16.4) / (A + T + G + C), see for example, Wallace,R.B., Shaffer,J., Murphy,R.F., Bonner,J., Hirose,T., and Itakura,K. (1979) Nucleic Acids Res 6:3543-3557). In some embodiments, a more complex algorithm may be used to account for the effects of base stacking entropy and enthalpy, ion concentration, and primer concentration (see, for example, SantaLucia J (1998), Proc Natl Acad Sci USA, 95:1460-5). In some embodiments an algorithm may use modified parameters (e.g., nearest-neighbor parameters for basepair entropy/enthalpy values). It should be appreciated that any suitable algorithm may be used as aspects of the invention are not limited in this respect. However, it also should be appreciated that different methodologies may results in different calculated or predicted Tms for the same sequences. Accordingly, in some embodiments, the same empirical and/or theoretical method is used to determine the Tms of different sequences for a set of probes to avoid a negative impact of any systematic difference in the Tm determination or prediction when designing a set of probes with predetermined similarities or differences for different Tms.

In some embodiments, the Tm of the first targeting arm may be about 1 °C, about 2 °C, about 3 °C, about 4 °C, about 5 °C, or more than about 5 °C higher than the Tm of the second targeting arm. In some embodiments, each probe in a plurality of probes (e.g., each probe in a set of 5-10, each probe in a set of at least 10, each probe in a set of 10-50, each probe in a set of 50-100, each probe in a set of 100-500, each probe in a set of 500-1,000, each probe in a set of 1,000-1,500, each probe in a set of 1,500-2,000, each probe in a set of 2,000-3,000, 3,000-5,000, 5,000-10,000 or each probe in a set of at least 5,000 different probes) has a unique first targeting arm (e.g., they all have different sequences) and a unique second targeting arm (e.g., they all

Petitioner Illumina, Inc., EX1004, p. 477

have different sequences). In some embodiments, for at least 10% of the probes (e.g., at least 25%, 25%-50%, 50%- 75%, 75%-90%, 90%-95% or over 95%, or all of the probes) the first targeting arm has a Tm for its complementary sequence that is higher (e.g., about 1 °C, about 2 °C, about 3 °C, about 4 °C, about 5 °C, or more than about 5 °C higher) than the Tm of the second targeting arm for its complementary sequence. In some embodiments, each of the first targeting arms have similar or identical Tms for their respective complementary sequences and each of the second targeting arms have similar or identical Tms for their respective complementary sequences (and the first targeting arms have higher Tms than the second targeting arms). For example, in some embodiments, the Tm of the first arm(s) may be about 58 °C and the Tm of the second arm(s) may be about 56 °C. In some embodiments, the Tm of the first arm(s) may be about 68 °C, and the Tm of the second arm(s) may be about 65 °C. It should be appreciated that in some embodiments the similarity (e.g., within a range of 1 °C, 2 °C, 3 °C, 4 °C, 5 °C) or identity of the Tms for the different targeting arms should be based either on empirical data for each arm or based on the same predictive algorithm for each arm (e.g., Wallace,R.B., Shaffer,J., Murphy,R.F., Bonner,J., Hirose,T., and Itakura,K. (1979) Nucleic Acids Res 6:3543-3557, SantaLucia J (1998), Proc Natl Acad Sci USA, 95:1460-5, or other algorithm).

In some embodiments, the Tm of the first targeting arm of a molecular inversion probe (at the 5' end of the molecular inversion probe) is selected to be sufficiently stable to prevent displacement of the first targeting arm from its complementary sequence on a target nucleic acid. In some embodiments, the Tm of the first targeting arm is 50-55 °C, at least 55 °C, 55-60 °C, at least 60 °C, 60-65 °C, at least 65 °C, at least 70 °C, at least 75 °C, or at least 80 °C. As discussed above, it should be appreciated that the for a particular targeting arm may be determined empirically or theoretically. Different theoretical models may be used to determine a Tm and it should be appreciated that the predicted Tm for a particular sequence may be different depending on the algorithm used for the prediction. In some embodiments, each probe in a plurality of probes (e.g., each probe in a set of 5-10, each probe in a set of at least 10, each probe in a set of 10-50, each probe in a set of 50-100, each probe in a set of 100-500, or each probe in a set of at least 500 different probes) has a different first targeting arm (e.g., different sequences) but each different first targeting arm has a similar or identical Tm for its complementary sequence on a target nucleic acid. It should be appreciated that in some embodiments the

Petitioner Illumina, Inc., EX1004, p. 478

similarity (e.g., within a range of I C, 2 C, 3 C, 4 C, 5 C) or identity of the Tms for the different targeting arms should be based either on empirical data for each arm or based on the same predictive algorithm for each arm (e.g., Wallace,R.B., Shaffer,J., Murphy,R.F., Bonner,J., Hirose,T., and Itakura,K. (1979) Nucleic Acids Res 6:3543- 3557, SantaLucia J (1998), Proc Natl Acad Sci USA, 95:1460-5, or other algorithm).

In some embodiments, the sub-target nucleic acid contains a nucleic acid repeat. In some embodiments, the nucleic acid repeat is a dinucleotide or trinucleotide repeat. In some embodiments, the sub-target nucleic acid contains 10-100 copies of the nucleic acid repeat in the absence of an abnormal increase or decrease in nucleic acid repeats. In some embodiments, the sub-target nucleic acid is a region of the Fragile-X locus that contains a nucleic acid repeat. In some embodiments, one or both targeting arms hybridize to a region on the target nucleic acid that is immediately adjacent to a region of nucleic acid repeats. In some embodiments, one or both targeting arms hybridize to a region on the target nucleic acid that is separated from a region of nucleic acid repeats by a region that does not contain any nucleic acid repeats. In some embodiments, the molecular inversion probe further comprises a primer-binding region that can be used to sequence the captured sub-target nucleic acid and optionally the first and/or second targeting arm.

In some embodiments, aspects of the invention relate to evaluating the length of a plurality of different target nucleic acids in a biological sample. In some embodiments, the plurality of target nucleic acids are analyzed using a plurality of different molecular inversion probes. In some embodiments, each different molecular inversion probe comprises a different pair of first and second targeting arms at each of the 3' and 5' ends. In some embodiments, each different molecular inversion probe comprises the same primer-binding sequence.

In some embodiments, aspects of the invention relate to analyzing nucleic acid from a biological sample obtained from a subject. In some embodiments, the biological sample is a blood sample. In some embodiments, the biological sample is a tissue sample, specific cell population, tumor sample, circulating tumor cells, or environmental sample. In some embodiments, the biological sample is a single cell. In some embodiments, nucleic acids are analyzed in biological samples obtained from a plurality of different subjects. In some embodiments, nucleic acids from a biological sample are analyzed in multiplex reactions. It should be appreciated that a biological sample contains a plurality of copies of a genome derived

Petitioner Illumina, Inc., EX1004, p. 479

from a plurality of cells in the sample. Accordingly, a sample may contain a plurality of independent copies of a target nucleic acid region of interest, the capture efficiency of which can be used to evaluate its size as described herein.

In some embodiments, aspects of the invention relate to evaluating a nucleic acid capture efficiency by determining an amount of target nucleic acid that is captured (e.g., an amount of sub-target nucleic acid sequences that are captured). In some embodiments, the amount of target nucleic acid that is captured is determined by determining a number of independently captured target nucleic acid molecules (e.g., the amount of independently captured molecules that have the sequence of the sub-target region). In some embodiments, the amount of target nucleic acid that is captured is compared to a reference amount of captured nucleic acid. In some embodiments, the reference amount is determined by determining a number of independently captured molecules of a reference nucleic acid. In some embodiments, the reference nucleic acid is a nucleic acid of a different locus in the biological sample that is not suspected of containing a deletion or insertion. In some embodiments, the reference nucleic acid is a nucleic acid of known size and amount that is added to the capture reaction. As described herein, a number of independently captured nucleic acid sequences can be determined by contacting a nucleic acid sample with a preparation of a probe (e.g., a MIP probe as described herein). It should be appreciated that the preparation may comprise a plurality of copies of the same probe and accordingly a plurality of independent copies of the target region may be captured by different probe molecules. The number of probe molecules that actually capture a sequence can be evaluated by determining an amount or number of captured molecules using any suitable technique. This number is a reflection of both the number of target molecules in the sample and the efficiency of capture of those target molecules, which in turn is related to the size of the target molecules as described herein. Accordingly, the capture efficiency can be evaluated by controlling for the abundance of the target nucleic acid, for example by comparing the number or amount of captured target molecules to an appropriate control (e.g., a known size and amount of control nucleic acid, or a different locus that should be present in the same amount in the biological sample and is not expected to contain any insertions or deletions). It should be appreciated that other factors may affect the capture efficiency of a particular target nucleic acid region (e.g., the sequence of the region, the GC content, the presence of secondary structures, etc.). However, these factors also can be accounted for by using appropriate controls (e.g.,

Petitioner Illumina, Inc., EX1004, p. 480

known sequences having similar properties, the same sequences, other genomic sequences expected to be present in the biological sample at the same frequency, etc., or any combination thereof).

In some embodiments, aspects of the invention relate to identifying a subject as having an insertion or deletion in one or more alleles of a genetic locus if the capture efficiency for that genetic locus is statistically significantly different than a reference capture efficiency.

It should be appreciated that hybridization conditions used for any of the capture techniques described herein (e.g., MIP capture techniques) can be based on known hybridization buffers and conditions.

In some embodiments, the methods disclosed herein are useful for any application where the detection of deletions or insertions is important.

In some embodiments, aspects of the invention relate to basing a nucleic acid sequence analysis on results from two or more different nucleic acid preparatory techniques that have different systematic biases in the types of nucleic acids that they sample. According to the invention, different techniques have different sequence biases that are systematic and not simply due to stochastic effects during nucleic acid capture or amplification. Accordingly, the degree of oversampling required to overcome variations in nucleic acid preparation needs to be sufficient to overcome the biases (e.g., an oversampling of 2-5 fold, 5-10 fold, 5-15 fold, 15-20 fold, 20-30 fold, 30-50 fold, or intermediate to higher fold).

According to some embodiments, different techniques have different characteristic or systematic biases. For example, one technique may bias a sample analysis towards one particular allele at a genetic locus of interest, whereas a different technique would bias the sample analysis towards a different allele at the same locus. Accordingly, the same sample may be identified as being different depending on the type of technique that is used to prepare nucleic acid for sequence analysis. This effectively represents a sensitivity limitation, because each technique has different relative sensitivities for polymorphic sequences of interest.

According to aspects of the invention, the sensitivity of a nucleic acid analysis can be increased by combining the sequences from different nucleic acid preparative steps and using the combined sequence information for a diagnostic assay (e.g., for a making a call as to whether a subject is homozygous or heterozygous at a genetic locus of interest).

Petitioner Illumina, Inc., EX1004, p. 481

In some embodiments, the invention provides a method of increasing the sensitivity of a nucleic acid detection assay by obtaining a first preparation of a target nucleic acid using a first preparative method on a biological sample, obtaining a second preparation of a target nucleic acid using a second preparative method on the biological sample, assaying the sequences obtained in both first and second nucleic acid preparations, and using the sequence information from both first and second nucleic acid preparations to determine the genotype of the target nucleic acid in the biological sample, wherein the first and second preparative methods have different systematic sequence biases. In some embodiments, the first and second nucleic acid preparations are combined prior to performing a sequence assay. In some embodiments, separate sequence assays are performed on the first and second nucleic acid preparations and the sequence information from both assays are combined to determine the genotype of the target nucleic acid in the biological sample. In some embodiments, the first preparative method is an amplification-based, a hybridization-based, or a circular probe-based preparative method. In some embodiments, the second method is an amplification-based, a hybridization-based, or a circular probe-based preparative method. In some embodiments, the first and second methods are of different types (e.g., only one of them is an amplification-based, a hybridization-based, or a circular probe-based preparative method, and the other one is one or the other two types of method). Accordingly, in some embodiments the second preparative method is an amplification-based, a hybridization-based, or a circular probe-based preparative method, provided that the second method is different from the first method. However, in some embodiments, both methods may be of the same type, provided they are different methods (e.g., both are amplification based or hybridization-based, but are different types of amplification or hybridization methods, e.g., with different relative biases).

In amplification-based (e.g., PCR-based or LCR-based, etc.) preparative methods, genomic loci (target nucleic acids) are isolated directly by means of a polymerase chain reaction or ligase chain reaction (or other amplification method) that selectively amplifies each locus using a pair of oligonucleotide primers. It is to be understood that primers will be sufficiently complementary to the target sequence to hybridize with and prime amplification of the target nucleic acid. Any one of a variety of art known methods may be utilized for primer design and synthesis. One or both of the primers may be perfectly complementary to the target sequence. Degenerate primers may also be used. Primers may also include additional nucleic acids that are

Petitioner Illumina, Inc., EX1004, p. 482

not complementary to target sequences but that facilitate downstream applications, including for example restriction sites and identifier sequences (e.g., source sequences). PCR based methods may include amplification of a single target nucleic acid and multiplex amplification (amplification of multiple target nucleic acids in parallel).

Hybridization-based preparative may methods involve selectively immobilizing target nucleic acids for further manipulation. It is to be understood that one or more oligonucleotides (immobilization oligonucleotides), which in some embodiments may be from 10 to 200 nucleotides in length, are used which hybridize along the length of a target region of a genetic locus to immobilize it. In some embodiments, immobilization oligonucleotides are either immobilized before hybridization is performed (e.g., Roche/Nimblegen 'sequence capture'), or are prepared such that they include a moiety (e.g., biotin) which can be used to selectively immobilize the target nucleic acid after hybridization by binding to e.g., streptavidin-coated microbeads (e.g., Agilent 'SureSelect').

Circularization selection-based preparative methods selectively convert each region of interest into a covalently-closed circular molecule which is then isolated by removal (usually enzymatic, e.g., with exonuclease) of any non-circularized linear nucleic acid. Oligonucleotide probes are designed which have ends that flank the region of interest. The probes are allowed to hybridize to the genomic target, and enzymes are used to first (optionally) fill in any gap between probe ends and second ligate the probe closed. In some embodiments, following circularization, any remaining (non-target) linear nucleic acid can be removed, resulting in isolation (capture) of target nucleic acid. Circularization selection-based preparative methods include molecular inversion probe capture reactions and 'selector' capture reactions. However, other techniques may be used as aspects of the invention are not limited in this respect. In some embodiments, molecular inversion probe capture of a target nucleic acid is indicative of the presence of a polymorphism in the target nucleic acid.

A variety of methods may be used to evaluate and compare bias profiles of each preparative technique. Next-generation sequencing may be used to quantitatively measure the abundance of each isolated target nucleic acid obtained from a certain preparative method. This abundance may be compared to a control abundance value (e.g., a known starting abundance of the target nucleic acid) and/or with an abundance determined through the use of an alternative preparative method. For example, a set of target nucleic acids may be isolated by one or more of

Petitioner Illumina, Inc., EX1004, p. 483

the three preparative methods; the target nucleic acid may be observed x times using the amplification technique, y times using the hybridization enrichment technique, and z times using the circularization selection technique. A pairwise correlation coefficient may be computed between each abundance value (e.g., x and y, x and z, and y and z) to assess bias in nucleic acid isolation between pairs of preparative methods. Since the mechanisms of isolation are different in each approach, the abundances will usually be different and largely uncorrelated with each other.

In some embodiments, the invention provides a method of obtaining a nucleic acid preparation that is representative of a target nucleic acid in a biological sample by obtaining a first preparation of a target nucleic acid using a first preparative method on a biological sample, obtaining a second preparation of a target nucleic acid using a second preparative method on the biological sample, and combining the first and second nucleic acid preparations to obtain a combined preparation that is representative of the target nucleic acid in the biological sample.

In some embodiments of any of the methods described herein, a third preparation of the target nucleic acid is obtained using a third preparative method that is different from the first and second preparative methods, wherein the first, second, and third preparative methods all have different systematic sequence biases. In some embodiments of any of the methods described herein, the different preparative methods are used for a plurality of different loci in the biological sample to increase the sensitivity of a multiplex nucleic acid analysis. In some embodiments, the target nucleic acid has a sequence of a gene selected from Table 1.

However, it should be appreciated that a genotyping method of the invention may include several steps, each of which independently may involve one or more different preparative techniques described herein. In some embodiments, a nucleic acid preparation may be obtained using one or more (e.g., 2, 3, 4, 5, or more) different techniques described herein (e.g., amplification, hybridization capture, circular probe capture, etc., or any combination thereof) and the nucleic acid preparation may be analyzed using one or more different techniques (e.g., amplification, hybridization capture, circular probe capture, etc., or any combination thereof) that are selected independently of the techniques used for the initial preparation.

In some embodiments, aspects of the invention also provide compositions, kits, devices, and analytical methods for increasing the sensitivity of nucleic acid assays. Aspects of the invention are particularly useful for increasing the confidence level of genotyping analyses.

Petitioner Illumina, Inc., EX1004, p. 484

However, aspects of the invention may be used in the context of any suitable nucleic acid analysis, for example, but not limited to, a nucleic acid analysis that is designed to determine whether more than one sequence variant is present in a sample.

In some embodiments, aspects of the invention relate to a plurality of nucleic acid probes (e.g., 10-50, 50-100, 100-250, 250-500, 500-1,000, 1,000-2,000, 2,000-5,000, 5,000-7,500, 7,500-10,000, or lower, higher, or intermediate number of different probes).

In some embodiments, each probe or each of a subset of probes (e.g., 10-25%, 25-50%, 50-75%, 75-90%, or 90-99%) has a different first targeting arm. In some embodiments, each probe or each probe of a subset of probes (e.g., 10-25%, 25-50%, 50-75%, 75-90%, or 90-99%) has a different second targeting arm. In some embodiments, the first and second targeting arms are separated by the same intervening sequence. In some embodiments, the first and second targeting arms are complementary to target nucleic acid sequences that are separated by the same or a similar length (e.g., number of nucleic acids, for example, 0-25, 25-50, 50-100, 100-250, 250-500, 500-1,000, 1,000-2,500 or longer or intermediate number of nucleotides) on their respective target nucleic acids (e.g., genomic loci). In some embodiments, each probe or a subset of probes (e.g., 10-30 25%, 25-50%, 50-75%, 75-90%, or 90-99%) includes a first primer binding sequence. In some embodiments, the primer binding sequence is the same (e.g., it can be used to prime sequencing or other extension reaction). In some embodiments, each probe or a subset of probes (e.g., 10-25%, 25-50%, 50-75%, 75-90%, or 90-99%) includes a unique identifier sequence tag (e.g., that is predetermined and can be used to distinguish each probe).

In some embodiments, the methods disclosed herein are useful for any application where sensitivity is important. For example, detection of cancer mutations in a heterogenous tissue sample, detection of mutations in maternally-circulating fetal DNA, and detection of mutations in cells isolated during a preimplantation genetic diagnostic procedure.

According to some aspects of the invention, methods of detecting a polymorphism in a nucleic acid in a biological sample are provided. In some embodiments, the methods comprise obtaining a nucleic acid preparation using a preparative method (e.g., any of the preparative methods disclosed herein) on a biological sample, and performing a molecular inversion probe capture reaction on the nucleic acid preparation, wherein a molecular inversion probe capture (e.g., using a mutation-detection MIP) of a target nucleic acid of the nucleic acid preparation is

Petitioner Illumina, Inc., EX1004, p. 485

indicative of the presence of a mutation (polymorphism) in the target nucleic acid, optionally wherein the polymorphism is selected from Table 2.

According to some aspects of the invention, methods of genotyping a nucleic acid in a biological sample are provided. In some embodiments, the methods comprise obtaining a nucleic acid preparation using a preparative method on a biological sample, sequencing a target nucleic acid of the nucleic acid preparation, and performing a molecular inversion probe capture reaction on the biological sample, wherein a molecular inversion probe capture of the target nucleic acid in the biological sample is indicative of the presence of a polymorphism in the target nucleic acid, genotyping the target nucleic acid based on the results of the sequencing and the capture reaction.

In some embodiments of the methods disclosed herein, the target nucleic acid has a sequence of a gene selected from Table 1.

It should be appreciated that any one or more embodiments described herein may be used for evaluating multiple genetic markers in parallel. Accordingly, in some embodiments, aspects of the invention relate to determining the presence of one or more markers (e.g., one or more alleles) at multiple different genetic loci in parallel.

Accordingly, the risk or presence of multiple heritable disorders may be evaluated in parallel. In some embodiments, the risk of having offspring with one or more heritable disorders may be evaluated. In some embodiments, an evaluation may be performed on a biological sample of a parent or a child (e.g., at a pre-implantation, prenatal, perinatal, or postnatal stage). In some embodiments, the disclosure provides methods for analyzing multiple genetic loci (e.g., a plurality of target nucleic acids selected from Table 1 or 2) from a patient sample, such as a blood, pre-implantation embryo, chorionic villus or amniotic fluid sample. A patient or subject may be a human. However, aspects of the invention are not limited to humans and may be applied to other species (e.g., mammals, birds, reptiles, other vertebrates or invertebrates) as aspects of the invention are not limited in this respect. A subject or patient may be male or female. In some embodiments, in connection with reproductive genetic counseling, samples from a male and female member of a couple may be analyzed. In some embodiments, for example, in connection with an animal breeding program, samples from a plurality of male and female subjects may be analyzed to determine compatible or optimal breeding partners or strategies for particular traits or to avoid one or more diseases or conditions. Accordingly,

Petitioner Illumina, Inc., EX1004, p. 486

reproductive risks may be determined and/or reproductive recommendations may be provided based on information derived from one or more embodiments of the invention.

However, it should be appreciated that aspects of the invention may be used in connection with any medical evaluation where the presence of one or more alleles at a genetic locus of interest is relevant to a medical determination (e.g., risk or detection of disease, disease prognosis, therapy selection, therapy monitoring, etc.). Further aspects of the invention may be used in connection with detection, in tumor tissue or circulating tumor cells, of mutations in cellular pathways that cause cancer or predict efficacy of treatment regimens, or with detection and identification of pathogenic organisms in the environment or a sample obtained from a subject, e.g., a human subject.

These and other aspects of the invention are described in more detail in the following description and non-limiting examples and drawings.

BRIEF DESCRIPTION OF DRAWINGS

FIG. 1 illustrates a non-limiting embodiment of a tiled probe layout;

FIG. 2 illustrates a non-limiting embodiment of a staggered probe layout;

FIG. 3 illustrates a non-limiting embodiment of an alternating staggered probe layout;

FIGS. 4A-C depict various non-limiting methods for combining differentiator tag sequence and target sequences (NNNN depicts a differentiator tag sequence);

FIG. 5 depicts a non-limiting method for genotyping based on target and 5 differentiator tag sequences;

FIG. 6 depicts non-limiting results of a simulation of a MIP capture reaction;

FIG. 7 depicts a non-limiting graph of sequencing coverage;

FIG. 8 illustrates that shorter sequences are captured with higher efficiency that longer sequences using MIPs;

FIG. 9 illustrates a non-limiting scheme of padlock (MIP) capture of a region that includes both repetitive regions (thick wavy line) and the adjacent unique sequence (thick strait line);

FIG. 10 illustrates a non-limiting hypothetical relationship between target gap size and the relative number of reads of the repetitive region;

FIG. 11A depicts MIP capture of FMR1 repeat regions from a diploid genome;

Petitioner Illumina, Inc., EX1004, p. 487

FIG. 11B depicts preparative methods for biallelic resolution of FMR I repeat region lengths in a diploid genome using MIP capture probes and unique differentiator tags;

FIG. 11C depicts an analysis of FMR1 repeat region lengths in a diploid genome;

FIG. 12 is a schematic of an embodiment of an algorithm of the invention;

FIG. 13 illustrates a non-limiting example of a graph of per-target abundance with MIP capture; and,

FIG. 14 shows a non-limiting a graph of correlation between two MIP capture reactions.

DETAILED DESCRIPTION

Aspects of the invention relate to preparative and analytical methods and compositions for evaluating genotypes, and in particular, for determining the allelic identity (or identities in a diploid organism) of one or more genetic loci in a subject.

Aspects of the invention are based, in part, on the identification of different sources of ambiguity and error in genetic analyses, and, in part, on the identification of one or more approaches to avoid, reduce, recognize, and/or resolve these errors and ambiguities at different stages in a genetic analysis. Aspects of the invention relate to methods and compositions for addressing bias and/or stochastic variation associated with one or more preparative and/or analytical steps of a nucleic acid evaluation technology. In some embodiments, preparative methods can be adapted to avoid or reduce the risk of bias skewing the results of a genetic analysis. In some embodiments, analytical methods can be adapted to recognize and correct for data variations that may give rise to misinterpretation (e.g., incorrect calls such as homozygous when the subject is actually heterozygous or heterozygous when the subject is actually homozygous). Methods of the invention may be used for any type of mutation, for example a single base change (e.g., insertion, deletion, transversion or transition, etc.), a multiple base insertion, deletion, duplication, inversion, and/or any other change or combination thereof.

In some embodiments, additional or alternative techniques may be used to address loci characterized by multiple repeats of a core sequence where the length of the repeat is longer than a typical sequencing read thereby making it difficult to determine whether a deletion or duplication of one or more core sequence units has occurred based solely on a sequence read.

In some embodiments, increased confidence in an assay result may be obtained by i) selecting two or more different preparative and/or analytical techniques that have different biases

Petitioner Illumina, Inc., EX1004, p. 488

(e.g., known to have different biases), ii) evaluating a patient sample using the two or more different techniques, iii) comparing the results from the two or more different techniques, and/or iv) determining whether the results are consistent for the two or more different techniques. In some embodiments, if determining in step (iv) indicates that the results are consistent (e.g., the same) then increased confidence in the assay result is obtained. In other embodiments, if determining in step (iv) indicates that the results are inconsistent (e.g., that the results are ambiguous) then one or more additional preparative and/or analytical techniques, which have a different bias (e.g., known to have a different bias) compared with the two or more different preparative and/or analytical techniques selected in step (i), are used to evaluate the patient sample, and the results of the one or more additional preparative and/or analytical techniques are compared with the results from step (ii) to resolve the inconsistency.

In some embodiments, two or more independent samples may be obtained from a subject and independently analyzed. In some embodiments, two or more independent samples are obtained at approximately the same time point. In some embodiments, two or more independent samples are obtained at multiple different time points. In some embodiments, the use of two or more independent sample facilitates the elimination, normalization, and/or quantification of stochastic measurement noise. It is to be appreciated that two or more independent samples may be obtained in connection with any of the methods disclosed herein, including, for example, methods for pathogen profiling in a human or other animal subjects, monitoring tumor progression/regression, analyzing circulating tumor cells, analyzing fetal cells in maternal circulation, and analyzing/monitoring/profiling of environmental pathogens.

In some embodiments, one or more of the techniques described herein may be combined in a single assay protocol for evaluating multiple patient samples in parallel.

It should be appreciated that aspects of the invention may be useful for high throughput, cost-effective, yet reliable, genotyping of multiple patient samples (e.g., in parallel, for example in multiplex reactions). In some embodiments, aspects of the invention are useful to reduce the error frequency in a multiplex analysis. Certain embodiments may be particularly useful where multiple reactions (e.g., multiple loci and/or multiple patient samples) are being processed. For example, 10-25, 25-50, 50-75, 75-100 or more loci may be evaluated for each subject out of any number of subject samples that may be processed in parallel (e.g., 1-25, 25-50, 50-100, 100-500, 500-1,000, 1,000-2,500, 2,500-5,000 or more or intermediate numbers of patient samples). It

Petitioner Illumina, Inc., EX1004, p. 489

should be appreciated that different embodiments of the invention may involve conducting two or more target capture reactions and/or two or more patient sample analyses in parallel in a single multiplex reaction. For example, in some embodiments a plurality of capture reactions (e.g., using different capture probes for different target loci) may be performed in a single multiplex reaction on a single patient sample. In some embodiments, a plurality of captured nucleic acids from each one of a plurality of patient samples may be combined in a single multiplex analysis reaction. In some embodiments, samples from different subjects are tagged with subject-specific (e.g., patient-specific) tags (e.g., unique sequence tags) so that the information from each product can be assigned to an identified subject. In some embodiments, each of the different capture probes used for each patient sample have a common patient-specific tag. In some embodiments, the capture probes do not have patient-specific tags, but the captured products from each subject may be amplified using one or a pair of amplification primers that are labeled with a patient-specific tag. Other techniques for associating a patient-specific tag with the captured product from a single patient sample may be used as aspects of the invention are not limited in this respect. It should be appreciated that patient-specific tags as used herein may refer to unique tags that are assigned to identified patients in a particular assay. The same tags may be used in a separate multiplex analysis with a different set of patient samples (e.g., from different patients) each of which is assigned one of the tags. In some embodiments, different sets of unique tags may be used in sequential (e.g., alternating) multiplex reactions in order to reduce the risk of contamination from one assay to the next and allow contamination to be detected on the basis of the presence of tags that are not expected to be present in a particular assay.

Embodiments of the invention may be used for any of a number of different settings: reproductive settings, disease screening, identifying subjects having cancer, identifying subjects having increased risk for a disease, stratifying a population of subjects according to one or more of a number of factors, for example responsiveness to a particular drug, lack or not of an adverse reaction (or risk therefore) to a particular drug, and/or providing information for medical records (e.g., homozygosity, heterozygosity at one or more loci). It should be appreciated that the invention is not limited to genomic analysis of patient samples. For example, aspects of the invention may be useful for high throughput genetic analysis of environment samples to detect pathogens.

Petitioner Illumina, Inc., EX1004, p. 490

In some embodiments, the methods disclosed herein are useful for diagnosis of one or more heritable disorders. In some embodiments, a heritable disorder that may be diagnosed with the methods disclosed herein is a genetic disorder that is prevalent in the Ashkenazi Jewish population. In some embodiments, the heritable disorders are selected from: 21-Hydroxylase-Defiocient Congenital Adrenal Hyperplasia; ABCC8-Related Hyperinsulinism; Alpha-Thalassemia, includes Constant Spring, & MR associated; Arylsulfatase A Deficiency-Metyachromatic Leukodystrophy; Biotinidase Deficiency Holocarboxylase Synthetase Deficiency; Bloom's Syndrome; Canavan Disease; CFTR-Related Disorders-cystic fibrosis; Citrullinemia Type 1; Combined MMA & Homocystinuria-dblC; Dystrophinopathies (DMD & BMD); Familial Dysautonomia; Fanconi Anemia-FANCC; Galactosemia-Classical: Galactokinase Defiency & Galactose Epimerase Deficiency; Gaucher Disease; GJB2-Related DFNB 1 Nonsyndromic Hearing Loss and Deafness; Glutaric acidemia Type 1; Hemoglobinopathies beta-chain disorders; Glycogen Storage Disease Type 1A; Maple Syrup Urine Disease; Types 1A, 1B, 2, 3; Medium Chain Acyl-Coenzyme A; Dehydrogenase Deficiency-MCADD; Methylmalonic Acidemia; Mucolipidosis IV; Nemaline Myopathy; Nieman-Pick Type A-Acid Sphingomyelinase Deficiency; Non-Ketotic Hyperglycinemia-Glycine Encephalopathy; Ornithine Transcarbamylase Deficiency; PKU Phenylalanine Hydroxylase Deficiency; Propionic Acidemia; Short Chain Acyl-CoA Dehydrogenase Deficiency-SCADD; Smith-Lemli-Opitz Syndrome; Spinal Muscular Atrophy (SMNI)-SMA; Tay Sachs-HexA Deficiency; Usher Synbdrome-Type I (Type IB, Type IC, Type ID, Type IF, Type IG) ; X-Linked Mental Retardation ARX-Related Disorders; X-Linked Mental Retardation with Cerebellar Cypoplasia and sistinctive Facial Appearance; X-Linked Mental Retardation; includes 9, 21, 30, 46, 58, 63, 88, 89; X linked mental retardation: FM1-Related Disorders-FRXA, Fragile X MR; X-linked SMR: Renpenning Syndrome 1; Zellweger Spectrum disorders - Peroxisomal Bifunctional Enzyme Deficiencies including Zellweger, NALD, and/or infantile Refsums. However, all of these, subsets of these, other genes, or combinations thereof may be used.

According to some aspects, the disclosure relates to multiplex diagnostic methods. In some embodiments, multiplex diagnostic methods comprise capturing a plurality of genetic loci in parallel (e.g., a genetic locus of Table 1). In some embodiments, genetic loci possess one or more polymorphisms (e.g., a polymorphism of Table 2) the genotypes of which correspond to

Petitioner Illumina, Inc., EX1004, p. 491

disease causing alleles. Accordingly, in some embodiments, the disclosure provides methods for assessing multiple heritable disorders in parallel.

In some embodiments, methods are provided for diagnosing multiple heritable disorders in parallel at a pre-implantation, prenatal, perinatal, or postnatal stage. In some embodiments, the disclosure provides methods for analyzing multiple genetic loci (e.g., a plurality of target nucleic acids selected from Table 1) from a patient sample, such as a blood, pre-implantation embryo, chorionic villus or amniotic fluid sample. A patient or subject may be a human. However, aspects of the invention are not limited to humans and may be applied to other species (e.g., mammals, birds, reptiles, other vertebrates or invertebrates) as aspects of the invention are not limited in this respect. A subject or patient may be male or female. In some embodiments, in connection with reproductive genetic counseling, samples from a male and female member of a couple may be analyzed. In some embodiments, for example, in connection with an animal breeding program, samples from a plurality of male and female subjects may be analyzed to determine compatible or optimal breeding partners or strategies for particular traits or to avoid one or more diseases or conditions.

However, it should be appreciated that any other diseases may be studied and/or risk factors for diseases or disorders including, but not limited to allergies, responsiveness to treatment, cancer tumor profiling for treatment and prognosis, monitoring and identification of patient infections, and monitoring of environmental pathogens.

1. Reducing representational bias in multiplex amplification reactions:

In some embodiments, aspects of the invention relate to methods that reduce bias and increase reproducibility in multiplex detection of genetic loci, e.g., for diagnostic purposes.

Molecular inversion probe technology is used to detect or amplify particular nucleic acid sequences in potentially complex mixtures. Use of molecular inversion probes has been demonstrated for detection of single nucleotide polymorphisms (Hardenbol et al. 2005 Genome Res 15:269-75) and for preparative amplification of large sets of exons (Porreca et al. 2007 Nat Methods 4:931-6, Krishnakumar et al. 2008 Proc Natl Acad Sci USA 105:9296-301). One of the main benefits of the method is in its capacity for a high degree of multiplexing, because generally thousands of targets may be captured in a single reaction containing thousands of probes. However, challenges associated with, for example, amplification efficiency (See, e.g.,

Petitioner Illumina, Inc., EX1004, p. 492

Turner EH, et al., Nat Methods. 2009 Apr 6:1-2.) have limited the practical utility of the method in research and diagnostic settings.

Aspects of the disclosure are based, in part, on the discovery of effective methods for overcoming challenges associated with systematic errors (bias) in multiplex genomic capture and sequencing methods, namely high variability in target nucleic acid representation and unequal sampling of heterozygous alleles in pools of captured target nucleic acids (e.g., isolated from a biological sample). Accordingly, in some embodiments, the disclosure provides methods that reduce variability in the detection of target nucleic acids in multiplex capture methods. In other embodiments, methods improve allelic representation in a capture pool and, thus, improve variant detection outcomes. In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve the use of different sets of multiple probes (e.g., molecular inversion probes MIPs) that capture overlapping regions of a target nucleic acid to achieve a more uniform representation of the target nucleic acids in a capture pool compared with methods of the prior art. In other embodiments, methods reduce bias, or the risk of bias, associated with large scale parallel capture of genetic loci, e.g., for diagnostic purposes. In other embodiments, methods are provided for increasing reproducibility (e.g., by reducing the effect of polymorphisms on target nucleic acid capture) in the detection of a plurality of genetic loci in parallel. In further embodiments, methods are provided for reducing the effect of probe synthesis and/or probe amplification variability on the analysis of a plurality of genetic loci in parallel.

In some aspects, the disclosure provides probe sets that comprise a plurality of different probes. As used herein, a 'probe' is a nucleic acid having a central region flanked by a 5' region and a 3' region that are complementary to nucleic acids flanking the same strand of a target nucleic acid or subregion thereof. An exemplary probe is a molecular inversion probe (MIP). A 'target nucleic acid' may be a genetic locus.

Exemplary genetic loci are disclosed herein in Table 1 (RefSeqGene Column).

While probes have been typically designed to meet certain constraints (e.g. melting temperature, G/C content, etc.) known to partially affect capture/amplification efficiency (Ball et al (2009) Nat Biotech 27:361-8 AND Deng et al (2009) Nat Biotech 27:353-60), a set of constraints which is sufficient to ensure either largely uniform or highly reproducible capture/amplification efficiency has not previously been achieved.

Petitioner Illumina, Inc., EX1004, p. 493

As disclosed herein, uniformity and reproducibility can be increased by designing multiple probes per target, such that each base in the target is captured by more than one probe. In some embodiments, the disclosure provides multiple MIPs per target to be captured, where each MIP in a set designed for a given target nucleic acid has a central region and a 5' region and 3' region ('targeting arms') which hybridize to (at least partially) different nucleic acids in the target nucleic acid (immediately flanking a subregion of the target nucleic acid). Thus, differences in efficiency between different targeting arms and fill-in sequences may be averaged across multiple MIPs for a single target, which results in more uniform and reproducible capture efficiency.

In some embodiments, the methods involve designing a single probe for each target (a target can be as small as a single base or as large as a kilobase or more of contiguous sequence).

It may be preferable, in some cases, to design probes to capture molecules (e.g., target nucleic acids or subregions thereof) having lengths in the range of 1 - 200 bp (as used herein, a bp refers to a base pair on a double-stranded nucleic acid - however, where lengths are indicated in bps, it should be appreciated that single-stranded nucleic acids having the same number of bases, as opposed to base pairs, in length also are contemplated by the invention). However, probe design is not so limited. For example, probes can be designed to capture targets having lengths in the range of up to 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 1000, or more bps, in some cases.

It is to be appreciated that the length of a capture molecule (e.g., a target nucleic acid or subregion thereof) is selected based upon multiple considerations. For example, where analysis of a target involves sequencing, e.g., with a next-generation sequencer, the target length should typically match the sequencing read-length so that shotgun library construction is not necessary. However, it should be appreciated that captured nucleic acids may be sequenced using any suitable sequencing technique as aspects of the invention are not limited in this respect.

It is also to be appreciated that some target nucleic acids are too large to be captured with one probe. Consequently, it may be necessary to capture multiple subregions of a target nucleic acid in order to analyze the full target.

In some embodiments, a subregion of a target nucleic acid is at least 1 bp. In other embodiments, a subregion of a target nucleic acid is at least 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 bp or more. In other embodiments, a

Petitioner Illumina, Inc., EX1004, p. 494

subregion of a target nucleic acid has a length that is up to 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or more percent of a target nucleic acid length.

The skilled artisan will also appreciate that consideration is made, in the design of MIPs, for the relationship between probe length and target length. In some embodiments, MIPs are designed such that they are several hundred basepairs (e.g., up to 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 bp or more) longer than corresponding target (e.g., subregion of a target nucleic acid, target nucleic acid).

In some embodiments, lengths of subregions of a target nucleic acid may differ.

For example, if a target nucleic acid contains regions for which probe hybridization is not possible or inefficient, it may be necessary to use probes that capture subregions of one or more different lengths in order to avoid hybridization with problematic nucleic acids and capture nucleic acids that encompass a complete target nucleic acid.

Aspects of the invention involve using multiple probes, e.g., MIPs, to amplify each target nucleic acid. In some embodiments, the set of probes for a given target can be designed to 'tile' across the target, capturing the target as a series of shorter sub targets. In some embodiments, where a set of probes for a given target is designed to 'tile' across the target, some probes in the set capture flanking non-target sequence). Alternately, the set can be designed to 'stagger' the exact positions of the hybridization regions flanking the target, capturing the full target (and in some cases capturing flanking non-target sequence) with multiple probes having different targeting arms, obviating the need for tiling. The particular approach chosen will depend on the nature of the target set. For example, if small regions are to be captured, a staggered-end approach might be appropriate, whereas if longer regions are desired, tiling might be chosen. In all cases, the amount of bias-tolerance for probes targeting pathological loci can be adjusted ('dialed in') by changing the number of different MIPs used to capture a given molecule.

In some embodiments, the 'coverage factor', or number of probes used to capture a basepair in a molecule, is an important parameter to specify. Different numbers of probes per target are indicated depending on whether one is using the tiling approach (see, e.g., FIG. 1) or one of the staggered approaches (see, e.g., FIGs. 2 or 3).

FIG. 1 illustrates a non-limiting embodiment of a tiled probe layout showing ten captured sub-targets tiled across a single target. Each position in the target is covered by three sub-targets such that MIP performance per base pair is averaged across three probes.

Petitioner Illumina, Inc., EX1004, p. 495

FIG. 2 illustrates a non-limiting embodiment of a staggered probe layout showing the targets captured by a set of three MIPs. Each MIP captures the full target, shown in black, plus (in some cases) additional extra-target sequence, shown in gray, such that the targeting arms of each MIP fall on different sequence. Each position in the target is covered by three sub-targets such that MIP performance per basepair is averaged across three probes. Targeting arms land immediately adjacent to the black or gray regions shown. It should be appreciated that in some embodiments, the targeting arms (not shown) can be designed so that they do not overlap with each other.

FIG. 3 illustrates a non-limiting embodiment of an alternating staggered probe layout showing the targets captured by a set of three MIPs. Each MIP captures the full target, shown in black, plus (in some cases) additional extra-target sequence, shown in gray, such that the targeting arms of each MIP fall on different sequence. Each position in the target is covered by three sub-targets such that MIP performance per basepair is averaged across three probes. Targeting arms land immediately adjacent to the black or gray regions shown.

It should be appreciated that for any of the layouts, the targeting arms on adjacent tiled or staggered probes may be designed to either overlap, not overlap, or overlap for only a subset of the probes.

In certain embodiments for any of the layouts, a coverage factor of about 3 to about 10 is used. However, the methods are not so limited and coverage factors of up to 2, 3, 4, 5, 6, 7, 8, 9, 10, 20 or more may be used. It is to be appreciated that the coverage factor selected may depend the probe layout being employed. For example, in the tiling approach, for a desired coverage factor, the number of probes per target is typically a function of target length, sub-target length, and spacing between adjacent sub-target start locations (step size). For example, for a desired coverage factor of 3, a 200 bp target with a start-site separation of 20 bp and sub-target length of 60 bp may be encompassed with 12 MIPs (FIG. 1). Thus, a specific coverage factor may be achieved by varying the number of probes per target nucleic acid and the length of the molecules captured. In the staggered approach, a fixed-length target nucleic acid is captured as several subregions or as 'super-targets', which are molecules comprising the target nucleic acid and additional flanking nucleic acids, which may be of varying lengths. For example, a target of 50 bp can be captured at a coverage factor of 3 with 3 probes in either a 'staggered' (FIG. 2) or 'alternating staggered' configuration (FIG. 3).

Petitioner Illumina, Inc., EX1004, p. 496

The coverage factor will be driven by the extent to which detection bias is tolerable. In some cases, where the bias tolerance is small, it may be desirable to target more subregions of target nucleic acid with, perhaps, higher coverage factors. In some embodiments, the coverage factor is up to 2, 3, 4, 5, 6, 7, 8, 9, 10 or more.

In some embodiments, when a tiled probe layout is used, when the target length is greater than 1 bp and when a step size (distance between the 5'-end of a target and the 5' end of its adjacent target) is less than the length of a target or subregion thereof, it is possible to compute probe number for a particular target based on target length (T), sub target length (S), and coverage factor (C), such that probe number = T/(S/C) + (C-1).

In some aspects, the disclosure provides methods to increase the uniformity of amplification efficiency when multiple molecules are amplified in parallel; methods to increase the reproducibility of amplification efficiency; methods to reduce the contribution of targeting probe variability to amplification efficiency; methods to reduce the effect on a given target nucleic acid of polymorphisms in probe hybridization regions; and/or methods to simplify downstream workflows when multiplex amplification by MIPs is used as a preparative step for analysis by nucleic acid sequencing.

Polymorphisms in the target nucleic acid under the regions flanking a target can interfere with hybridization, polymerase fill-in, and/or ligation. Furthermore, this may occur for only one allele, resulting in allelic drop-out, which ultimately decreases downstream sequencing accuracy. In some embodiments, using a set of MIPs having multiple hybridization sites for the capture of any given target, the probability of loss from polymorphism is substantially decreased because not all targeting arms in the set of MIPs will cover the location of the mutation.

Probes for MIP capture reactions may be synthesized on programmable microarrays because of the large number of sequences required. Because of the low synthesis yields of these methods, a subsequent amplification step is required to produce sufficient probe for the MIP amplification reaction. The combination of multiplex oligonucleotide synthesis and pooled amplification results in uneven synthesis error rates and representational biases. By synthesizing multiple probes for each target, variation from these sources may be averaged out because not all probes for a given target will have the same error rates and biases.

Multiplex amplification strategies disclosed herein may be used analytically, as in detection of SNPs, or preparatively, often for next-generation sequencing or other sequencing

Petitioner Illumina, Inc., EX1004, p. 497

techniques. In the preparative setting, the output of an amplification reaction is generally the input to a shotgun library protocol, which then becomes the input to the sequencing platform. The shotgun library is necessary in part because next-generation sequencing yields reads significantly shorter than amplicons such as exons. In addition to the bias-reduction afforded by the multi-tiled approach described here, tiling also obviates the need for shotgun library preparation. Since the length of the capture molecule can be specified when the probes, e.g., MIPs, are designed, it can be chosen to match the readlength of the sequencer. In this way, reads can 'walk' across an exon by virtue of the start position of each capture molecule in the probe set for that exon. Reducing analytical errors associated with bias in nucleic acid preparations:

In some embodiments, aspects of the invention relate to preparative steps in DNA sequencing-related technologies that reduce bias and increase the reliability and accuracy of downstream quantitative applications.

There are currently many genomics assays that utilize next-generation (polony-based) sequencing to generate data, including genome resequencing, RNA-seq for gene expression, bisulphite sequencing for methylation, and Immune-seq, among others. In order to make quantitative measurements (including genotype calling), these methods utilize the counts of sequencing reads of a given genomic locus as a proxy for the representation of that sequence in the original sample of nucleic acids. The majority of these techniques require a preparative step to construct a high-complexity library of DNA molecules that is representative of a sample of interest. This may include chemical or biochemical treatment of the DNA (e.g., bisulphite treatment), capture of a specific subset of the genome (e.g., padlock probe capture, solution hybridization), and a variety of amplification techniques (e.g., polymerase chain reaction, whole genome amplification, rolling circle amplification).

Systematic and random errors are common problems associated with genome amplification and sequencing library construction techniques. For example, genomic sequencing library may contain an over- or under-representation of particular sequences from a source genome as a result of errors (bias) in the library construction process. Such bias can be particularly problematic when it results in target sequences from a genome being absent or undetectable in the sequencing libraries. For example, an under representation of particular allelic sequences (e.g., heterozygotic alleles) from a genome in a sequencing library can result in an apparent homozygous representation in a sequencing library. As most downstream

Petitioner Illumina, Inc., EX1004, p. 498

sequencing library quantification techniques depend on stochastic counting processes, these problems have typically been addressed by sampling enough (over-sampling) to obtain a minimum number of observations necessary to make statistically significant decisions. However, the strategy of oversampling is generally limited to elimination of low-count Poisson noise, and the approach wastes resources and increases the expense required to perform such experiments. Moreover, oversampling can result in a reduced statistical confidence in certain conclusions (e.g., diagnostic calls) based on the data. Accordingly, new approaches are needed for overcoming bias in sequencing library preparatory methods.

Aspects of the disclosure are based, in part, on the discovery of methods for overcoming problems associated with systematic and random errors (bias) in genome capture, amplification and sequencing methods, namely high variability in the capture and amplification of nucleic acids and disproportionate representation of heterozygous alleles in sequencing libraries. Accordingly, in some embodiments, the disclosure provides methods that reduce variability in the capture and amplification of nucleic acids. In other embodiments, the methods improve allelic representation in sequencing libraries and, thus, improve variant detection outcomes. In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve the use of differentiator tag sequences to uniquely tag individual nucleic acid molecules. In some embodiments, the differentiator tag sequence permits the detection of bias based on the frequency with which pairs of differentiator tag and target sequences are observed in a sequencing reaction. In other embodiments, the methods reduce errors caused by bias, or the risk of bias, associated with the capture, amplification and sequencing of genetic loci, e.g., for diagnostic purposes.

Aspects of the invention relate to associating unique sequence tags (referred to as differentiator tag sequences) with individual target molecules that are independently captured and/or analyzed (e.g., prior to amplification or other process that may introduce bias). These tags are useful to distinguish independent target molecules from each other thereby allowing an analysis to be based on a known number of individual target molecules. For example, if each of a plurality of target molecule sequences obtained in an assay is associated with a different differentiator tag, then the target sequences can be considered to be independent of each other and a genotype likelihood can be determined based on this information. In contrast, if each of the plurality of target molecule sequences obtained in the assay is associated with the same

Petitioner Illumina, Inc., EX1004, p. 499

differentiator tag, then they probably all originated from the same target molecule due to over-representation (e.g., due to biased amplification) of this target molecule in the assay. This provides less information than the situation where each nucleic acid was associated with a different differentiator tag. In some embodiments, a threshold number of independently isolated molecules (e.g., unique combinations of differentiator tag and target sequences) is analyzed to determine the genotype of a subject.

In some embodiments, the invention relates to compositions comprising pools (libraries) of preparative nucleic acids that each comprise "differentiator tag sequences" for detecting and reducing the effects of bias, and for genotyping target nucleic acid sequences. As used herein, a "differentiator tag sequence" is a sequence of a nucleic acid (a preparative nucleic acid), which in the context of a plurality of different isolated nucleic acids, identifies a unique, independently isolated nucleic acid. Typically, differentiator tag sequences are used to identify the origin of a target nucleic acid at one or more stages of a nucleic acid preparative method. For example, in the context of a multiplex nucleic acid capture reaction, differentiator tag sequences provide a basis for differentiating between multiple independent, target nucleic acid capture events. Also, in the context of a multiplex nucleic acid amplification reaction, differentiator tag sequences provide a basis for differentiating between multiple independent, primary amplicons of a target nucleic acid, for example. Thus, combinations of target nucleic acid and differentiator tag sequence (target:differentiator tag sequences) of an isolated nucleic acid of a preparative method provide a basis for identifying unique, independently isolated target nucleic acids. FIG. 4A-C depict various non-limiting examples of methods for combining differentiator tag sequence and target sequences.

It will be apparent to the skilled artisan that differentiator tags may be synthesized using any one of a number of different methods known in the art. For example, differentiator tags may be synthesized by random nucleotide addition.

Differentiator tag sequences are typically of a predefined length, which is selected to control the likelihood of producing unique target:differentiator tag sequences in a preparative reaction (e.g., amplification-based reaction, a circularization selection-based reaction, e.g., a MIP reaction). Differentiator tag sequences may be, up to 5, up to 6, up to 7 up to 8, up to 9, up to 10, up to 11, up to 12, up to 13, up to 14, up to 15, up to 16, up to 17, up to 18, up to 19, up to 20, up to 21, up to 22, up to 23, up to 24, up to 25, or more nucleotides in length. For purposes of

Petitioner Illumina, Inc., EX1004, p. 500

genotyping, isolated nucleic acids are identified as independently isolated if they comprise unique combinations of target nucleic acid and differentiator tag sequences, and observance of threshold numbers of unique combinations of target nucleic acid and differentiator tag sequences provide a certain statistical confidence in the genotype.

During a library preparation process, each nucleic acid molecule may be tagged with a unique differentiator tag sequence in a configuration that permits the differentiator tag sequence to be sequenced along with the target nucleic acid sequence of interest (the nucleic acid sequence for which the library is being prepared, e.g., a polymorphic sequence). The incorporation of the nucleic acid comprising a differentiator tag sequence at a particular step allows the detection and correction of biases in subsequent steps of the protocol.

A large library of unique differentiator tag sequences may be created by using degenerate, random-sequence polynucleotides of defined length. The differentiator tag sequences of the polynucleotides may be read at the final stage of the sequencing. The observations of the differentiator tag sequences may be used to detect and correct biases in the final sequencing read-out of the library. For example, the total possible number of differentiator tag sequences, which may be produced, e.g., randomly, is 4^N, where N is the length of the differentiator tag sequence. Thus, it is to be understood that the length of the differentiator tag sequence may be adjusted such that the size of the population of MIPs having unique differentiator tag sequences is sufficient to produce a library of MIP capture products in which identical independent combinations of target nucleic acid and differentiator tag sequence are rare. As used herein combinations of target nucleic acid and differentiator tag sequences, may also be referred to as "target:differentiator tag sequences".

In the final readout of a sequencing process, each read may have an additional unique differentiator tag sequence. In some embodiments, when differentiator tag sequences are distributed randomly in a library, all the unique differentiator tag sequences will be observed about an equal number of times. Accordingly, the number of occurrences of a differentiator tag sequence may follow a Poisson distribution.

In some embodiments, overrepresentation of target:differentiator tag sequences in a pool of preparative nucleic acids (e.g., amplified MIP capture products) is indicative of bias in the preparative process (e.g., bias in the amplification process). For example, target:differentiator tag sequence combinations that are statistically overrepresented are indicative of bias in the

Petitioner Illumina, Inc., EX1004, p. 501

protocol at one or more steps between the incorporation of the differentiator tag sequences into MIPs and the actual sequencing of the MIP capture products.

The number of reads of a given target:differentiator tag sequence may be indicative (may serve as a proxy) of the amount of that target sequence present in the originating sample. In some embodiments, the numbers of occurrence of sequences in the originating sample is the quantity of interest. For example, using the methods disclosed herein, the occurrence of differentiator tag sequences in a pool of MIPs may be predetermined (e.g., may be the same for all differentiator tag sequences). Accordingly, changes in the occurrence of differentiator tag sequences after amplification and sequencing may be indicative of bias in the protocol. Bias may be corrected to provide an accurate representation of the composition of the original MIP pool, e.g., for diagnostic purposes.

According to some aspects, a library of preparative nucleic acid molecules (e.g., MIPs, each nucleic acid in the library having a unique differentiator tag sequence, may be constructed such that the number of nucleic acid molecules in the library is significantly larger than the number prospective target nucleic acid molecules to be captured using the library. This ensures that products of the preparative methods include only unique target:differentiator tag sequence; e.g., in a MIP reaction the capture step would undersample the total population of unique differentiator tag sequences in the MIP library. For example, an experiment utilizing 1 ug of genomic DNA will contain about ~150,000 copies of a diploid genome. For a MIP library, each MIP in the library comprising a randomly produced 12-mer differentiator tag sequence (~1.6 million possible unique differentiator tag sequences), there would be more than 100 unique differentiator tag sequences per genomic copy. For a MIP library, each MIP in the library comprising a randomly produced 15-mer differentiator tag sequence (~1 billion possible unique differentiator tag sequences), there would be more than 7000 unique differentiator tag sequences per genomic copy. Therefore, the probability of the same differentiator tag sequence being incorporated multiple times is incredibly small. Thus, it is to be appreciated that the length of the differentiator tag sequence is to be selected based on the amount of target sequence in a MIP capture reaction and the desired probability for having multiple, independent occurrences of target:differentiator tag sequence combinations.

Petitioner Illumina, Inc., EX1004, p. 502

FIG. 5 depicts a non-limiting method for genotyping based on target and differentiator tag sequences. Sequencing reads of target and differentiator tags sequences are collapsed to make diploid genotype calls. FIG. 6 depicts non-limiting results of a simulation of a MIP capture reaction in which MIP probes, each having a differentiator tag sequence of 15 nucleotides, are combined with 10000 target sequence copies (e.g., genome equivalents). In this simulated reaction, the probability of capturing one or more copies of a target sequence having the same differentiator tag sequence is 0.05. The Y axis reflects the number of observations. The X axis reflects the number of independent occurrences of target:differentiator tag combinations. FIG. 7 depicts a non-limiting graph of sequencing coverage, which can help ensure that alleles are sampled to sufficient depth (e.g., either 10x or 20x minimum sampling per allele, assuming 1000 targets). In this non-limiting example, the X axis is total per-target coverage required, and the Y axis is the probability that a given total coverage will result in at least 10x or 20x coverage for each allele.

The skilled artisan will appreciate that as part of a MIP library preparation process, adapters may be ligated onto the ends of the molecules of interest. Adapters often contain PCR primer sites (for amplification or emulsion PCR) and/or sequencing primer sites. In addition, barcodes may be included, for example, to uniquely identify individual samples (e.g., patient samples) that may be mixed together. (See, e.g., USPTO Publication Number US 2007/0020640 Al (McCloskey et al.)

The actual incorporation of the random differentiator tag sequences can be performed through various methods known in the art. For example, nucleic acids comprising differentiator tag sequences may be incorporated by ligation. This is a flexible method, because molecules having differentiator tag sequence can be ligated to any blunt-ended nucleic acids. The sequencing primers must be incorporated subsequently such that they sequence both the differentiator tag sequence and the target sequence. Alternatively, the sequencing adaptors can be synthesized with the random differentiator tag sequences at their 3' end (as degenerate bases), so that only one ligation must be performed. Another method is to incorporate the differentiator tag sequence into a PCR primer, such that the primer structure is arranged with the common adaptor sequence followed by the random differentiator tag sequence followed by the PCR priming sequence (in 5' to 3' order). A differentiator tag sequence and adaptor sequence (which may contain the sequencing primer site) are incorporated as tags. Another method to incorporate

Petitioner Illumina, Inc., EX1004, p. 503

the differentiator tag sequences is to synthesize them into a padlock probe prior to performing a gene capture reaction. The differentiator tag sequence is incorporated 3' to the targeting arm but 5' to the amplification primer that will be used downstream in the protocol. Another method to incorporate the differentiator tag sequences is as a tag on a gene-specific or poly-dT reverse-transcription primer. This allows the differentiator tag sequence to be incorporated directly at the cDNA level.

In some embodiments, at the incorporation step, the distribution of differentiator tag sequences can be assumed to be uniform. In this case, bias in any part of the protocol would change the uniformity of this distribution, which can be observed after sequencing. This allows the differentiator tag sequence to be used in any preparative process where the ultimate output is sequencing of many molecules in parallel.

Differentiator tag sequences may be incorporated into probes (e.g., MIPs) of a plurality when they are synthesized on-chip in parallel, such that degeneracy of the incorporated nucleotides is sufficient to ensure near-uniform distribution in the plurality of probes. It is to be appreciated that amplification of a pool of unique differentiator tag sequences may itself introduce bias in the initial pool. However, in most practical cases, the scale of synthesis (e.g., by column synthesis, chip based synthesis, etc.) is large enough that amplification of an initial pool of differentiator tag sequences is not necessary. By avoiding amplification or selection steps on the pool of unique differentiator tag sequences, potential bias may be minimized.

One example of the use of the differentiator tag sequences is in genome re-sequencing. Considering that the raw accuracy of most next-generation sequencing instruments is relatively low, it is crucial to oversample the genomic loci of interest.

Furthermore, since there are two alleles at every locus, it is important to sample enough to ensure that both alleles have been observed a sufficient number of times to determine with a sufficient degree of statistical confidence whether the sample is homozygous or heterozygous. Indeed, the sequencing is performed to sample the composition of molecules in the originating sample. However, after multiple reads have been collected for a given locus, it is possible that due to bias (e.g., caused by PCR amplification steps), a large fraction of the reads are derived from a single originating molecule. This would skew the population of target sequences observed, and would affect the outcome of the genotype call. For example, it is possible that a locus that is heterozygous is called as homozygous, because there are only a few observations of

Petitioner Illumina, Inc., EX1004, p. 504

the second allele out of many observations of that locus. However, if information is available on differentiator tag sequences, this situation could be averted, because the over-represented allele would be seen to also have an over-represented differentiator tag sequence (i.e., the sequences with the overrepresented differentiator tag sequence all originated from the same single molecule). Therefore, the sequences and corresponding distribution of differentiator tag sequences can be used as an additional input to the genotype-calling algorithm to significantly improve the accuracy and confidence of the genotype calls.

In some aspects, the disclosure provides methods for analyzing a plurality of target sequences which are genetic loci or portions of genetic loci (e.g., a genetic locus of Table 1). The genetic loci may be analyzed by sequencing to obtain a genotype at one or more polymorphisms (e.g., SNPs). Exemplary polymorphisms are disclosed in Table 2. The skilled artisan will appreciate that other polymorphisms are known in the art and may

Petitioner Illumina, Inc., EX1004, p. 505

be identified, for example, by querying the Entrez Single Nucleotide Polymorphism database, for example, by searching with a GeneID from Table 1.

Table 1: Target Nucleic Acids

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
CYP21A2	1589	cytochrome P450, family 21, subfamily A, polypeptide 2	CAH1; CPS1; CA21H; CYP21; CYP21B; P450c21B; MGC150536; MGC150537; CYP21A2	201910	NG_008337.1	6p21.3
ABCC8	6833	ATP-binding cassette, sub-family C (CFTR/MRP), member 8	HI; SUR; HHF1; MRP8; PHHI; SUR1; ABC36; HRINS; TNDM2; ABCC8	600509	NG_008867.1	11p15.1
ATRX	546	alpha thalassemia/mental retardation syndrome X-linked (RAD54 homolog, S. cerevisiae)	SHS; XH2; XNP; ATR2; SFM1; RAD54; MRXHF1; RAD54L; ZNF-HX; MGC2094; ATRX	300032	NG_008838.1	Xq13.1-q21.1
ARSA	410	arylsulfatase A	MLD; ARSA	607574	NG_009260.1	22q13.31-qter; 22q13.33
PSAP	5660	Prosaposin	GLBA; SAP1; FLJ00245; MGC10993; PSAP	176801	NG_008835.1	10q21-q22
BTD	686	Biotinidase	BTD	609019	NG_008819.1	3p25

5

Petitioner Illumina, Inc., EX1004, p. 506

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
HLCS	3141	holocarboxylase synthetase (biotin-(proprionyl-Coenzyme A-carboxylase (ATP-hydrolysing)) ligase)	HCS; HLCS	609018	NC_000021.7	21q22.1; 21q22.13
BLM	641	Bloom syndrome, RecQ helicase-like	BS; RECQ2; RECQL2; RECQL3; MGC126616; MGC131618; MGC131620; BLM	604610	NG_007272.1	15q26.1
ASPA	443	aspartoacylase (Canavan disease)	ASP; ACY2; ASPA	608034	NG_008399.1	17pter-p13
CFTR	1080	cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7)	CF; MRP7; ABC35; ABCC7; CFTR/MRP; TNR-CFTR; dJ760C5.1; CFTR	602421	NC_000007.12	7q31.2
ASS1	445	argininosuccinate synthetase 1	ASS; CTLN1; ASS1	603470	NG_011542.1	9q34.1
MMACHC	25974	methylmalonic aciduria (cobalamin deficiency) cblC type, with homocystinuria	cblC; FLJ25671; DKFZp564I122; RP11-291L19.3; MMACHC	609831	NC_000001.9	1p34.1
IKBKAP	8518	inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase complex-associated protein	FD; DYS; ELP1; IKAP; IKI3; TOT1; FLJ12497; DKFZp781H1425; IKBKAP	603722	NG_008788.1	9q31
FANCC	2176	Fanconi anemia, complementation group C	FA3; FAC; FACC; FLJ14675; FANCC	227645	NG_011707.1	9q22.3
GALK1	2584	galactokinase 1	GK1; GALK; GALK1	604313	NG_008079.1	17q24
GALT	2592	galactose-1-phosphate uridylyltransferase	GALT	606999	NC_000009.10	9p13
GALE	2582	UDP-galactose-4-epimerase	SDR1E1; FLJ95174; FLJ97302; GALE	606953	NG_007068.1	1p36-p35
GBA	2629	glucosidase, beta; acid (includes glucosylceramidase)	GCB; GBA1; GLUC; GBA	606463	NG_009783.1	1q21
GJB2	2706	gap junction protein, beta 2, 26kDa	HID; KID; PPK; CX26; DFNA3; DFNB1; NSRD1; DFNA3A; DFNB1A; GJB2	121011	NG_008358.1	13q11-q12
GCDH	2639	glutaryl-Coenzyme A dehydrogenase	GCD; ACAD5; GCDH	608801	NG_009292.1	19p13.2
G6PC	2538	glucose-6-phosphatase, catalytic subunit	G6PT; GSD1; GSD1a; MGC163350; G6PC	232200	NG_011808.1	17q21
HBB	3043	hemoglobin, beta	CD113-C; beta-globin; HBB	141900	NG_000007.3	11p15.5

Petitioner Illumina, Inc., EX1004, p. 507

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
BCKDHA	593	branched chain keto acid dehydrogenase E1, alpha polypeptide	MSU; MSUD1; OVD1A; BCKDE1A; FLJ45695; BCKDHA	608348	NC_000019.8	19q13.1-q13.2
BCKDHB	594	branched chain keto acid dehydrogenase E1, beta polypeptide	E1B; FLJ17880; dJ279A18.1; BCKDHB	248611	NG_009775.1	6q13-q15
DBT	1629	dihydrolipoamide branched chain transacylase E2	E2; E2B; BCATE2; MGC9061; DBT	248610	NG_011852.1	1p31
DLD	1738	dihydrolipoamide dehydrogenase	E3; LAD; DLDH; GCSL; PHE3; DLD	238331	NG_008045.1	7q31-q32
ACADM	34	acyl-Coenzyme A dehydrogenase, C-4 to C-12 straight chain	MCAD; ACAD1; MCADH; FLJ18227; FLJ93013; FLJ99884; ACADM	607008	NG_007045.1	1p31
MMAA	166785	methylmalonic aciduria (cobalamin deficiency) cblA type	cblA; MGC120010; MGC120011; MGC120012; MGC120013; MMAA	607481	NG_007536.1	4q31.22
MMAB	326625	methylmalonic aciduria (cobalamin deficiency) cblB type	ATR; cblB; MGC20496; MMAB	607568	NG_007096.1	12q24
MUT	4594	methylmalonyl Coenzyme A mutase	MCM; MUT	609058	NG_007100.1	6p12.3
MCOLN1	57192	mucolipin 1	ML4; MLIV; MST080; TRPML1; MSTP080; TRP-ML1; TRPM-L1; MCOLN1	605248	NC_000019.8	19p13.3-p13.2
ACTA1	58	actin, alpha 1, skeletal muscle	ACTA; ASMA; CFTD; MPFD; NEM1; NEM2; NEM3; CFTD1; CFTDM; ACTA1	102610	NG_006672.1	1q42.13
TPM3	7170	tropomyosin 3	TM3; TRK; NEM1; TM-5; TM30; TM30nm; TPMsk3; hscp30; MGC3261; FLJ41118; MGC14582; MGC72094; OK/SW-cl.5; TPM3	191030	NG_008621.1	1q21.2
TNNT1	7138	troponin T type 1 (skeletal, slow)	ANM; TNT; STNT; TNTS; FLJ98147; MGC104241; TNNT1	191041	NG_011829.1	19q13.4
NEB	4703	nebulin	NEM2; NEB177D; FLJ11305; FLJ36536; FLJ39568; FLJ39584; DKFZp686C1456; NEB	161650	NG_009382.1	2q22
SMPD1	6609	sphingomyelin phosphodiesterase 1, acid lysosomal	ASM; NPD; SMPD1	607608	NG_011780.1	11p15.4-p15.1
GLDC	2731	glycine dehydrogenase (decarboxylating)	GCE; NKH; GCSP; HYGN1; MGC138198; MGC138200; GLDC	238300	NC_000009.10	9p22

Petitioner Illumina, Inc., EX1004, p. 508

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
GCSH	2653	glycine cleavage system protein H (aminomethyl carrier)	GCE; NKH; GCSH	238330	NC_000016.8	16q23.2
AMT	275	aminomethyltransferase	GCE; NKH; GCST; AMT	238310	NC_000003.10	3p21.2-p21.1
OTC	5009	ornithine carbamoyltransferase	OCTD; MGC129967; MGC129968; MGC138856; OTC	300461	NG_008471.1	Xp21.1
PAH	5053	phenylalanine hydroxylase	PH; PKU; PKU1; PAH	612349	NG_008690.1	12q22-q24.2
DHPR	5860	quinoid dihydropteridine reductase	DHPR; PKU2; SDR33C1; FLJ42391; QDPR	612676	NG_008763.1	4p15.31
PTS	5805	6-pyruvoyltetrahydropterin synthase	PTPS; FLJ97081; PTS	261640	NG_008743.1	11q22.3-q23.3
PCCA	5095	propionyl Coenzyme A carboxylase, alpha polypeptide	PCCA	232000	NG_008768.1	13q32
PCCB	5096	propionyl Coenzyme A carboxylase, beta polypeptide	DKFZp451E113; PCCB	232050	NG_008939.1	3q21-q22
ACADS	35	acyl-Coenzyme A dehydrogenase, C-2 to C-3 short chain	SCAD; ACAD3; ACADS	606885	NG_007991.1	12q22-qter
DHCR7	1717	7-dehydrocholesterol reductase	SLOS; DHCR7	602858	NC_000011.8	11q13.2-q13.5
SMN1	6606	survival of motor neuron 1, telomeric	SMA; SMN; SMA1; SMA2; SMA3; SMA4; SMA@; SMNT; BCD541; T-BCD541; SMN1	600354	NG_008691.1	5q13
HEXA	3073	hexosaminidase A (alpha polypeptide)	TSD; MGC99608; HEXA	606869	NG_009017.1	15q23-q24
MYO7A	4647	myosin VIIA	DFNB2; MYU7A; NSRD2; USH1B; DFNA11; MYOVIIA; MYO7A	276903	NG_009086.1	11q13.5
USH1C	10083	Usher syndrome 1C (autosomal recessive, severe)	PDZ73; AIE-75; DFNB18; PDZ-45; PDZ-73; NY-CO-37; NY-CO-38; ush1cpst; PDZ-73/NY-CO-38; USH1C	605242	NC_000011.8	11p15.1-p14
CDH23	64072	cadherin-like 23	USH1D; DFNB12; FLJ00233; FLJ36499; KIAA1774; KIAA1812; MGC102761; DKFZp434P2350; CDH23	605516	NG_008835.1	10q21-q22
PCDH15	65217	protocadherin 15	USH1F; DFNB23; DKFZp667A1711; PCDH15	605514	NG_009191.1	10q21.1

Petitioner Illumina, Inc., EX1004, p. 509

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
SANS	124590	Usher syndrome 1G (autosomal recessive)	SANS; ANKS4A; FLJ33924; USH1G	607696	NG_007882.1	17q25.1
ARX	170302	aristaless related homeobox	ISSX; PRTS; MRX29; MRX32; MRX33; MRX36; MRX38; MRX43; MRX54; MRX76; MRX87; MRXS1; ARX	300382	NG_008281.1	Xp21
OPHN1	4983	oligophrenin 1	OPN1; MRX60; OPHN1	300127	NG_008960.1	Xq12
JARID1C	8242	lysine (K)-specific demethylase 5C	MRXJ; SMCX; MRXSJ; XE169; JARID1C; DXS1272E; KDM5C	314690	NG_008085.1	Xp11.22-p11.21
FTSJ1	24140	FtsJ homolog 1 (E. coli)	JM23; MRX9; SPB1; TRM7; CDLIV; MRX44; FTSJ1	300499	NG_008879.1	Xp11.23
SLC6A8	6535	solute carrier family 6 (neurotransmitter transporter, creatine), member 8	CRT; CT1; CRTR; MGC87396; SLC6A8	300036	NC_000023.9	Xq28
DLG3	1741	discs, large homolog 3 (Drosophila)	MRX; MRX90; NEDLG; NE-Dlg; SAP102; SAP-102; KIAA1232; DLG3	300189	NC_000023.9	Xq13.1
TM4SF2	7102	tetraspanin 7	AI5; MXS1; CD231; MRX58; CCG-B7; TM4SF2; TALLA-1; TM4SF2b; DXS1692E; TSPAN7	300096	NG_009160.1	Xp11.4
ZNF41	7592	zinc finger protein 41	MRX89; MGC8941; ZNF41	314995	NG_008238.1	Xp11.23
FACL4	2182	acyl-CoA synthetase long-chain family member 4	ACS4; FACL4; LACS4; MRX63; MRX68; ACSL4	300157	NG_008053.1	Xq22.3-q23
PQBP1	10084	polyglutamine binding protein 1	SHS; MRX55; MRXS3; MRXS8; NPW38; RENS1; PQBP1	300463	NC_000023.9	Xp11.23
PEX1	5189	peroxisomal biogenesis factor 1	ZWS1; PEX1	602136	NG_008341.1	7q21.2
PXMP3	5828	peroxisomal membrane protein 3, 35kDa	PAF1; PEX2; PMP3; PAF-1; PMP35; RNF72; PXMP3	170993	NG_008371.1	8q21.1
PEX6	5190	peroxisomal biogenesis factor 6	PAF2; PAF-2; PXAAA1; PEX6	601498	NG_008370.1	6p21.1
PEX10	5192	peroxisomal biogenesis factor 10	NALD; RNF69; MGC1998; PEX10	602859	NG_008342.1	1p36.32
PEX12	5193	peroxisomal biogenesis factor 12	PAF-3; PEX12	601758	NG_008447.1	17q12
PEX5	5830	peroxisomal biogenesis factor 5	PXR1; PTS1R; PTS1-BP; FLJ50634; FLJ50721; FLJ51948; PEX5	600414	NG_008448.1	12p13.31
PEX26	55670	peroxisomal biogenesis factor 26	FLJ20695; PEX26M1T; Pex26pM1T; PEX26	608666	NG_008339.1	22q11.21

Petitioner Illumina, Inc., EX1004, p. 510

The mutations listed in Table 2 are documented polymorphisms in several disease-associated genes (CFTR is mutated in cystic fibrosis, GBA is mutated in Gaucher disease, ASPA is mutated in Canavan disease, HEXA is mutated in Tay Sachs disease). The polymorphisms are of several types: insertion/deletion polymorphisms which will cause frameshifts (and thus generally interrupt protein function) unless the insertion/deletion length is a multiple of 3bp, and substitutions which can alter the amino acid sequence of the protein and in some cases cause complete inactivation by introduction of a stop codon.

Table 2: Non-limiting examples of polymorphisms

Gene name	GeneID	SNP ID	Mutation	SEQ ID NO:
CFTR	1080	rs63500661	TCACATCACCAAGTTAAAAAAAAAAA[A/G]G GGGCGGGGGGGCAGAATGAAAATT	1
CFTR	1080	rs63107760	AAACAAGGATGAATTAAGTTTTTTTT[-/T]AAAAAAGAAACATTTGGTAAGGGGA	2
CFTR	1080	rs62469443	ATCACCAAGTTAAAAAAAAAAAAGGG[A/G]C GGGGGGGCAGAATGAAAATTGCAT	3
CFTR	1080	rs62469442	CTATTGAACCAGAACCAAACAGGAAT[A/G]C CATAGCATTTTGTAAACTAAACTG	4
CFTR	1080	rs62469441	CAGGAGTTCAAGACCAGCCTACTAAA[A/C]C ACACACACACACACACACACACAC	5
CFTR	1080	rs62469439	GATTAAATAATAGTGTTTATGTACCC[C/G]GC TTATAGGAGAAGAGGGTGTGTGT	6
CFTR	1080	rs62469438	ATTGTTATCTTTTCATATAAGGTAAC[A/T]GA GGCCCAGAGAGATTAAATAACAT	7
CFTR	1080	rs62469437	TAATTTTAATTAAGTAAATTTAATTG[A/G]TA GATAAATAAGTAGATAAAAAATA	8
CFTR	1080	rs62469436	GTATAAAAAAAAAAAAAAAAAAAGTT[A/T]G AATGTTTTCTTGCATTCAGAGCCT	9
CFTR	1080	rs62469435	ATACTAAAAATTTAAAGTTCTCTTGC[A/G]AT ATATTTTCTTAATATCTTACATC	10
CFTR	1080	rs62469434	TGCTGGGATTACAGGCGTGAGCCACC[A/G]C GCCTGGCCTGATGGGACATATTTT	11
CFTR	1080	rs62469433	CTACAATATAAGTATAGTATTGCAAA[A/C]CC ATCAGGAAGGGTGTTAACTATTT	12
CFTR	1080	rs61763210	GTTGTCTCCAAACTTTTTTTCAGGTG[-/AGA]AGGTGGCCAACCGAGCTTCGGAAAG	13
CFTR	1080	rs61720488	TTTTTTCATAAAAGATTATATAAAGG[A/C]TA TTGCTTTTGAATCACAAACACTA	14
CFTR	1080	rs61481156	ATCTAGTGAGCAGTCAGGAAAGAGAA[C/T]T TCCAGATCCTGGAAATCAGGGTTA	15
CFTR	1080	rs61443875	TAGAGTATAAAAAAAAAAAAAAAAAA[-/A]GTTTGAATGTTTTCTTGCATTCAGA	16

Petitioner Illumina, Inc., EX1004, p. 511

CFTR	1080	rs61312222	TGCAAATGCCAACTATCAAAGATATT[C/G]GA GTATACTGTCAATAAACTTCATA	17
CFTR	1080	rs61159372	TCCTCAACAGTTAGAAACAATATTTT[C/G]AG TGATTTCCCATGCCAACTTTACT	18
CFTR	1080	rs61094145	TTTTTGGTATTGTTGTTAAATAAGTG[A/G]GA ATTCAATACAGTATAATGTCTGT	19
CFTR	1080	rs61086387	CTTGAAATCGGATATATATATATATATA[- /TGTATATATATATATATATATATATATAT ACATATATATATATA]GTATTATCCCTGTTTTC ACAGTTTT	20
CFTR	1080	rs60996744	AGAGGGGCTGTGAAGGACACCAAGGA[A/G]G AGACTAAGAGCCAGGAGGGAAAAC	21
CFTR	1080	rs60960860	TAGAGTTTATTAGCTTTTACTACTCT[A/G]CTT AGTTACTTTGTGTTACAGAATA	22
CFTR	1080	rs60923902	ACTAGTGATGATGAGCTTCTTTTCAT[- /AT]GTTTGTTGGCTGCATAAATGTCTTC	23
CFTR	1080	rs60912824	GCAGAGAAAAGAGGGGCTGTGAAGGA[C/G]A CCAAGGAGGAGACTAAGAGCCAGG	24
CFTR	1080	rs60887846	TTCAGAGGTCTACCACTGGTGCATAC[G/T]CT AATCACAGTGTCGAAAATTTTAC	25
CFTR	1080	rs60793174	AAGAAAGAGCAAAAGAGGGCAAACTT[C/T]T CATACATTTTTGATGTCGAAACCA	26
CFTR	1080	rs60788575	CCTAAAGTTTAAAAAGAAAAAAAAAA[- /A]GGAAGAAGGAATTAAAAATCCAAAG	27
CFTR	1080	rs60760741	GTGTGTGTGTGTATATATATATATATAT[A/T]TA TATATTTTTTTTTTCCTGAGCCA	28
CFTR	1080	rs60456599	AAACTGTTGATGTTTTCATTTATTTA[C/G]ATC ATTGGAAAACTTTAGATTCTAG	29
CFTR	1080	rs60363249	TTTATCCATTCTTAACCAGAACAGAC[A/G]TT TTTTCAGAGCTGGTCCAGGAAAA	30
CFTR	1080	rs60355115	TTGAAATCGGATATATATATATATATAT[A/G]TA TATATATATATATATATATATAT	31
CFTR	1080	rs60308689	TAGTTTTTTATTTCCTCATATTATTT[- /T]CAGTGGCTTTTTCTTCCACATCTTT	32
CFTR	1080	rs60271242	ACATAGTTCTCAGTGGTACAACTACA[A/G]GT GATTTCTCTTTTCTTATTTCTGG	33
CFTR	1080	rs60010318	AGAGCAATGGCATCCCTTGTCTTGTG[C/T]TA TACAGGATGCAGCAATTTATAGG	34
CFTR	1080	rs59961323	TTCTGTCTACATAAGATGTCATACTA[A/G]AT TATCTTTTCCAGCATGCATTCAG	35
CFTR	1080	rs59961270	CAGGGTGGCATGTTAGGCAGTGCTTA[A/G]A ATAAATGAGTTGGTTATACAAGTA	36
CFTR	1080	rs59837506	AGGACACACACACACACACACACACA[- /CA]TGCACACACATTTAAATAGATGCAT	37
CFTR	1080	rs59572090	TAAAAAATTGGTATAATGAAATTGCA[C/T]TT GTAGTCTTTGGACATTTAAATCC	38
CFTR	1080	rs59548252	TTTCAATACTTAAGAGGTACGCAGAG[A/G]A AAGAGGGGCTGTGAAGGACACCAA	39
CFTR	1080	rs59519859	CAGCAATGAATATTTTGAGGCTGAGG[C/T]GC TGAGGGGTAAAATTGCAGCCTGG	40
CFTR	1080	rs59509837	TTATGGTTTATATTTTTGTGTCTTCT[- /CTTT]AACACATCTTTTCTAGCAGAATTCA	41

Petitioner Illumina, Inc., EX1004, p. 512

CFTR	1080	rs59417037	GTATTTTAGTTTTTTTTTTTTGTTTG[-/T]TTTGTTTTGTTTTGTTTTGTTTTTG	42
CFTR	1080	rs59159458	TGGGTGACTCCATTTTTACTTTTAGT[C/T]TGG TCTGTTGAGGCCTCGTGAGAGA	43
CFTR	1080	rs59048119	TATTTTCATGTATTTTAGTTTTTTTTT[-/TTTT]GTTTGTTTTGTTTTGTTTTGTTTTG	44
CFTR	1080	rs58970500	GTGTGTGTGTATATATATATATATAT[A/T]TA TATTTTTTTTTTCCTGAGCCAAA	45
CFTR	1080	rs58942292	AACCTATTAGCATGTCTGGCAGAAAA[-/A]TAGATACTTAATAAATTTCTTAAAT	46
CFTR	1080	rs58917054	GAGGCTTAGACAGTTTAAGTAACTCA[A/G]G CATGGTTACACAACTAGCTAGGGC	47
CFTR	1080	rs58837484	GTGTGAGTATTATGAGACCATATGTT[A/G]GG AGATTTTATTTGGTATTGAGGAT	48
CFTR	1080	rs58829491	GAAACCCCACCCCTTCTATAGTTTTC[C/T]CTT TAATATTTACAATGGAACCATT	49
CFTR	1080	rs58805195	CATATATATATAGTGTGTGTGTGTGT[A/G]TA TATATATATATATATATATTTTT	50
GBA	2629	rs60866785	CGAGCGAGAGAGAGAGAGAGAGAGAG[-/AG]GAGCCGGCGCGAGAACTACGCATGC	51
GBA	2629	rs60239603	GGCAGGTAATATCTAGTACCTTACTT[A/T]TA TTTCCTGAGCACATTCTACATTT	52
GBA	2629	rs56310840	GGCCAGGAATGGGAGTGCTTAGGTGC[A/G]G AGGTGGCACTGTTCCCGCAGCTGC	53
GBA	2629	rs41264927	TGAAAACTCCATCCCCTCAGGGTCAT[C/T]AG ATGAAGAGAAGACCACAGGGGTT	54
GBA	2629	rs41264925	TGTAGGTAAGGGTCACATGTGGGAGA[C/G]G CAGCTGTGGGTAGGTCAGCCCTGT	55
GBA	2629	rs36024691	CCAAGAAGGCGCCATTACACTCCAGC[-/C]TGGGCGACAGGGCGAGACTCCCTCA	56
GBA	2629	rs36024092	TGCCACACCCAGCTAATTTGTGTGTG[-/G]TATGTGTGTGTATGTATGTGTGTGT	57
GBA	2629	rs35682967	GTTCCTCCAGTAATTTTTTTTTTTTT[-/T]CGTTTTGAGACAGAGTCTTGCCCTG	58
GBA	2629	rs35033592	ATCATGCCCAGATAATTTTTTTTTTT[-/T]GTATTTTAGTAGACACAGGGTTTCA	59
GBA	2629	rs34732744	CGAGCGAGAGAGAGAGAGAGAGAGAG[-/AG]GAGCCGGCGCGAGAACTACGCATGC	60
GBA	2629	rs34620635	CCTGTGAGGGGCACATTCCTTAGTAG[-/C]TAAGGAGTTGGGGGTGTGAAGATCC	61
GBA	2629	rs34302637	ACAGGCTACTGGCTGGGCCCAGGCAA[-/A]GGGGGCCTTGGCAGGAAAAGTTCCT	62
GBA	2629	rs33949225	GCGAGAGAGAGAGAGAGAGAGAGAGG[-/AG]AGCCGGCGCGAGAACTACGCATGCG	63
GBA	2629	rs28678003	AAGAAGAAAAATAAAAAGAAAGTGGG[C/T]C AGACCGAGAGAACAGGAAGCCTGA	64
GBA	2629	rs28559737	AAGGACAAAGGCAAAGAGACAAAGGC[G/T]C AACACTGGGGGTCCCCAGAGAGTG	65
GBA	2629	rs28373017	TACCTAGTCACTTCCTGCCTCCATGG[C/T]GC AAAAGGGGATGGGTGTGCCTCTT	66
GBA	2629	rs12752133	CTCTTCCGAGGTTCCACCCTGAACAC[C/T]TT CCTGCTCCCTCGTGGTGTAGAGT	67

Petitioner Illumina, Inc., EX1004, p. 513

GBA	2629	rs12747811	TTCTGACTGGCAACCAGCCCCACTCT[C/T]TG GGAGCCCTCAGGAATGAACTTGC	68
GBA	2629	rs12743554	gctcagcctcccaggctggagtgcag[A/T]ggcgcgatctcggctcac cgcaacc	69
GBA	2629	rs12041778	CATGAACCACATCAAATGAGATTTAG[C/T]GG GAGTGGCACACACAGTCATGACC	70
GBA	2629	rs12034326	AAGCAGCCCTGGGGAGTCGGGGCGGG[A/G]C CTGGATTGGAAAAGAGACGGTCAC	71
GBA	2629	rs11558184	CTCCAAGTTCTGGGAGCAGAGTGTGC[A/G]G CTAGGCTCCTGGGATCGAGGGATG	72
GBA	2629	rs11430678	GTTCCTCCAGTAAttttttttttt[- /G/T]gtttgagacagagtcttgccctgt	73
GBA	2629	rs11264345	CTAGTACCTTACTTCCCTCAAGTTCA[A/T]TC ATCTCACAGATATTTCCTGAGCA	74
GBA	2629	rs10908459	aattagccgtgcggtggtggcgggtgc[C/T]gtaatcccacgtacttgg gaggct	75
GBA	2629	rs10796940	CCATGGCCAGCCGGGGAGGGGACGGG[A/C]A CACACAGACCCACACAGAGACTCA	76
GBA	2629	rs10668496	agcgagagagagagagagagagagag[- /AG]gagCCGGCGCGAGAACTACGCATGC	77
GBA	2629	rs7416991	CGTAGCAGTTAGCAGATGATAGGCGG[C/G/T] GAAATCTTATTTCACAGGGCATTAA	78
GBA	2629	rs4024049	CTGGCCCTGGTGACAGTGGGGCTGTG[C/T]GT GGGGCCAGAGCCTTCTCAGAGGT	79
GBA	2629	rs4024048	CAGATACTGGCCCTGGTGACAGTGGG[A/G]C TGTGCGTGGGGCCAGAGCCTTCTC	80
GBA	2629	rs4024047	GACAGATACTGGCCCTGGTGACAGTG[G/T]G GCTGTGCGTGGGGCCAGAGCCTTC	81
GBA	2629	rs3841430	GGCTCctctctctctctctctctc[- /TC]gctcgctctctcgctctctcgctct	82
GBA	2629	rs3754485	GTTTCAGACCAGCCTGGCCAACATAG[C/T]GA AACCCCATCTCTACTAAAAATAA	83
GBA	2629	rs3205619	AGTGGGCGATTGGATGGAGCTGAGTA[C/T]G GGGCCCATCCAGGCTAATCACACC	84
GBA	2629	rs2990227	CCGGGCTCCGTGAATGTTTGTCACAT[C/G]TC TGAAGAACGTATGAATTACATAA	85
GBA	2629	rs2990226	GAATCCCAACCCCGACGCTCGTCGCC[C/G]G GCTCCGTGAATGTTTGTCACATGT	86
GBA	2629	rs2990225	GCGAATCCCAACCCCGACGCTCGTCG[C/T]CG GGCTCCGTGAATGTTTGTCACAT	87
GBA	2629	rs2990224	TGGGCAGAAGTCAGGGTCCAAAGAAA[G/T]G GCAAAGAAAAGTGTcagtggctca	88
ASPA	443	rs63751297	TAAGAAAGACGTTTTTGATTTTTTTC[A/G]GA CTTCTCTGGCTCCACTACCCTGC	89
ASPA	443	rs62071301	CTGATTCCTGGCCAGGAGCGGTGGCT[C/T]AC GCCTGTAATCCCAGCGCTTTGGG	90
ASPA	443	rs62071300	TAAAAATGCTGATTCCTGGCCAGGAG[C/T]GG TGGCTCACGCCTGTAATCCCAGC	91
ASPA	443	rs62071299	TTTAAAAATGCTGATTCCTGGCCAGG[A/C]GC GGTGGCTCACGCCTGTAATCCCA	92
ASPA	443	rs62071297	CAAGACCTGTCAAAGATCTGAGAAAT[A/T]TT ACCCGACTTACAAGCTAACCATT	93

Petitioner Illumina, Inc., EX1004, p. 514

ASPA	443	rs61697033	ACTGTAATAAGTGCTGTAAAAGAAAT[A/G]C ACAAAATAATATAGCAGAGGGTAT	94
ASPA	443	rs60743592	CTTGAGGTCAGGAGTTCAAGACCAGT[C/T]TG GGCAACATGGGGAAAACCTTGTC	95
ASPA	443	rs60666840	AGGTTGCAGTGAGCCGAGATCATGCC[A/G]TT GCACTCCAGCCGGGGCAACAAAA	96
ASPA	443	rs60147514	ACAAGTGTCTTGAAATTATCTGTGAT[C/T]TG CTATAGAGCAATACTTTTGTAAA	97
ASPA	443	rs59930743	GTGGGTATATGCAGCTCTATGCACTA[C/T]CT GCTCATTTATTTGGTAAATCTAA	98
ASPA	443	rs59690349	TGTGTGTGTGTGCGTGTGTGTGTGTG[-/TGTGTGTG]ATCATAAGAGTGGCTGCAGCAA ACT	99
ASPA	443	rs59676360	AGTCTGGAGTGCAATGGTGCAATCTC[A/G]GC TCACTGCAGCCTCCACCTCCGGG	100
ASPA	443	rs59335404	CTCCTAATGGATATTTCCTAAATTTT[G/T]CTG AACAGAATTTAACTTGAGCTGG	101
ASPA	443	rs58879097	ATTTAAAAATGGATTTCTAGAAAAAC[A/G]AT CACATACTTGAATATTTTAGCAA	102
ASPA	443	rs58686774	CTATAAATGGGTAGCATGAGGGATTC[A/G]A GGAGGTGGCTGAAAGAAGCACGTA	103
ASPA	443	rs57511162	AAGAAACCAAGCATAGTAGAGTGTTA[A/G]A AAACCAAAGCAACTAAACAACTGT	104
ASPA	443	rs55859596	CGGGGCTCAGAACTTGTAACAGAAAA[A/T]T AAAATATACTCCACTCAAGGGAAT	105
ASPA	443	rs55742972	TACTACACTTCACGGATACTGTACTT[-/GTACTT]TTTTTCCAAATTGAAGGTTTTTGGC	106
ASPA	443	rs55640436	TTGTTTTTGTTTTTGTTTTTGTTTTT[-/GTTTTTGTTTTT]TGAGATGGAGTCTCGCTCT GTCGCC	107
ASPA	443	rs36225687	TTTGCCTTACTACACTTCACGGATAC[-/TGTACT]TGTACTTTTTTTCCAAATTGAAGGT	108
ASPA	443	rs36051310	GAGGTGGCTGAAAGAAGCACGTATCC[-/C]TGATGGCATGGTTGCGGGTTATATG	109
ASPA	443	rs36034906	GAGAAAAGCAGTTCCTGGAACACCCC[-/C]ACCCCTTAACCCCTTATCTCTGCTT	110
ASPA	443	rs36033666	TTACATATGTATACATGTGCCATGTT[-/T]GGTGTGCCGCACCCATTAACTCGTC	111
ASPA	443	rs35730123	CTTTTTCCAGATTTTTTTTTTTTTTT[-/T]GAGACAGAGTTTCACTCTTGTTGCC	112
ASPA	443	rs35629100	TTTGGAAATCTTAAGCTTTTATTTGG[-/G]TGTCACAGAGAAACAGGATCTGTAT	113
ASPA	443	rs35614631	TACTTTAAGTTTTAGGGTACATGTGC[-/A]CCATGTGCAGGTTTGTTACATATGT	114
ASPA	443	rs35225782	ATTCATGACCAGCCACATAAATGCAC[-/A]GTATTACTTCGCAAGCATGCCAATG	115
ASPA	443	rs35178659	GTGCACTAGAATTAGCTAAAGTGGGG[-/G]AAAAAAAGATGCATTTGATGGTCTA	116
ASPA	443	rs35095578	AACCTCCACCTCCCAGGTTCAAGAGA[-/A]TTCTCCTGCCTCAGCCTCCCAAGTA	117
ASPA	443	rs35002210	CCTCCCTGTGATCCGAAGTAGCAGAC[A/G]TA CTTAACTTCCATGGTGGATTGTT	118

Petitioner Illumina, Inc., EX1004, p. 515

ASPA	443	rs34744839	AAAACATTATTATATCTAGAAAAAAA[-/A]TGTATCTTAACCATTGTGGGAAGTG	119
ASPA	443	rs34680506	TTGAAGGTAAAATCATAGGGAGTTGG[-/G]AGCTGTCCTCTTGCGCTGAATCAGT	120
ASPA	443	rs34365618	ACTTGTGGCCTTTTTGGAGAGGTTAG[-/CA]ACTCTGAAAACTCTGTCCCTGGACC	121
ASPA	443	rs34275920	GAAGGAGAAAAAGAGAGGAAATAAGT[-/T]AAAATAATAAACACAATTAATAAAG	122
ASPA	443	rs34109510	TGTATACATGTGCCATGTTGGTGTGC[C/T]GC ACCCATTAACTCGTCATTTAGCA	123
ASPA	443	rs34054576	TCACCTGTCACCTCCTATAGAACTTT[-/C]CCCTGACCCTCCTCTATAGCATTAA	124
ASPA	443	rs34015272	ATAAATGATCATCATTCACAGTAGGG[-/G]TTTTGTTTTGTTTTTTTTCTGGAAA	125
ASPA	443	rs34002091	ACAGACATATCTACAAACACACTTTT[-/T]CACATATTTGTGTAAGTCATTTATG	126
ASPA	443	rs28940574	AAAGACAACTAAACTAACGCTCAATG[A/C]A AAAAGTATTCGCTGCTGTTTACAT	127
ASPA	443	rs28940279	TACCGTGTACCCCGTGTTTGTGAATG[A/C]GG CCGCATATTACGAAAAGAAAGAA	128
ASPA	443	rs17850703	CAGGGCTGGAGGTAAAACCATTTATT[A/G]CT AACCCCAGAGCAGTGAAGAAGTG	129
ASPA	443	rs17222495	TTCTTCATTGCCTATTGAAGAGAGAG[C/T]GG AATGCTTTGGTTGCCAGATATGG	130
ASPA	443	rs17175228	CACAAGATCTCATTACTCAGGAGCTG[C/T]CC AAGTGTCTAATGTACTTAGTTAA	131
ASPA	443	rs16953074	TTCTGTGTAACATTTCATTTAAGCAA[A/G]GG ATTCGGCAAATCAAAAATTGTCA	132
ASPA	443	rs16953070	TAAAACGTATTGAAGGTATTATTGAC[G/T]CT GTTGAAGCAAAGAGAACAAAACA	133
HEXA	3073	rs62022858	ATCTGCTCTTCCAGTTGGATGACAAG[C/T]CT TGCTGTCTAACACCTGCTGCAGA	134
HEXA	3073	rs62022857	CCATTTTTTGTTGTATTTTTTTTTTC[C/T]TGAA TACTTTTTATCGCAGTTGGTT	135
HEXA	3073	rs62017872	CCCTGTCTCTAAAAGAAAAAAAAAAA[A/G]A AAAAAAAAAGAAAACAAAACCCAA	136
HEXA	3073	rs62017871	AGTGGCTCCAAAAAGGTCATGGAACC[C/T]CT TGAGGATGATGCAAATTGACTCT	137
HEXA	3073	rs61662730	TAAAGTTACTTTTCTTTTATTGACTT[C/T]CCC TTATTTTTTAACCTTATGCTTT	138
HEXA	3073	rs61329913	CAGAGTTAAAAAAAAAAAAAAAAAAA[-/A]GGAAGTAGCAGCAACAGCTTGGAAA	139
HEXA	3073	rs60920713	GTTGCCCAGGGTTGAGTGCAGAGGCA[C/T]AT CTTGGCTCACAGCAACCTCTGCC	140
HEXA	3073	rs60783213	AAGGCTTTTTTTTTTTTTTTTTTTT[-/TTTT]GAGACAGAGTCTTGCTGTGTCACCC	141
HEXA	3073	rs60644867	GCCTACATTCTGCAAAGAGGAGGGAA[C/G]A TTCACAGCTCCATACTTGAACCCT	142
HEXA	3073	rs60288568	CCAAAGGAGAATAGCTCTAGGGGAGG[C/G]A GGTGGATGAGTATGCATGGGGGAG	143
HEXA	3073	rs59888548	GACTCCATCTCAAAAAAAAAAAAAAA[-/A]TGCAGTCTAATGGCAGAATTAGACT	144

Petitioner Illumina, Inc., EX1004, p. 516

HEXA	3073	rs59733856	TTATTTATTTATTTATTTATTTTTGA[A/G]ACA GGGTCTCTGTTGTCCAGGCTGG	145
HEXA	3073	rs59427837	TTTTGAGGCAGGGTCTCACTCTGTTG[C/T]CC AGGGTTGAGTGCAGAGGCACATC	146
HEXA	3073	rs59171976	CGCCTTGCGAAGGCCCCACAGCTTGC[C/T]TG TGACAAACGTTCATAGGCAAATG	147
HEXA	3073	rs58706602	GGAGGTCTGTACAAAGCACCACCTAC[C/T]TC ATGGGTCAGTTTCCACAGCAGAA	148
HEXA	3073	rs58696963	GAATCTTATAATTCACTGTGTACCTC[-/CCTC]TGTTTCATATTTTCGCAATTGAACT	149
HEXA	3073	rs58610850	AACATAGTATCTAATATAGCTTTACA[C/T]CC AAAGCCAAAATATGAATACACTG	150
HEXA	3073	rs58016062	TTGTTTTGTTTTGTTTGGGGGGGGGG[-/G]TTGTTTTTCTGAGAGGGAGTCTTGC	151
HEXA	3073	rs57733983	CATACCAAAGGGCAGCTGGAGGGATA[C/T]A GACGGAAGTCATGTGGAGAGTGAA	152
HEXA	3073	rs57476645	CAGGTGTGAGCCACCACGACCACCAA[A/T]T TAGCTCTTTTTACTCCTTCCCTTC	153
HEXA	3073	rs56870003	AGTGGTAGCTGATTTTGCTTCTGGAT[A/C]CT TTGCCACCTTCCCACTCTTTAAT	154
HEXA	3073	rs56338339	AAAGACCTGTTTCTTAAAAAAAAAAA[-/AGAAAAAAAAAAA]GAAAGAAAAGAAAAG AAAAAAACAG	155
HEXA	3073	rs55995352	TAAAAAATCTTTCAATGAGGAGATGT[C/T]CC CAGAGCAAGACAGCTGTAGGATG	156
HEXA	3073	rs55860138	AAAAGAAAAAAAAAAAAAAAAAAAA[-/A]GAAAACAAAACCCAAACCCATAAAG	157
HEXA	3073	rs55743646	CCTGTCTCTAAAAGAAAAAAAAAAAA[A/G]A AAAAAAAAGAAAACAAAACCCAAA	158
HEXA	3073	rs55665666	GTTATCATAGAAAAATATCACACTCT[-/GT]CTGTATCCCCACTTCCAGAAACTGT	159
HEXA	3073	rs36106892	CAGGAGCTCATAGAATTACATACAAT[-/C]TTTTTTTTTTTTTTTTGAGACAGCG	160
HEXA	3073	rs36091525	TTGAGAATCTTATAATTCACTGTGTA[-/CCTC]CCTCTGTTTCATATTTTCGCAATTG	161
HEXA	3073	rs35949555	CCACTACCACAGTGCCTAGAGAACAA[C/T]A TGTGTTTAATAATATTTAAATAAT	162
HEXA	3073	rs35827424	CCCTGTCTCTAAAAGAAAAAAAAAAA[-/A]AAAAAAAAAAGAAAACAAAACCCAA	163
HEXA	3073	rs35729578	CCATTATATCATTCATTTCCCACTCA[-/T]TTTCTTCATTCCAACCAAGATATAT	164
HEXA	3073	rs35649102	TCCGTCTCAAAAAAAAAAAAAAAAAG[-/A]GAAAGGAATTATTCTCATGTATACA	165
HEXA	3073	rs35118677	CTGGGGCAGTTAAAAAGAAAAACAAA[-/C]CCCTGGTCCCTGCCCTTGAGGAGAT	166
HEXA	3073	rs35005352	CTCCAGGGTCCCATTCCAGGACCACA[-/C]GCCTGCTACCTCTGCAGCTCACTCA	167
HEXA	3073	rs34736306	GGATTGACATATACCAGTTAGACGGA[-/T]TTTTTTTTTCCATAAACCAGGCTCA	168
HEXA	3073	rs34607939	ACAAATAATTACTACATATCTACAAC[A/G]TT CCAGATACAGAAGAAATGGCCAA	169

Petitioner Illumina, Inc., EX1004, p. 517

HEXA	3073	rs34496117	TAAACACACTTGAAACATCATATAAA[-/ATG]ATATTACTACAAGACTTAACCGTAA	170
HEXA	3073	rs34300017	ACACAGGTAATCCATGTTTATTATAG[-/A]AAAATGCCACATTACTCTTTATTGA	171
HEXA	3073	rs34206496	AGTTATCATAGAAAAATATCACACTC[-/TG]TCTGTATCCCCACTTCCAGAAACTG	172
HEXA	3073	rs34110830	AATGAACTTACAGGAAGGTAATATAT[-/G]GGAAATAAACATCTTATTGAATTTA	173
HEXA	3073	rs34093438	GGACCCCTGAAAGGCACAAGACACCC[-/T]TTCAGGTTCACACTTCCTGAAAGCT	174
HEXA	3073	rs34085965	CCACCAATCACCAGAGCCTTCTGCTC[A/G]GG GGTACCTGAGGGAAAACAAGCAA	175
HEXA	3073	rs34004907	AAAGACTGAAAAAACATTCATAACTA[-/T]TTTTCTTGTTATCCTCGGAAATGTC	176
HEXA	3073	rs28942072	TATCTTCATCTTGGAGGAGATGAGGT[C/T]GA TTTCACCTGCTGGAAGTCCAACC	177
HEXA	3073	rs28942071	TTGCCTATGAACGTTTGTCACACTTC[C/T]GCT GTGAGTTGCTGAGGCGAGGTGT	178
HEXA	3073	rs28941771	GCTTGCTGTTGGATACATCTCGCCAT[C/T]AC CTGCCACTCTCTAGCATCCTGGA	179
HEXA	3073	rs28941770	CCGGGGCTTGCTGTTGGATACATCTC[G/T]CC ATTACCTGCCACTCTCTAGCATC	180

3. Nucleic acid target length evaluation:

In some embodiments, aspects of the invention relate to methods for detecting nucleic acid deletions or insertions in regions containing nucleic acid sequence repeats.

Genomic regions that contain nucleic acid sequence repeats are often the site of genetic instability due to the amplification or contraction of the number of sequence repeats (e.g., the insertion or deletion of one or more units of the repeated sequence). Instability in the length of genomic regions that contain high numbers of repeat sequences has been associated with a number of hereditary and non hereditary diseases and conditions.

For example, "Fragile X syndrome, or Martin-Bell syndrome, is a genetic syndrome which results in a spectrum of characteristic physical, intellectual, emotional and behavioral features which range from severe to mild in manifestation. The syndrome is associated with the expansion of a single trinucleotide gene sequence (CGG) on the X chromosome, and results in a failure to express the FMR-1 protein which is required for normal neural development. There are four generally accepted forms of Fragile X syndrome which relate to the length of the repeated CGG sequence; Normal (29-31 CGG repeats), Premutation (55-200 CGG repeats), Full

Petitioner Illumina, Inc., EX1004, p. 518

Mutation (more than 200 CGG repeats), and Intermediate or Gray Zone Alleles (40 - 60 repeats)."

Other examples include cancer, which has been associated with microsatellite instability (MSI) involving an increase or decrease in the genomic copy number of nucleic acid repeats at one or more microsatellite loci (e.g., BAT-25 and/or BAT-26). The are currently many sequencing-based assays for determining the number of nucleic acid sequence repeats at a particular locus and identifying the presence of nucleic acid insertions or deletions. However, such techniques are not useful in a high throughput multiplex analysis where the entire length of a region may not be sequenced.

In contrast, in some embodiments, aspects of the invention relate to detecting the presence of an insertion or deletion at a genomic locus without requiring the locus to be sequenced (or without requiring the entire locus to be sequenced). Aspects of the invention are particularly useful for detecting an insertion or deletion in a nucleic acid region that contains high levels of sequence repeats. The presence of sequence repeats at a genetic locus is often associated with relatively high levels of polymorphism in a population due to insertions or deletions of one or more of the sequence repeats at the locus. The polymorphisms can be associated with diseases or predisposition to diseases (e.g., certain polymorphic alleles are recessive alleles associated with a disease or condition). However, the presence of sequence repeats often complicates the analysis of a genetic locus and increases the risk of errors when using sequencing techniques to determine the precise sequence and number of repeats at that locus.

In some embodiments, aspects of the invention relate to determining the size of a genetic locus by evaluating the capture frequency of a portion of that locus suspected of containing an insertion or deletion (e.g., due to the presence of sequence repeats) using a nucleic acid capture technique (e.g., a nucleic acid sequence capture technique based on molecular inversion probe technology). According to aspects of the invention, a statistically significant difference in capture efficiency for a genetic locus of interest in different biological samples (e.g., from different subjects) is indicative of different relative lengths in those samples. It should be appreciated that the length differences may be at one or both alleles of the genetic locus. Accordingly, aspects of the invention may be used to identify polymorphisms regardless of whether biological samples being interrogated at heterozygous or homozygous for the

Petitioner Illumina, Inc., EX1004, p. 519

polymorphisms. According to aspects of the invention, subjects that contain one or more loci with an insertion or deletion can be identified by analyzing capture efficiencies for nucleic acids obtained from one or more biological samples using appropriate controls (e.g., capture efficiencies for known nucleic acid sizes, capture efficiencies for other regions that are not suspected of containing an insertion or deletion in the biological sample(s), or predetermined reference capture efficiencies, or any combination thereof. However, it should be appreciated that aspects of the invention are not limited by the nature or presence of the control. In some embodiments, if a statistically significant variation in capture efficiency is detected, a subject may be identified as being at risk for a disease or condition associated with insertions or deletions at that genetic locus. In some embodiments, the subject may be analyzed in greater detail in order to determine the precise nature of the insertion or deletion and whether the subject is heterozygous or homozygous for one or more insertions or deletions. For example, gel electrophoresis of an amplification (e.g., PCR) product of the locus, or Southern blotting, or any combination thereof can be used as an orthogonal approach to verify the length of the locus. In some embodiments, a more exhaustive and detailed sequence analysis of the locus can be performed to identify the number and types of insertions and deletions. However, other techniques may be used to further analyze a locus identified as having an abnormal length according to aspects of the invention.

Accordingly, aspects of the invention relate to detecting abnormal nucleic acid lengths in genomic regions of interest. In some embodiments, the invention aims to estimate the size of genomic regions that are hard to be accessed, such as repetitive elements. However, it should be appreciated that methods of the invention do not require that the precise length be estimated. In some embodiments, it is sufficient to determine that one or more alleles with abnormal lengths are present at a locus of interest (e.g., based on the detection of abnormal capture efficiencies).

In a non-limiting example, fragile X can be used to illustrate aspects of the invention where the size of trinucleotide repeats (genotype) is linked to a symptom (phenotype). However, it should be appreciated that fragile X is a non-limiting example and similar analyses may be performed for other genetic loci (e.g., independently or simultaneously in multiplex analyses).

Use of molecular inversion probes (MIPs) has been demonstrated for detection of single nucleotide polymorphisms (Hardenbol et al. 2005 Genome Res 15:269-75) and for preparative amplification of large sets of exons (Porreca et al. 2007 Nat Methods 4:931-6, Krishnakumar et

Petitioner Illumina, Inc., EX1004, p. 520

al. 2008 Proc Natl Acad Sci USA 105:9296-301). In both cases, oligonucleotide probes are designed which have ends ('targeting arms') that hybridize up-stream and down-stream of the locus that is to be amplified.

In some embodiments, aspects of the invention are based on the recognition that the effect of length on probe capturing efficiency can be used in the context of an assay (e.g., a high throughput and/or multiplex assay) to allow the length of sequences to be determined without requiring sequencing of the entire region being evaluated. This is particularly useful for repeat regions that are prone to changes in size. As illustrated in FIG. 8, which is reproduced from Deng et al., Nature Biotech. 27:353-60, (see Supplemental Fig. 1G of Deng et al.,) illustrates that shorter sequences are captured with higher efficiency that longer sequences using MIPs. The statistical package R and its effects module were used for this analysis. A linear model was used, and each individual factor was assumed to be independent. The dashed lines represent a 95% confidence interval. Shorter target sequences were captured with higher efficiency than long target sequences ($p<2x10^{-16}$). However, the use of this differential capture efficiency for systematic sequence length analysis was not previously recognized.

In some embodiments, following probe hybridization, polymerase fill-in and ligation reactions are performed to convert the hybridized probe to a covalently-closed, circular molecule containing the desired target. PCR or rolling circle amplification plus exonuclease digestion of non-circularized material is performed to isolate and amplify the circular targets from the starting nucleic acid pool. Since one of the main benefits of the method is the potential for a high degree of multiplexing, generally thousands of targets are captured in a single reaction containing thousands of probes.

According to aspects of the invention, repetitive regions are surrounded by non repetitive unique sequences, which can be used to amplify the repeat-containing regions using, for example, PCR or padlock (MIP)-based method.

In addition to the repetitive regions, a probe (e.g., a MIP or padlock probe) can be designed to include at least a sequence that is sufficient to be uniquely identified in the genome (or target pool). After the probe is circularized and amplified, the amplicon can be end-sequenced so that the unique sequence can be identified and served as the "representative" of the repetitive region as illustrated in FIG. 9. FIG. 9 illustrates a non-limiting scheme of padlock (MIP) capture of a region that includes both repetitive regions (thick wavy line) and the adjacent

Petitioner Illumina, Inc., EX1004, p. 521

unique sequence (thick strait line). The regions of the probe are indicated with the targeting arms shown as regions "1" and "3." An intervening region that may be, or include, a sequencing primer binding site is shown as "2." After the padlock is circularized and amplified, it can be end-sequenced to obtain the sequence of the unique sequence, which represents the repetitive region of interest.

Although capturing efficiency is overall negatively correlated with target length, different probe sequences may have unique features. Therefore, multiple probes could be designed and tested so that an optimal one is chosen to be sensitive enough to differentiate repetitive sizes of roughly 0-150bp, 150-600bp, and beyond, which represent normal, premutation and full mutation of fragile X syndrome, respectively. However, it should be appreciated that other probe sizes and sequences can be designed, and optionally optimized, to distinguish a range of repeat region size differences (e.g., length differences of about 3-30 bases, about 30-60 bases, about 60-90 bases, about 90-120 bases, about 120-150 bases, about 150-300 bases, about 300-600 bases, about 600-900 bases, or any intermediate or longer length difference). It should be appreciated that a length difference may be an increase in size or a decrease in size.

In some embodiments, an initial determination of an unexpected capture frequency is indicative of the presence of size difference. In some embodiments, an increase in capture frequency is indicative of a deletion. In some embodiments, a decrease in capture frequency is indicative of an insertion. However, it should be appreciated that depending on specific sequence parameters and the relative sizes of the capture probes, the target region, and the deletions or insertions, a change in capture frequency can be associated with either an increase or decrease in target region length. In some embodiments, the precise nature of the change can be determined using one or more additional techniques as described herein.

Accordingly, in some aspects a MIP probe includes a linear nucleic acid strand that contains two hybridization sequences or targeting arms, one at each end of the linear probe, wherein each of the hybridization sequences is complementary to a separate sequence on a the same strand of a target nucleic acid, and wherein these sequences on the target nucleic acid flank the two ends of the target nucleic acid sequence of interest. It should be appreciated that upon hybridization, the two ends of the probe are inverted with respect to each other in the sense that both 5' and 3' ends of the probe hybridize to the same strand to separate regions flanking the target region (as illustrated in FIG. 9 for example).

Petitioner Illumina, Inc., EX1004, p. 522

In some embodiments, the hybridization sequences are between about 10-100 nucleotides long, for example between about 10-30, about 30-60, about 60-90, or about 20, about 30, about 40, or about 50 nucleotides long. However, other lengths may be used depending on the application. In some embodiments, the hybridization Tms of both targeting arms of a probe are designed or selected to be similar. In some embodiments, the hybridization Tms of the targeting arms of a plurality of probes designed to capture different target regions are selected or designed to be similar so that they can be used together in a multiplex reaction. Accordingly, a typical size of a MIP probe prior to fill in is about 60-80 nucleotides long. However, other sizes can be used depending on the sizes of the targeting arms and any other sequences (e.g., primer binding or tag sequences) that are present in the MIP probe. In some embodiments, MIP probes are designed to avoid sequence-dependent secondary structures. In some embodiments, MIP probes are designed such that the targeting arms do not overlap with known polymorphic regions. In some embodiments, targeting arms that can be used for capturing the repeat region of the Fragile X locus can have the following sequences or complementary to these sequences depending on the strand that is captured.

left: CTCCUITYCGGTITCACTTC (SEQ ID NO: 181)

right: ATCTTCTCTTCAGCCCTGCT (SEQ ID NO: 182)

The typical captured size using these targeting arms is about 100 nucleotides in length (e.g., about 30 repeats of a tri-nucleotide repeat).

In some embodiments, the number of reads obtained for the "representative" of the repetitive region is not informative to estimate the target length because it is dependent on the total number of reads obtained. To overcome this, it is useful to include one or more probes that target other "control" regions where no or minimal polymorphism exists among populations. Because of the systematic consistency of capturing efficiency (see, e.g., FIG. 9), the ratio of reads obtained for the repetitive "representative" to reads obtained for the control region(s) will be tuned using DNA with defined numbers of repeats. Ultimately, the ratio can serve as a measure of the repeat length as illustrated in FIG. 10. FIG. 10 illustrates a non-limiting hypothetical relationship between target gap size and the relative number of reads of the repetitive region, which is measured by the ratio of the repeat "representative" reads vs. the "control" region reads. The unit of y-axis is arbitrary.

Petitioner Illumina, Inc., EX1004, p. 523

In some embodiments, to better tell targets with similar size range apart, the whole repetitive region can be sequenced by making a shotgun library (e.g., by making a shotgun library from a captured sequence, for example a sequence captured using a MIP probe). The longer the repeat is, the more short reads of repeats will be obtained. Therefore, the target length will contribute twice to the relative number of "repetitive" reads, which will gain better resolution of differentiating targets. In some embodiments, the expectation is that the number of reads from any given repeat will be a direct function of the number of repeats present. However, in some embodiments, a Poisson sampling induced spread may need to be considered and in some embodiments may be sufficiently large to limit the resolution.

When a precise measurement of the length of both alleles from a diploid sample is desired, further manipulations may be required. This is because the capture efficiency measured will actually be the average efficiency of the two alleles. To effectively achieve separate measurements for each allele, barcodes (e.g., sequence tags) can be used that allow the efficiency of individual capture events (from individual genomic loci) to be followed. FIG. 11A-C shows the approach. For a given locus, MIPs are synthesized to contain one of a large number differentiator tags in their backbone such that the probability of any two MIPs in a reaction having the same differentiator tag sequence is low. MIP capture is performed on the sample; the reaction will be biased for shorter target lengths, and therefore the reaction product will be comprised of more 'short' circles than 'long' circles. Each circle should bear a unique differentiator tag sequence. Then, linear RCA (1RCA) is performed on the circles. In the 1RCA reaction, circles are converted into long, linear concatemers of themselves. The 1RCA reaction for a given circle stops when the concatemer has reached a 'fixed' length (based on the processivity/error rate of the polymerase). Concatemers derived from smaller circles will therefore contain more copies of the differentiator tag, and concatemers derived from larger circles will contain fewer copies of the differentiator tag. The number of each differentiator tag sequence is counted, for example, by next-generation sequencing.

When number of occurrences is plotted against differentiator tag ID, the data will naturally cluster into two groups reflecting the lengths of the two alleles in the diploid sample. The allele lengths can therefore be read directly off this graph, after absolute length calibration using known standards. In some embodiments, a sequencing technique (e.g., a next-generation sequencing technique) is used to sequence part of one or more captured targets (e.g., or

Petitioner Illumina, Inc., EX1004, p. 524

amplicons thereof) and the sequences are used to count the number of different barcodes that are present. Accordingly, in some embodiments, aspects of the invention relate to a highly-multiplexed qPCR reaction.

Other non-limiting examples of loci at which insertions or deletions or repeat sequences may be associated with a disease or condition are provided in Tables 3 and 4. It should be appreciated that the presence of an abnormal length at any one or more of these loci may be evaluated according to aspects of the invention. In some embodiments, two or more of these loci or other loci may be evaluated in a single multiplex reaction using different probes designed to hybridize under the same reaction conditions to different target nucleic acid in a biological sample.

Table 3.

Polyglutamine (PolyQ) Diseases

Type	Gene	Normal/wildtype	Pathogenic
DRPLA (Dentatorubropallidoluysian atrophy)	ATN1 or DRPLA	6 - 35	49 - 88
HD (Huntington's disease)	HTT (Huntingtin)	10 - 35	35+
SBMA (Spinobulbar muscular atrophy or Kennedy disease)	Androgen receptor on the X chromosome.	9 - 36	38 - 62
SCA1 (Spinocerebellar ataxia Type 1)	ATXN1	6 - 35	49 - 88
SCA2 (Spinocerebellar ataxia Type 2)	ATXN2	14 - 32	33 - 77
SCA3 (Spinocerebellar ataxia Type 3 or Machado-Joseph disease)	ATXN3	12 - 40	55 - 86
SCA6 (Spinocerebellar ataxia Type 6)	CACNA1A	4 - 18	21 - 30
SCA7 (Spinocerebellar ataxia Type 7)	ATXN7	7 - 17	38 - 120
SCA17 (Spinocerebellar ataxia Type 17)	TBP	25 - 42	47 - 63

Table 4.

Petitioner Illumina, Inc., EX1004, p. 525

Non-Polyglutamine Diseases

Type	Gene	Codon	Normal/wildtype	Pathogenic
FRAXA (Fragile X syndrome)	*FMR1*, on the X-chromosome	CGG	6 - 53	230+
FXTAS (Fragile X-associated tremor/ ataxia syndrome)	*FMR1*, on the X-chromosome	CGG	6 - 53	55-200
FRAXE (Fragile XE mental retardation)	AFF2 or FMR2, on the X-chromosome	GCC	6 - 35	200+
FRDA (Friedreich's ataxia)	FXN or X25, (frataxin)	GAA	7 - 34	100+
DM (Myotonic dystrophy)	DMPK	CTG	5 - 37	50+
SCA8 (Spinocerebellar ataxia Type 8)	OSCA or SCA8	CTG	16 - 37	110 – 250
SCA12 (Spinocerebellar ataxia Type 12)	PPP2R2B or SCA12	CAG On 5' end	7 - 28	66 - 78

The following examples illustrate aspects and embodiments of the invention and are not intended to be limiting or restrictive. Many variations of the invention will become apparent to those skilled in the art upon review of this specification. The full scope of the invention should be determined by reference to the claims, along with their full scope of equivalents, and the specification, along with such variations.

4. Increasing detection sensitivity:

In some embodiments, aspects of the invention relate to methods for increasing the sensitivity of nucleic acid detection assays.

There are currently many genomic assays that utilize next-generation (e.g., polony-based) sequencing to generate data, including genome resequencing, RNA-seq for gene expression, bisulphite sequencing for methylation, and Immune-seq, among others. In order to make

Petitioner Illumina, Inc., EX1004, p. 526

quantitative measurements (including genotype calling), these methods utilize the counts of sequencing reads of a given genomic locus as a proxy for the representation of that sequence in the original sample of nucleic acids. The majority of these techniques require a preparative step to construct a high-complexity library of DNA molecules that is representative of a sample of interest. Current assays use one of several alternative nucleic acid preparative techniques (e.g., amplification, for example PCR-based amplification; sequence-specific capture, for example, using immobilized capture probes; or target capture into a circularized probe followed by a sequence analysis step. In order to reduce errors associated with the unpredictability (stochastic nature) of nucleic acid isolation and sequence analysis techniques, current methods involve oversampling a target nucleic acid preparation in order to increase the likelihood that all sequences that are present in the original nucleic acid sample will be represented in the final sequence data. For example, a genomic sequencing library may contain an over- or under-representation of particular sequences from a source nucleic acid sample (e.g., genome preparation) as a result of stochastic variations in the library construction process. Such variations can be particularly problematic when they result in target sequences from a genome being absent or undetectable in a sequencing library. For example, an under-representation of particular allelic sequences (e.g., heterozygotic alleles) from a genome in a sequencing library can result in an apparent homozygous representation in a sequencing library.

In contrast, aspects of the invention relate to basing a nucleic acid sequence analysis on results from two or more different nucleic acid preparatory techniques that have different systematic biases in the types of nucleic acids that they sample rather than simply oversampling the target nucleic acid. According to some embodiments, different techniques have different sequence biases that are systematic and not simply due to stochastic effects during nucleic acid capture or amplification. Accordingly, in some embodiments, the degree of oversampling required to overcome variations in nucleic acid preparation needs to be sufficient to overcome the biases. In some embodiments, the invention provides methods that reduce the need for oversampling by combining nucleic acid and/or sequence results obtained from two or more different nucleic acid preparative techniques that have different biases.

According to the invention, different techniques have different characteristic or systematic biases. For example, one technique may bias a sample analysis towards one particular allele at a genetic locus of interest, whereas a different technique would bias the

Petitioner Illumina, Inc., EX1004, p. 527

sample analysis towards a different allele at the same locus. Accordingly, the same sample may be identified as being different depending on the type of technique that is used to prepare nucleic acid for sequence analysis. This effectively represents a sensitivity issue, because each technique has a different relative sensitivities for polymorphic sequences of interest.

According to aspects of the invention, the sensitivity of a nucleic acid analysis can be increased by combining the sequences from different nucleic acid preparative steps and using the combined sequence information for a diagnostic assay (e.g., for a making a call as to whether a subject is homozygous or heterozygous at a genetic locus of interest).

Currently, the ability of DNA sequencing to detect mutations is limited by the ability of the upstream sample isolation (e.g., by amplification, immobilization enrichment, circularization capture, etc.) methods to reliably isolate the locus of interest.

If one wishes to make heterozygote base-calls for a diploid genome (e.g. a human sample presented for molecular diagnostic sequencing), it is important in some embodiments that the isolation method produces near- or perfectly- uniform amounts of the two alleles to be sequenced (at least sufficiently uniform to be "called" unambiguously as a heterozygote or a homozygote for a locus of interest).

Sample preparative methods may fall into three classes: 1) single- or several target amplification (e.g., uniplex PCR, 'multiplex' PCR), 2) multi-target hybridization enrichment (e.g., Agilent SureSelect 'hybrid capture' [Gnirke et al 2009, *Nature methods* 27:182-9], Roche/Nimblegen 'sequence capture' [Hodges et al 2007, *Nature genetics* 39:1522-7], and 3) multi-target circularization selection (e.g. molecular inversion probes or padlock probes, [Porreca et al 2007, *Nature methods* 4:931-6, Turner et al 2009, *Nature methods* 6:315-6], 'selectors' [Dahl et al 2005, *Nucleic acids research* 33:e71]). Each of these methods can result in a pool of isolated product that does not adequately represent the input abundance distribution. For example, the two alleles at a heterozygous position can become skewed far from their input 50:50 ratio to something that results in a missed basecall during downstream sequencing. For example, if the ratio was skewed from 50:50 to 10:90, and the sample was sequenced to 10X average coverage, there is a high probability that one of the two alleles would not be observed once in the ten sequencing reads. This would reduce the sensitivity of the sequencing method by converting a heterozygous position to homozygous (where potentially the 'mutant' allele was the one not observed). In some embodiments, a skewed ratio is a particular issue that decreases the

Petitioner Illumina, Inc., EX1004, p. 528

sensitivity of detecting mutations present in a heterogeneous tumor tissue. For example, if only 10% of the cells analyzed in a heterogeneous sample harbored a heterozygous mutation, the mutation would be expected to be present in 5% of sequence reads, not 50%. In this scenario, the need for robust, sensitive detection may be even more acute.

The methods disclosed herein are based, in part, on the discovery that certain classes of isolation methods have different modes of bias. The disclosure provide methods for increasing the sensitivity of the downstream sequencing by using a combination of multiple isolation methods (e.g., one or more from at least two of the classes disclosed herein) for a sample. This is particularly important in molecular diagnostics where high sensitivity is required to minimize the chances of 'missing' a disease-associated mutation. For example, given a nominal false-negative error rate of $1x10^{-3}$ for sequencing following circularization selection, and a false-negative error rate of $1x10^{-3}$ for sequencing following hybridization enrichment, one can achieve a final false-negative rate of $1x10^{-6}$ by performing both techniques on the sample (assuming failures in each method are fully independent). For a recessive disease with carrier frequency of 0.1, caused by a single fully-penetrant mutant allele, the number of missed carrier diagnoses would decrease from 1000 per million patients tested to 1 per million patients tested. Furthermore, if the testing was used in the context of prenatal carrier screening, the number of affected children born as a result of missing the carrier call in one parent would decrease from 25 per million to 25 per billion born.

Additionally, the disclosure provides combinations of preparative methods to effectively increase sequencing coverage in regions containing disease-associated alleles. Since heterozygote error rate is largely tied to both deviations from 50:50 allele representation, and in the case of next-generation DNA sequencing deviations from average abundance (such that less abundant isolated targets are more likely to be undersampled at one or both alleles), selectively increasing coverage in these regions will also selectively increase sensitivity. Furthermore, MIPs that detect presence or absence of specific known disease-associated mutations can be used to increase sensitivity selectively. In some embodiments, these MIPs would have a targeting arm whose 3' -most region is complementary to the expected mutation, and has a fill-in length of 0 or more bp. Thus, the MIP will form only if the mutation is present, and its presence will be detected by sequencing.

Petitioner Illumina, Inc., EX1004, p. 529

Additionally, algorithms disclosed herein may be used to determine base identity with varying levels of stringency depending on whether the given position has any known disease-associated alleles. Stringency can be reduced in such positions by decreasing the minimum number of observed mutant reads necessary to make a consensus base-call. This will effectively increase sensitivity for mutant allele detection at the cost of decreased specificity.

An embodiment of the invention combines MIPs plus hybridization enrichment, plus optionally extra MIPs targeted to specific known, common disease-associated loci, e.g., to detect the presence of a polymorphism in a target nucleic acid. A non-limiting example is illustrated in FIG. 12 that illustrates a schematic using MIPs plus hybridization enrichment, plus optionally extra MIPs targeted to specific known, common disease-associated loci, e.g., to detect the presence of a polymorphism in a target nucleic acid.

FIGs. 13 and 14 illustrate different capture efficiencies for MIP-based captures. FIG. 13 shows a graph of per-target abundance with MIP capture. In this graph, bias largely drives the heterozygote error rate, since targets which are less abundant here are less likely to be covered in sufficient depth during sequencing to adequately sample both alleles. This is from Turner et al 2009, Nature methods 6:315-6. Hybridization enrichment results in a qualitatively similar abundance distribution, but the abundance of a given target is likely not correlated between the two methods. FIG. 14 shows a graph of correlation between two MIP capture reactions from Ball et al 2009, Nature biotechnology 27:361-8. Each point represents the target abundance in replicate 1 and replicate 2. Pearson correlation r=0.956. This indicates that MIP capture reproducibly biases targets to specific abundances. Hybridization enrichment is similarly correlated from one capture to the next.

According to aspects of the invention, such biases can be detected or overcome by systematically combining different capture and/or analytical techniques in an assay that interrogates a plurality of loci in a plurality of subject samples.

Accordingly, it should be appreciated that in any of the embodiments described herein (e.g., tiling/staggering, tagging, size-detection, sensitivity enhancing algorithms, or any combination thereof), aspects of the invention involve preparing genomic nucleic acid and/or contacting them with one or more different probes (e.g., capture probes, hybridization probes, MIPs, others etc.). In some embodiments, the amount of genomic nucleic acid used per subject ranges from 1 ng to 10 micrograms (e.g., 500 ng to 5 micrograms). However, higher or lower

Petitioner Illumina, Inc., EX1004, p. 530

amounts (e.g., less than 1 ng, more than 10 micrograms, 10-50 micrograms, 50-100 micrograms or more) may be used. In some embodiments, for each locus of interest, the amount of probe used per assay may be optimized for a particular application. In some embodiments, the ratio (molar ratio, for example measured as a concentration ratio) of probe to genome equivalent (e.g., haploid or diploid genome equivalent, for example for each allele or for both alleles of a nucleic acid target or locus of interest) ranges from 1/100, 1/10, 1/1, 10/1, 100/1, 1000/1. However, lower, higher, or intermediate ratios may be used.

In some embodiments, the amount of target nucleic acid and probe used for each reaction is normalized to avoid any observed differences being caused by differences in concentrations or ratios. In some embodiments, in order to normalize genomic DNA and probe, the genomic DNA concentration is read using a standard spectrophotometer or by fluorescence (e.g., using a fluorescent intercalating dye). The probe concentration may be determined experimentally or using information specified by the probe manufacturer.

Similarly, once a locus has been captured (e.g., on a MIP or other probe or in another form), it may be amplified and/or sequenced in a reaction involving one or more primers. The amount of primer added for each reaction can range from 0.1 pmol to 1 nmol, 0.15 pmol to 1.5 nmol (for example around 1.5 pmol). However, other amounts (e.g., lower, higher, or intermediate amounts) may be used.

In some embodiments, it should be appreciated that one or more intervening sequences (e.g., sequence between the first and second targeting arms on a MIP capture probe), identifier or tag sequences, or other probe sequences that are not designed to hybridize to a target sequence (e.g., a genomic target sequence) should be designed to avoid excessive complementarity (to avoid cross-hybridization) to target sequences or other sequences (e.g., other genomic sequences) that may be in a biological sample. For example, these sequences may be designed have a sufficient number of mismatches with any genomic sequence (e.g., at least 5, 10, 15, or more mismatches out of 30 bases) or as having a Tm (e.g., a mismatch Tm) that is lower (e.g., at least 5, 10, 15, 20, or more degrees C lower) than the hybridization reaction temperature.

It should be appreciated that a targeting arm as used herein may be designed to hybridize (e.g., be complementary) to either strand of a genetic locus of interest if the nucleic acid being analyzed is DNA (e.g., genomic DNA). However, in the context of MIP probes, whichever strand is selected for one targeting arm will be used for the other one. However, in the context of

Petitioner Illumina, Inc., EX1004, p. 531

RNA analysis, it should be appreciated that a targeting arm should be designed to hybridize to the transcribed RNA. It also should be appreciated that MIP probes referred to herein as "capturing" a target sequence are actually capturing it by template-based synthesis rather than by capturing the actual target molecule (other than for example in the initial stage when the arms hybridize to it or in the sense that the target molecule can remain bound to the extended MIP product until it is denatured or otherwise removed).

It should be appreciated that in some embodiments a targeting arm may include a sequence that is complementary to one allele or mutation (e.g., a SNP or other polymorphism, a mutation, etc.) so that the probe will preferentially hybridize (and capture) target nucleic acids having that allele or mutation. However, in many embodiments, each targeting arm is designed to hybridize (e.g., be complementary) to a sequence that is not polymorphic in the subjects of a population that is being evaluated. This allows target sequences to be captured and/or sequenced for all alleles and then the differences between subjects (e.g., calls of heterozygous or homozygous for one or more loci) can be based on the sequence information and/or the frequency as described herein.

It should be appreciated that sequence tags (also referred to as barcodes) may be designed to be unique in that they do not appear at other positions within a probe or a family of probes and they also do not appear within the sequences being targeted. Thus they can be used to uniquely identify (e.g., by sequencing or hybridization properties) particular probes having other characteristics (e.g., for particular subjects and/or for particular loci).

It also should be appreciated that in some embodiments probes or regions of probes or other nucleic acids are described herein as comprising or including certain sequences or sequence characteristics (e.g., length, other properties, etc.). However, it should be appreciated that in some embodiments, any of the probes or regions of probes or other nucleic acids consist of those regions (e.g., arms, central regions, tags, primer sites, etc., or any combination thereof) of consist of those sequences or have sequences with characteristics that consist of one or more characteristics (e.g., length, or other properties, etc.) as described herein in the context of any of the embodiments (e.g., for tiled or staggered probes, tagged probes, length detection, sensitivity enhancing algorithms or any combination thereof).

It should be appreciated that probes, primers, and other nucleic acids designed or used herein may be synthetic, natural, or a combination thereof. Accordingly, as used herein, the term

Petitioner Illumina, Inc., EX1004, p. 532

"nucleic acid" refers to multiple linked nucleotides (i.e., molecules comprising a sugar (e.g., ribose or deoxyribose) linked to an exchangeable organic base, which is either a pyrimidine (e.g., cytosine (C), thymidine (T) or uracil (U)) or a purine (e.g., adenine (A) or guanine (G)). "Nucleic acid" and "nucleic acid molecule" may be used interchangeably and refer to oligoribonucleotides as well as oligodeoxyribonucleotides. The terms shall also include polynucleosides (i.e., a polynucleotide minus a phosphate) and any other organic base containing nucleic acid.

The organic bases include adenine, uracil, guanine, thymine, cytosine and inosine. Unless otherwise stated, nucleic acids may be single or double stranded. The nucleic acid may be naturally or non-naturally occurring. Nucleic acids can be obtained from natural sources, or can be synthesized using a nucleic acid synthesizer (i.e., synthetic).

Harvest and isolation of nucleic acids are routinely performed in the art and suitable methods can be found in standard molecular biology textbooks. (See, for example, Maniatis' Handbook of Molecular Biology.) The nucleic acid may be DNA or RNA, such as genomic DNA, mitochondrial DNA, mRNA, cDNA, rRNA, miRNA, or a combination thereof. Non-naturally occurring nucleic acids such as bacterial artificial chromosomes (BACs) and yeast artificial chromosomes (YACs) can also be used.

The invention also contemplates the use of nucleic acid derivatives. As will be described herein, the use of certain nucleic acid derivatives may increase the stability of the nucleic acids of the invention by preventing their digestion, particularly when they are exposed to biological samples that may contain nucleases. As used herein, a nucleic acid derivative is a non-naturally occurring nucleic acid or a unit thereof. Nucleic acid derivatives may contain non-naturally occurring elements such as non-naturally occurring nucleotides and non-naturally occurring backbone linkages.

Nucleic acid derivatives may contain backbone modifications such as but not limited to phosphorothioate linkages, phosphodiester modified nucleic acids, phosphorothiolate modifications, combinations of phosphodiester and phosphorothioate nucleic acid, methylphosphonate, alkylphosphonates, phosphate esters, alkylphosphonothioates, phosphoramidates, carbamates, carbonates, phosphate triesters, acetamidates, carboxymethyl esters, methylphosphorothioate, phosphorodithioate, p-ethoxy, and combinations thereof. The backbone composition of the nucleic acids may be homogeneous or heterogeneous.

Petitioner Illumina, Inc., EX1004, p. 533

Nucleic acid derivatives may contain substitutions or modifications in the sugars and/or bases. For example, they include nucleic acids having backbone sugars which are covalently attached to low molecular weight organic groups other than a hydroxyl group at the 3' position and other than a phosphate group at the 5' position (e.g., an 2'-0-alkylated ribose group). Nucleic acid derivatives may include non-ribose sugars such as arabinose. Nucleic acid derivatives may contain substituted purines and pyrimidines such as C-5 propyne modified bases, 5-methylcytosine, 2-aminopurine, 2-amino-6-chloropurine, 2,6-diaminopurine, hypoxanthine, 2-thiouracil and pseudoisocytosine. In some embodiments, substitution(s) may include one or more substitutions/modifications in the sugars/bases, groups attached to the base, including biotin, fluorescent groups (fluorescein, cyanine, rhodamine, etc), chemically-reactive groups including carboxyl, NHS, thiol, etc., or any combination thereof.

A nucleic acid may be a peptide nucleic acid (PNA), locked nucleic acid (LNA), DNA, RNA, or co-nucleic acids of the same such as DNA-LNA co-nucleic acids. PNA are DNA analogs having their phosphate backbone replaced with 2-aminoethyl glycine residues linked to nucleotide bases through glycine amino nitrogen and methylenecarbonyl linkers. PNA can bind to both DNA and RNA targets by Watson-Crick base pairing, and in so doing form stronger hybrids than would be possible with DNA or RNA based oligonucleotides in some cases.

PNA are synthesized from monomers connected by a peptide bond (Nielsen, P.E. et al. Peptide Nucleic Acids, Protocols and Applications, Norfolk: Horizon Scientific Press, p. 1-19 (1999)). They can be built with standard solid phase peptide synthesis technology. PNA chemistry and synthesis allows for inclusion of amino acids and polypeptide sequences in the PNA design. For example, lysine residues can be used to introduce positive charges in the PNA backbone. All chemical approaches available for the modifications of amino acid side chains are directly applicable to PNA. Several types of PNA designs exist, and these include single strand PNA (ssPNA), bisPNA and pseudocomplementary PNA (pcPNA).

The structure of PNA/DNA complex depends on the particular PNA and its sequence. ssPNA binds to single stranded DNA (ssDNA) preferably in antiparallel orientation (i.e., with the N-terminus of the ssPNA aligned with the 3' terminus of the ssDNA) and with a Watson-Crick pairing. PNA also can bind to DNA with a Hoogsteen base pairing, and thereby forms triplexes with double stranded DNA (dsDNA) (Wittung, P. et al., *Biochemistry* 36:7973 (1997)).

Petitioner Illumina, Inc., EX1004, p. 534

A locked nucleic acid (LNA) is a modified RNA nucleotide. An LNA form hybrids with DNA, which are at least as stable as PNA/DNA hybrids (Braasch, D.A. et al., *Chem & Biol.* 8(1):1-7(2001)). Therefore, LNA can be used just as PNA molecules would be. LNA binding efficiency can be increased in some embodiments by adding positive charges to it. LNAs have been reported to have increased binding affinity inherently.

Commercial nucleic acid synthesizers and standard phosphoramidite chemistry are used to make LNAs. Therefore, production of mixed LNA/DNA sequences is as simple as that of mixed PNA/peptide sequences. The stabilization effect of LNA monomers is not an additive effect. The monomer influences conformation of sugar rings of neighboring deoxynucleotides shifting them to more stable configurations (Nielsen, P.E. et al. Peptide Nucleic Acids, Protocols and Applications, Norfolk: Horizon Scientific Press, p. 1-19 (1999)). Also, lesser number of LNA residues in the sequence dramatically improves accuracy of the synthesis. Most of biochemical approaches for nucleic acid conjugations are applicable to LNA/DNA constructs.

These and other aspects of the invention are illustrated by the following non-limiting examples.

EXAMPLES

The following examples illustrate non-limiting embodiments of the invention.

Example 1: Design a set of capture probes for a human target exon

All targets are captured as a set of partially-overlapping subtargets. For example, in the tiling approach, a 200 bp target exon might be captured as a set of 12 subtargets, each 60bp in length (FIG. 1). Each subtarget is chosen such that it partially overlaps two or three other targets.

In some embodiments, all probes are composed of three regions: 1) a 20 bp 'targeting arm' comprised of sequence which hybridizes immediately upstream from the sub-target, 2) a 30 bp 'constant region' comprised of sequence used as a pair of amplification priming sites, and 3) a second 20 bp 'targeting arm' comprised of sequence which hybridizes immediately downstream from the sub-target. Targeting arm sequences will be different for each capture probe in a set, while constant region sequence will be the same for all probes in the set, allowing all captured targets to be amplified with a single set of primers. Targeting arm sequences should be designed

Petitioner Illumina, Inc., EX1004, p. 535

such that any given pair of 20 bp sequences is unique in the target genome (to prevent spurious capture of undesired sites). Additionally, melting temperatures should be matched for all probes in the set such that hybridization efficiency is uniform for all probes at a constant temperature (e.g., 60 C). Targeting arm sequences should be computationally screened to ensure they do not form strong secondary structure that would impair their ability to basepair with the genomic target.

Hybridize capture probes to human genomic sample

Assemble hybridization reaction:

- 1.0 ul capture probe mix (~2.5 pmol)
- 2.0 ul 10X Ampligase buffer (Epicentre)
- 6.0 ul 500 ng/ul human genomic DNA (~16.7 fmol)
- 11 ul dH20

In a thermal cycler, heat reaction to 95C for 5 min to denature genomic DNA, then cool to 60C. Allow to incubate at 60C for 40 hours.

Convert hybridized probes into covalently-closed circular products containing subtargets

Prepare fill-in / ligation reaction mixture:

- 0.25 ul 2 mM dNTP mix (Invitrogen)
- 2.5 ul 10X Ampligase buffer (Epicentre)
- 5.0 ul 5 U/ul Taq Stoffel fragment (Applied Biosystems)
- 12.5 ul 5 U/ul Ampligase (Epicentre)
- 4.75 ul dH20

Add 1.0 ul of this mix to the hybridized probe reaction, and incubate at 60C for 10 hours.

Purify circularized probe / subtarget products from un-reacted probes and genomic DNA

Prepare exonuclease reaction mixture:

- 21 ul fill-in / ligation reaction product
- 2.0 ul 10X exonuclease I buffer (New England Biolabs)

Petitioner Illumina, Inc., EX1004, p. 536

- 2.0 ul 20 U/ul exonuclease I (New England Biolabs)
- 2.0 ul 100 U/ul exonuclease III (New England Biolabs)

Incubate at 37C for 60 min, then heat-inactivate by incubating at 80C for 15 min. Immediately cool to 4C for storage.

Amplify circular material by PCR using primers specific to the 'constant region' of the probes

Prepare PCR mixture:

- 5.0 ul 10X Accuprime reaction buffer (Invitrogen)
- 1.5 ul 10 uM CP-2-FA (5'-GCACGATCCGACGGTAGTGT-3') (SEQ ID NO:183)
- 1.5 ul 10 uM CP-2-RA (5'-CCGTAATCGGGAAGCTGAAG-3') (SEQ ID NO:184)
-
- 0.4 ul 25 mM dNTP mix (Invitrogen)
- 2.0 ul heat-inactivated exonuclease reaction mix
- 1.5 ul 10X SybrGreen (Invitrogen)
- 0.4 ul 2.5 U/ul Accuprime Pfx polymerase (Invitrogen)
- 37.7 ul dH20

Thermal cycle in real-time thermal cycler according to the following protocol, but stop cycling before amplification yield plateaus (generally 8-12 cycles):

1. 95C for 5 min
2. 95C for 30 sec
3. 58C for 60 sec
4. 72C for 60 sec
5. goto 2, N more times

Prepare a shotgun next-generation sequencing library for analysis

- Purify desired amplicon population from non-specific amplification products by gel extraction.

Petitioner Illumina, Inc., EX1004, p. 537

- Concatemerize amplicons into high-molecular weight products suitable for shearing
- Mechanically shear, using either a nebulizer, BioRuptor, Hydroshear, Covaris, or similar instrument. DNA should be sheared into fragments several hundred basepairs in length.
- Ligate adapters required for amplification by the sequencing platform used. If necessary, purify ligated product from unligated product and adapters.

Example 2: Use of differentiator tag sequences to detect and correct bias in a MIP-capture reaction of a set of exon targets

The first step in performing the detection/correction is to determine how many differentiator tag sequences are necessary for the given sample. In this example, 1000 genomic targets corresponding to 1000 exons were captured. Since the differentiator tag sequence is part of the probe, it will measure/report biases that occur from the earliest protocol steps. Also, being located in the backbone, the differentiator tag sequence can easily be sequenced from a separate priming site, and therefore not impact the total achievable read-length for the target sequence. MIP probes are synthesized using standard column-based oligonucleotide synthesis by any number of vendors (e.g. IDT), and differentiator tag sequences are introduced as 'degenerate' positions in the backbone. Each degenerate position increases the total number of differentiator tag sequences synthesized by a factor of 4, so a 10 nt degenerate region implies a differentiator tag sequence complexity of ~1e6 species.

Hybridize capture probes to human genomic sample

Assemble hybridization reaction:

- 1.0 ul capture probe mix (~2.5 pmol)
- 2.0 ul 10X Ampligase buffer (Epicentre)
- 6.0 ul 500 ng/ul human genomic DNA (~16.7 fmol)
- 11 ul dH20

In a thermal cycler, heat reaction to 95C for 5 min to denature genomic DNA, then cool to 60C. Allow to incubate at 60C for 40 hours.

Convert hybridized probes into covalently-closed circular products containing subtargets

Prepare fill-in / ligation reaction mixture:

- 0.25 ul 2 mM dNTP mix (Invitrogen)
- 2.5 ul 10X Ampligase buffer (Epicentre)
- 5.0 ul 5 U/ul Taq Stoffel fragment (Applied Biosystems)
- 12.5 ul 5 U/ul Ampligase (Epicentre)
- 4.75 ul dH20

Add 1.0 ul of this mix to the hybridized probe reaction, and incubate at 60C for 10 hours.

Purify circularized probe / subtarget products from un-reacted probes and genomic DNA

Prepare exonuclease reaction mixture:

- 21 ul fill-in / ligation reaction product
- 2.0 ul 10X exonuclease I buffer (New England Biolabs)
- 2.0 ul 20 U/ul exonuclease I (New England Biolabs)
- 2.0 ul 100 U/ul exonuclease III (New England Biolabs)

Incubate at 37C for 60 min, then heat-inactivate by incubating at 80C for 15 min. Immediately cool to 4C for storage.

Amplify circular material by PCR using primers specific to the 'constant region' of the probes

Prepare PCR mixture:

- 5.0 ul 10X Accuprime reaction buffer (Invitrogen)
- 1.5 ul 10 uM CP-2-FA (5'-GCACGATCCGACGGTAGTGT-3') (SEQ ID NO: 183)
-
- 1.5 ul 10 uM CP-2-RA (5'-CCGTAATCGGGAAGCTGAAG-3') (SEQ ID NO: 184)
-
- 0.4 ul 25 mM dNTP mix (Invitrogen)
- 2.0 ul heat-inactivated exonuclease reaction mix
- 1.5 ul 10X SybrGreen (Invitrogen)
- 0.4 ul 2.5 U/ul Accuprime Pfx polymerase (Invitrogen)

Petitioner Illumina, Inc., EX1004, p. 539

- 37.7 ul dH20

Thermal cycle in real-time thermal cycler according to the following protocol, but stop cycling before amplification yield plateaus (generally 8-12 cycles):

6. 95C for 5 min
7. 95C for 30 sec
8. 58C for 60 sec
9. 72C for 60 sec
10. goto 2, N more times

Prepare a shotgun next-generation sequencing library for analysis

- Purify desired amplicon population from non-specific amplification products by gel extraction.
- Concatemerize amplicons into high-molecular weight products suitable for shearing
- Mechanically shear, using either a nebulizer, BioRuptor, Hydroshear,
- Covaris, or similar instrument. DNA should be sheared into fragments several hundred basepairs in length.
- Ligate adapters required for amplification by the sequencing platform used. If necessary, purify ligated product from unligated product and adapters.

Perform sequencing of library according to manufacturer's directions (e.g. Illumina,ABI, etc), reading both the target sequence and the differentiator tag sequence.

Analyze data by correcting for any biases detected by quantitation of differentiator tag sequence abundance

Construct a table of target: differentiator tag abundances from the read data, e.g.:

Target ID	Differentiator Tag sequence ID	Count
1	3547	1
2	4762	1
1	9637	1
1	1078	5
3	4762	1
1	2984	1

Petitioner Illumina, Inc., EX1004, p. 540

All 'count' entries should be '1', since any particular target:differentiator tag mapping will not occur more than once by chance, and therefore will only be observed if bias was present somewhere in the sample preparation process. For any target:differentiator tag combination observed more than once, all such reads are 'collapsed' into a single read before consensus basecalls are determined. This will cancel the effect of bias on consensus basecall accuracy. FIG. 5 depicts a method for making diploid genotype calls in which repeat target:differentiator tag combination are collapsed.

Example 3: Differentiator tag sequence design for MIP capture reactions

For a set of targets, the number of differentiator tag sequences necessary to be confident (within some statistical bounds) that a certain differentiator tag sequence will not be observed more than once by chance in combination with a certain target sequence was determined. The total number of unique differentiator tag sequences for a certain differentiator tag sequence length is determined as $4^{(\text{Length in nucleotides of the differentiator tag sequence})}$. For a molecular inversion probe capture reaction that uses MIP probes having differentiator tag sequences, the probability of performing the capture reaction and capturing one or more copies of a target sequence having the same differentiator tag sequence is calculated as: $p = 1 - [N! / (N-M)!] / [NAM]$, wherein N is the total number of possible unique differentiator tag sequences and M is the number of target sequence copies in the capture reaction. Thus, by varying the differentiator tag sequence length it is possible to perform a MIP capture reaction in which the probability of capturing one or more copies of a target sequence having the same differentiator tag sequence is set at a predetermined probability value.

For example, for a differentiator tag sequence of 15 nucleotides in length, there are 1,073,741,824 possible differentiator tag sequences. A MIP capture reaction in which MIP probes, each having a differentiator tag sequence of 15 nucleotides, are combined with 10000 target sequence copies (e.g., genome equivalents), the probability of capturing one or more copies of a target sequence having the same differentiator tag sequence is 0.05. In this example, the MIP reaction will produce very few (usually 0, but occasionally 1 or more) targets where multiple copies are tagged with the same differentiator tag sequence. FIG. 6 depicts results of a simulation for 100000 capture reactions having 15 nucleotide differentiator tag sequences and 10000 target sequences.

Petitioner Illumina, Inc., EX1004, p. 541

Example 4: Assessment of the probability for obtaining enough sequencing reads to make accurate base-calls at multiple independent loci, as a function of sequencing coverage.

Monte Carlo simulations were performed to determine sequencing coverage requirements. The simulations assume 10000 genomic copies of a given locus (target) half mom alleles and half dad alleles. The simulations further assume 1% efficiency of capture for the MIP reaction. The simulation samples from a capture mix 100 times without replacement to create a set of 100 capture products. The simulation then samples from the set of 100 capture products with replacement (assuming unbiased amplification) to generate 'reads' from either mom or dad. The number of reads sampled depends on the coverage. The number of independent reads from both mom and dad necessary to make a high-quality base-call (assumed to be 10 or 20 reads) were then determined. The process was repeated 1000 times for each coverage level, and the fraction of times that enough reads from both parents were successfully obtained was determined. This fraction was raised to the power 1000, assuming we have 1000 independent loci that must obtain successful base-calls, plotted (See FIG. 7). Result show that roughly 50x coverage is required to capture each allele >= 10x with > 0.95 probability.

Example 5: MIP capture of 'target' locus and 'control' loci

In some embodiments, to accurately quantify the efficiency of target locus capture, at least three sets of control loci are captured in parallel that have a priori been shown to serve as proxies for various lengths of target locus. For example, if the target locus is expected to have a length between 50 and 1000 bp, then sets of control loci having lengths of 50, 250, and 1000 bp could be captured (e.g. 20 loci per set should provide adequate protection from outliers), and their abundance digitally measured by sequencing. These loci should be chosen such that minimal variation in efficiency between samples and on multiple runs of the same sample is observed (and are therefore 'efficiency invariant'). These will serve as 'reference' points that define the shape of the curve of abundance-vs-length. Determining the length of the target is then simply a matter of 'reading' the length from the appropriate point on the calibration curve.

In some embodiments, the statistical confidence one has in the estimate of target length from this method is driven largely by three factors: 1) reproducibility / variation of the abundance data used to generate the calibration curve; 2) goodness of fit of the regression to the 'control' datapoints; 3) reproducibility of abundance data for the target locus being measured. Statistical bounds on 1) and 2) will be known in advance, having been measured during

Petitioner Illumina, Inc., EX1004, p. 542

development of the assay. Additionally, statistical bounds on 3) will be known in general in advance, since assay development should include adequate population sampling and measure of technical reproducibility. Standard statistical methods should be used to combine these three measures into a single P value for any given experimental measure of target abundance.

In some embodiments, given the set of calibration observations, and a linear regression fit to that data, the regression can be used to predict the length value for n observations of the target locus whose length is unknown. First, choose an acceptable range for the confidence interval of the length estimate. For example, in the case of distinguishing "normal" (87-93bp) from "premutation" (165-600bp) potential cases of Fragile X, the goal is to measure length to sufficient precision to distinguish 93 bp from 165 bp. The predicted response value, computed when n observations is substituted into the equation for the regressed line, will have arbitrary precision. However, if for example a 95% confidence level is desired, that 95% confidence interval must be sufficiently short that it does not overlap both the "normal" and "premutation" length ranges. Continuing the example, if one calculates a length of 190 from n=400 MIP observations, and based on the regression from calibration data, the 95% confidence interval is 190 +/- 20 bp, one can conclude the sample represents a "premutation" length with 95% certainty. Conversely, if the calibration data were less robust, error estimates of the regression would be higher, leading to larger confidence intervals on the predicted response value. In some embodiments, if the 95% CI were calculated as 190 +/- 100 bp from n=400, one could not determine whether the predicted response value corresponds to a "normal" or "premutation" length.

In some embodiments, the confidence interval for a predicted response is calculated as:

The estimate for the response $\hat{y}$ *is identical to the estimate for the mean of the response:* $\hat{y} = b_0 + b_1 x^*$. *The confidence interval for the predicted value is given by* $\hat{y} \pm t^* s_{\hat{y}}$, *where* $\hat{y}$ *is the fitted value corresponding to* x^*. *The value* t^* *is the upper (1 - C)/2 critical value for the t(n - 2) distribution.*

In some embodiments, a technique for analyzing a locus of interest can involve the following steps.

Petitioner Illumina, Inc., EX1004, p. 543

Convert hybridized probes into covalently-closed circular products containing subtargets

Prepare fill-in / ligation reaction mixture:

- 0.25 ul 2 mM dNTP mix (Invitrogen)
- 2.5 ul 10X Ampligase buffer (Epicentre)
- 5.0 ul 5 U/ul Taq Stoffel fragment (Applied Biosystems)
- 12.5 ul 5 U/ul Ampligase (Epicentre)
- 4.75 ul dH20

Add 1.0 ul of this mix to the hybridized probe reaction, and incubate at 60C for 10 hours.

Purify circularized probe / subtarget products from un-reacted probes and 30 genomic DNA

Prepare exonuclease reaction mixture:

- 21 ul fill-in / ligation reaction product
- 2.0 ul 10X exonuclease I buffer (New England Biolabs)
- 2.0 ul 20 U/ul exonuclease I (New England Biolabs)
- 2.0 ul 100 U/ul exonuclease III (New England Biolabs)

Incubate at 37C for 60 min, then heat-inactivate by incubating at 80C for 15 min. Immediately cool to 4C for storage.

Amplify circular material by PCR using primers specific to the 'constant region' of the probes

Prepare PCR mixture:

- 5.0 ul 10X Accuprime reaction buffer (Invitrogen)
- 1.5 ul 10 uM CP-2-FA-Ilmn (platform-specific amplification sequence plus 'circle constant region'-specific sequence)
- 1.5 ul 10 uM CP-2-RA-Ilmn (platform-specific amplification sequence plus 'circle constant region'-specific sequence)
- 0.4 ul 25 mM dNTP mix (Invitrogen)
- 2.0 ul heat-inactivated exonuclease reaction mix
- 1.5 ul 10X SybrGreen (Invitrogen)
- 0.4 ul 2.5 U/ul Accuprime Pfx polymerase (Invitrogen)

Petitioner Illumina, Inc., EX1004, p. 544

- 37.7 ul dH20

Thermal cycle in real-time thermal cycler according to the following protocol, but stop cycling before amplification yield plateaus (generally 8-12 cycles):

11. 95C for 5 min
12. 95C for 30 sec
13. 58C for 60 sec
14. 72C for 60 sec
15. goto 2, N more times

Perform sequencing (e.g., next-generation sequencing) on sample for digital quantitation according to manufacturer's instructions (e.g., Illumina, ABI)

Example 6: MIP-capture reaction of a set of exon target nucleic acids

MIP probes are synthesized using standard column-based oligonucleotide synthesis by any number of vendors (e.g. IDT).

Hybridize capture probes to human genomic sample

Assemble hybridization reaction:

- 1.0 ul capture probe mix (-2.5 pmol)
- 2.0 ul 10X Ampligase buffer (Epicentre)
- 6.0 ul 500 ng/ul human genomic DNA (-16.7 fmol)
- 11 ul dH20

In a thermal cycler, heat reaction to 95C for 5 min to denature genomic DNA, then cool to 60C. Allow to incubate at 60C for 40 hours.

Convert hybridized probes into covalently-closed circular products containing target nucleic acids

Prepare fill-in / ligation reaction mixture:

- 0.25 ul 2 mM dNTP mix (Invitrogen)
- 2.5 ul 10X Ampligase buffer (Epicentre)
- 5.0 ul 5 U/ul Taq Stoffel fragment (Applied Biosystems)
- 12.5 ul 5 U/ul Ampligase (Epicentre)
- 4.75 ul dH20

Add 1.0 ul of this mix to the hybridized probe reaction, and incubate at 60C for 10 hours.

Purify, circularized probe/target nucleic acid products from un-reacted probes and genomic DNA

Prepare exonuclease reaction mixture:

- 21 ul fill-in / ligation reaction product
- 2.0 ul 10X exonuclease I buffer (New England Biolabs)
- 2.0 ul 20 U/ul exonuclease I (New England Biolabs)
- 2.0 ul 100 U/ul exonuclease III (New England Biolabs)

Incubate at 37C for 60 min, then heat-inactivate by incubating at 80C for 15 min. Immediately cool to 4C for storage.

Amplify circular material by PCR using primers specific to the 'constant region' of the probes

Prepare PCR mixture:

- 5.0 ul 10X Accuprime reaction buffer (Invitrogen)
- 1.5 ul 10 uM CP-2-FA (5'-GCACGATCCGACGGTAGTGT-3') (SEQ ID NO: 183)
-
- 1.5 ul 10 uM CP-2-RA (5'-CCGTAATCGGGAAGCTGAAG-3') (SEQ ID NO: 184)
- 0.4 ul 25 mM dNTP mix (Invitrogen)
- 2.0 ul heat-inactivated exonuclease reaction mix
- 1.5 ul 10X SybrGreen (Invitrogen)
- 0.4 ul 2.5 U/ul Accuprime Pfx polymerase (Invitrogen)
- 37.7 ul dH20

Thermal cycle in real-time thermal cycler according to the following protocol, but stop cycling before amplification yield plateaus (generally 8-12 cycles):

16. 95C for 5 min
17. 95C for 30 sec
18. 58C for 60 sec
19. 72C for 60 sec
20. goto 2, N more times

Petitioner Illumina, Inc., EX1004, p. 546

Prepare a shotgun next-generation sequencing library for analysis

- Purify desired amplicon population from non-specific amplification products by gel extraction.
- Concatemerize amplicons into high-molecular weight products suitable for shearing
- Mechanically shear, using either a nebulizer, BioRuptor, Hydroshear, Covaris, or similar instrument. DNA should be sheared into fragments several hundred basepairs in length.
- Ligate adapters required for amplification by the sequencing platform used. If necessary, purify ligated product from unligated product and adapters.

Perform sequencing of library according to manufacturer's directions (e.g. Illumina, ABI, etc), reading the target sequence to determine abundance of the target nucleic acid.

Example 7: Use of MIPs, hybridization, and mutation-detection MIPs to genotype a set of 1000 targets

MIPs, hybridization, and mutation-detection MIPs are used to genotype a set of 1000 targets. The protocol permits detection of any of 50 specific known point mutations

First, separate MIP, hybridization, and mutation-detection MIP reactions are performed on a biological sample. A MIP capture reaction is performed essentially as described in Turner et al 2009, Nature methods 6:315-6. A set of MIPs is designed such that each probe in the set flanks one of the 1000 targets. Separately, a hybridization enrichment reaction is performed using the Agilent SureSelect procedure. Prior to selection, the genomic DNA to be enriched is converted into a shotgun sequencing library using Illumina's 'Fragment Library' kit and protocol. Agilent's web interface is used to design a set of probes which will hybridize to the target nucleic acids. Separately, a set of probes are designed (mutation-detection MIPs) which will form MIPs only if mutations (e.g., specific polymorphisms) are present. Each mutation-detection MIP has a 3'-most base identity that is specific for a single known mutation. A reaction with this set of mutation-detection MIPs is performed to selectively detect the presence of any mutant alleles.

Once all three reactions have been performed, the two MIP reactions are combined (e.g., at potentially non-equimolar ratios to further increase sensitivity of mutation detection) into a

Petitioner Illumina, Inc., EX1004, p. 547

single tube, and run as one sample on the next-generation DNA sequencing instrument. The hybridization-enriched reaction is run as a separate sample on the next-generation DNA sequencing instrument. Reads from each 'sample' are combined by a software algorithm which forms a consensus diploid genotype at each position in the target set by evaluating the total coverage at each position, the origin of each read in that total coverage, the quality score of each individual read, and the presence (or absence) of any reads derived from mutation-specific MIPs overlapping the region.

It should be appreciated that the preceding examples are non-limiting and aspects of the invention may be implemented as described herein using alternative techniques and/or protocols that are available to one or ordinary skill in the art.

It will be clear that the methods may be practiced other than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present disclosure are possible in light of the above teachings and, therefore, are within the scope of the claims. Preferred features of each aspect of the Disclosure are as for each of the other aspects mutatis mutandis. The documents including patents, patent applications, journal articles, or other disclosures mentioned herein are hereby incorporated by reference in their entirety. In the event of conflict, the disclosure of present application controls, other than in the event of clear error.

RAW SEQUENCE LISTING

Loaded by SCORE, no errors detected.

Application Serial Number: 16952764
Source: OPAP
Date Processed by SCORE: 12/08/2020

ENTERED

SEQUENCE LISTING

<110> Good Start Genetics, Inc.
Porreca, Gregory
Laserson, Uri

<120> METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS

<130> GSGE-002/04US 29258/78

<140> US 16/952,764
<141> 2020-11-19

<150> US 15/231,687
<151> 2016-08-08

<160> 184

<170> PatentIn version 3.5

<210> 1
<211> 52
<212> DNA
<213> Homo sapiens

<400> 1
tcacatcacc aagttaaaaa aaaaaarggg gcgggggggc agaatgaaaa tt 52

<210> 2
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 2
aaacaaggat gaattaagtt ttttttaaa aaagaaacat ttggtaaggg ga 52

<210> 3
<211> 52
<212> DNA
<213> Homo sapiens

<400> 3
atcaccaagt taaaaaaaaa aaagggrcgg gggggcagaa tgaaaattgc at 52

<210> 4
<211> 52
<212> DNA
<213> Homo sapiens

<400> 4
ctattgaacc agaaccaaac aggaatrcca tagcattttg taaactaaac tg 52

<210> 5
<211> 52
<212> DNA
<213> Homo sapiens

<400> 5
caggagttca agaccagcct actaaamcac acacacacac acacacacac ac 52

<210> 6
<211> 52
<212> DNA
<213> Homo sapiens

<400> 6
gattaaataa tagtgtttat gtacccsgct tataggagaa gagggtgtgt gt 52

<210> 7
<211> 52
<212> DNA
<213> Homo sapiens

<400> 7
attgttatct tttcatataa ggtaacwgag gcccagagag attaaataac at 52

<210> 8
<211> 52
<212> DNA
<213> Homo sapiens

<400> 8
taattttaat taagtaaatt taattgrtag ataaataagt agataaaaaa ta 52

Petitioner Illumina, Inc., EX1004, p. 551

<210> 9
<211> 52
<212> DNA
<213> Homo sapiens

<400> 9
gtataaaaaa aaaaaaaaaa aaagttwgaa tgttttcttg cattcagagc ct 52

<210> 10
<211> 52
<212> DNA
<213> Homo sapiens

<400> 10
atactaaaaa tttaaagttc tcttgcrata tattttctta atatcttaca tc 52

<210> 11
<211> 52
<212> DNA
<213> Homo sapiens

<400> 11
tgctgggatt acaggcgtga gccaccrcgc ctggcctgat gggacatatt tt 52

<210> 12
<211> 52
<212> DNA
<213> Homo sapiens

<400> 12
ctacaatata agtatagtat tgcaaamcca tcaggaaggg tgttaactat tt 52

<210> 13
<211> 54
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(29)
<223> This region may or may not be present in its entirety

Petitioner Illumina, Inc., EX1004, p. 552

<400> 13
gttgtctcca aacttttttt caggtgagaa ggtggccaac cgagcttcgg aaag 54

<210> 14
<211> 52
<212> DNA
<213> Homo sapiens

<400> 14
ttttttcata aaagattata taaaggmtat tgcttttgaa tcacaaacac ta 52

<210> 15
<211> 52
<212> DNA
<213> Homo sapiens

<400> 15
atctagtgag cagtcaggaa agagaayttc cagatcctgg aaatcagggt ta 52

<210> 16
<211> 52
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(27)
<223> May or may not be present

<400> 16
tagagtataa aaaaaaaaaa aaaaaaagtt tgaatgtttt cttgcattca ga 52

<210> 17
<211> 52
<212> DNA
<213> Homo sapiens

<400> 17
tgcaaatgcc aactatcaaa gatattsgag tatactgtca ataaacttca ta 52

<210> 18

Petitioner Illumina, Inc., EX1004, p. 553

<211> 52
<212> DNA
<213> Homo sapiens

<400> 18
tcctcaacag ttagaaacaa tattttsagt gatttcccat gccaacttta ct 52

<210> 19
<211> 52
<212> DNA
<213> Homo sapiens

<400> 19
tttttggtat tgttgttaaa taagtgrgaa ttcaatacag tataatgtct gt 52

<210> 20
<211> 87
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (27)..(72)
<223> This region may or may not be present in its entirety

<400> 20
cttgaaatcg gatatatata tatatatgta tatatatata tatatatata tatatatata 60

tagtattatc cctgttttca cagtttt 87

<210> 21
<211> 52
<212> DNA
<213> Homo sapiens

<400> 21
agaggggctg tgaaggacac caaggargag actaagagcc aggagggaaa ac 52

<210> 22
<211> 52
<212> DNA
<213> Homo sapiens

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
16/952,764	11/19/2020	Gregory Porreca	MLB-002/06US 34468/40

CONFIRMATION NO. 7812

21710
BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

FORMALITIES LETTER

OC000000121944662

Date Mailed: 12/11/2020

NOTICE OF INCOMPLETE REPLY (NONPROVISIONAL)

Filing Date Granted

The U.S. Patent and Trademark Office has received your reply on 12/03/2020 to the NOTICE TO FILE CORRECTED APPLICATION PAPERS (Notice) mailed 12/01/2020 and it has been entered into the application. The reply, however, is not a complete reply for the reason(s) listed below. A complete reply must be timely filed to prevent ABANDONMENT of the above-identified application. Replies should be mailed to: Mail Stop Missing Parts, Commissioner for Patents, P.O.Box 1450, Alexandria VA 22313-1450.

The period for reply continues to run from the date of the Notice mailed 12/01/2020. Applicant must submit all required items and pay any fees required below within two months from the date of the NOTICE TO FILE CORRECTED APPLICATION PAPERS to avoid abandonment. If the reply is submitted after two months from the date of the NOTICE TO FILE CORRECTED APPLICATION PAPERS, extensions of time may be obtained by filing a petition accompanied by the extension fee under the provisions of 37 CFR 1.136(a).

Items Required to Avoid Abandonment:

The required items noted below SHOULD be filed along with any items required above.

The application is informal since it does not comply with the regulations for the reason(s) indicated below.

The required item(s) identified below must be timely submitted to avoid abandonment:

- Applicant did not provide a statement that the substitute specification contains no new matter.

Applicant is cautioned that correction of the above items may cause the specification and drawings page count to exceed 100 pages. If the specification and drawings exceed 100 pages, applicant will need to submit the required application size fee.

Petitioner Illumina, Inc., EX1004, p. 555

Mail date of Notice: 12/01/2020

Last date that extension may be obtained: (Note: The petition and fee must be received by this date, or include a proper certificate of mailing under 37 CFR 1.8 with a date on or before this date, and extend the time to include this date.) 07/01/2021

Length of Extension of Time	Fee under 37 CFR 1.17(a) effective March 19, 2013		
(response due on or before)	Undiscounted	Small Entity	Micro Entity
One Month (03/01/2021)	$220	$110	$55
Two Months (04/01/2021)	$640	$320	$160
Three Months (05/01/2021)	$1480	$740	$370
Four Months (06/01/2021)	$2320	$1160	$580
Five Months (07/01/2021)	$3160	$1580	$790

Replies must be received in the USPTO within the set time period or must include a proper Certificate of Mailing or Transmission under 37 CFR 1.8 with a mailing or transmission date within the set time period. For more information and a suggested format, see Form PTO/SB/92 and MPEP 512.

Replies should be mailed to:

Mail Stop Missing Parts
Commissioner for Patents
P.O. Box 1450
Alexandria VA 22313-1450

Registered users of EFS-Web may alternatively submit their reply to this notice via EFS-Web, including a copy of this Notice and selecting the document description "Applicant response to Pre-Exam Formalities Notice". https://sportal.uspto.gov/authenticate/AuthenticateUserLocalEPF.html

For more information about EFS-Web please call the USPTO Electronic Business Center at 1-866-217-9197 or visit our website at http://www.uspto.gov/ebc.

If you are not using EFS-Web to submit your reply, you must include a copy of this notice.

Questions about the contents of this notice and the requirements it sets forth should be directed to the Office of Data Management, Application Assistance Unit, at **(571) 272-4000** or **(571) 272-4200** or **1-888-786-0101**.

/kgebremichael/

Petitioner Illumina, Inc., EX1004, p. 556

Electronic Acknowledgement Receipt	
EFS ID:	41364589
Application Number:	16952764
International Application Number:	
Confirmation Number:	7812
Title of Invention:	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS
First Named Inventor/Applicant Name:	Gregory Porreca
Customer Number:	21710
Filer:	Thomas C. Meyers/Sandra Richards
Filer Authorized By:	Thomas C. Meyers
Attorney Docket Number:	MLB-002/06US 34468/40
Receipt Date:	11-DEC-2020
Filing Date:	19-NOV-2020
Time Stamp:	16:18:35
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Response to Pre-Exam Sequence Notice	MLB-002-06US-Response-NTFCAP121120.pdf	15965 a4f24e4c2fa3e83867f53c4d5a03aeacf8536059	no	2

Warnings:

Information:					
2	Specification	MLB-002-06US-Marked-Specification.pdf	2930349 13a31fde55edc23775d2143b1f87857e965e692c	no	87
Warnings:					
Information:					
3	Specification	MLB-002-06US-Clean-Specification.pdf	2930693 b4bbab84bfbe363a55bd69e7e5697e5120e33ca9	no	87
Warnings:					
Information:					
4	CRF Statement Paper and CRF are the same	MLB-002-06-CRF-Transfer-Req.pdf	255540 b151b4723dc02f5a08ad2aebf0c778e03fe5bc19	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	6132547		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

<u>New Applications Under 35 U.S.C. 111</u>
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.
<u>National Stage of an International Application under 35 U.S.C. 371</u>
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.
<u>New International Application Filed with the USPTO as a Receiving Office</u>
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Attorney Docket No.: MLB-002/06US 34468/40

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANT:	Molecular Loop Biosolutions, LLC	ART UNIT:	Not Assigned
SERIAL NUMBER:	16/952,764	CONF. NO.:	7812
FILING DATE:	November 19, 2020	EXAMINER:	Not Assigned
TITLE:	**METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS**		

FILED ELECTRONICALLY

Commissioner for Patents
P. O. Box 1450
Alexandria, VA 22313-1450

RESPONSE TO NOTICE OF INCOMPLETE REPLY AND NOTICE TO FILE CORRECTED APPLICATION PAPERS

This paper is in response to the Notice to File Corrected Application Papers mailed from the United States Patent and Trademark Office on December 1, 2020 and the Notice of Incomplete Reply mailed on December 11, 2020. Applicant resubmits herewith both marked-up and clean copies of the specification as amended in the preliminary amendment submitted on November 19, 2020 and as provided in the response filed on December 3, 2020. As stated in the preliminary amendment filed on November 19, no new matter is or was added by way of the amendments to the specification.

Regarding the sequence listing requirements of 37 CFR 1.821(e), as stated in the preliminary amendment submitted on November 19, 2020 and reiterated in the response filed on December 3, Applicant wishes the computer readable form (CRF) of the sequence listing in this application to be identical with the CRF filed in application 15/231,687. To that end, Applicant re-submits herewith the Request for Transfer of a Computer Readable Form Under 37 CFR 1.821(e) (PTO/SB/93).

If there are any questions regarding the above remarks, the Examiner is invited and encouraged to contact Applicant's representative.

Dated: December 11, 2020

Respectfully submitted,
BROWN RUDNICK LLP

BROWN RUDNICK LLP
One Financial Center
Boston, MA 02111
Tel: (617) 856-8483
Fax: (617) 856-8201

/Thomas C. Meyers/
Thomas C. Meyers, Reg. No. 36,989
Attorney for Applicant
Email: tmeyers@brownrudnick.com

63926044 v1

METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS

RELATED APPLICATIONS

This application is a continuation of U.S. Application No. 15/231,687, filed August 8, 2016, which is a continuation of U.S. Application No. 13/266,862, filed March 13, 2012, which is a national phase application that claims the benefit of, and priority to, international PCT Application No. PCT/US2010/001293, filed April 30, 2010, which claims the benefit under 35 U.S.C. §119(e) of U.S. Provisional Application No. ~~provisional application serial number~~ 61/174,470, filed April 30, 2009, U.S. Provisional Application No. ~~provisional application serial number~~ 61/178,923, filed May 15, 2009, U.S. Provisional Application No. ~~provisional application serial number~~ 61/179,358, filed May 18, 2009, and U.S. Provisional Application No. ~~provisional application serial number~~ 61/182,089, filed May 28, 2009, the entire contents of each of which are incorporated herein by reference.

FIELD OF INVENTION

The invention relates to methods and compositions for determining genotypes in patient samples.

BACKGROUND OF THE INVENTION

Information about the genotype of a subject is becoming more important and relevant for a range of healthcare decisions as the genetic basis for many diseases, disorders, and physiological characteristics is further elucidated. Medical advice is increasingly personalized, with individual decisions and recommendations being based on specific genetic information. Information about the type and number of alleles at one or more genetic loci impacts disease risk, prognosis, therapeutic options, and genetic counseling amongst other healthcare considerations.

For cost-effective and reliable medical and reproductive counseling on a large scale, it is important to be able correctly and unambiguously identify the allelic status for many different genetic loci in many subjects.

Petitioner Illumina, Inc., EX1004, p. 561

Numerous technologies have been developed for detecting and analyzing nucleic acid sequences from biological samples. These technologies can be used to genotype subjects and determine the allelic status of any locus of interest. However, they are not sufficiently robust and cost-effective to be scaled up for reliable high throughput analysis of many genetic loci in large numbers of patients. The frequency of incorrect or ambiguous calls is too high for current technology to manage large numbers of patient samples without involving expensive and time-consuming steps to resolve uncertainties and provide confidence in the information output.

SUMMARY OF THE INVENTION

Aspects of the invention relate to preparative and analytical methods and compositions for evaluating genotypes, and in particular, for determining the allelic identity (or identities in a diploid organism) of one or more genetic loci in a subject.

Aspects of the invention are based, in part, on the identification of different sources of ambiguity and error in genetic analyses, and, in part, on the identification of one or more approaches to avoid, reduce, recognize, and/or resolve these errors and ambiguities at different stages in a genetic analysis.

According to aspects of the invention, certain types of genetic information can be under-represented or over-represented in a genetic analysis due to a combination of stochastic variation and systematic bias in any of the preparative stages (e.g., capture, amplification, etc.), determining stages (e.g., allele-specific detection, sequencing, etc.), data interpretation stages (e.g., determining whether the assay information is sufficient to identify a subject as homozygous or heterozygous), and/or other stages.

According to aspects of the invention, error or ambiguity may be apparent in a genetic analysis, but not readily resolved without running additional samples or more expensive assays (*e.g.*, array-based assays may report no-calls due to noisy/low signal). According to further aspects of the invention, error or ambiguity may not be accounted for in a genetic analysis and incorrect base calls may be made even when the evidence for them is limited and/or not statistically significant (*e.g.*, next-generation sequencing technologies may report base calls even if the evidence for them is not statistically significant). According to further aspects of the invention error or ambiguity may be problematic for a multi-step genetic analysis because it is

Petitioner Illumina, Inc., EX1004, p. 562

apparent but not readily resolved in one or more steps of the analysis and not apparent or accounted for in other steps of the analysis.

In some embodiments, sources of error and ambiguity in one or more steps can be addressed by capturing and/or interrogating each target locus of interest with one or more sets of overlapping probes that are designed to overcome any systematic bias or stochastic effects that may impact the complexity and/or fidelity of the genetic information that is generated.

In some embodiments, sources of error and ambiguity in one or more steps can be addressed by capturing and/or interrogating each target locus of interest with at least one set of probes, wherein different probes are labeled with different identifiers that can be used to track the assay reactions and determine whether certain types of genetic information are under-represented or over-represented in the information that is generated.

In some embodiments, errors and ambiguities associated with the analysis of regions containing large numbers of sequence repeats are addressed by systematically analyzing frequencies of certain nucleic acids at particular stages in an assay (e.g., at a capture, sequencing, or detection stage). It should be appreciated that such techniques may be particularly useful in the context of a standardized protocol that is designed to allow many different loci to be evaluated in parallel without requiring different assay procedures for each locus. In some embodiments, the use of a single detection modality (*e.g.*, sequencing) to assay multiple types of genetic lesions (*e.g.*, point mutations, insertions/deletions, length polymorphisms) is advantageous in the clinical setting. In some embodiments of the invention, methods are provided that facilitate the use of multiple sample preparation steps in parallel, coupled with multiple analytical processes following sequence detection. Thus, in some embodiments of the invention, an improved workflow is provided that reduces error and uncertainty when simultaneously assaying different types of genetic lesions across multiple loci in multiple patients.

In some embodiments, aspects of the invention provide methods for overcoming preparative and/or analytical bias by combining two or more techniques, each having a different bias (e.g., a known bias towards under-representation or over-representation of one or more types of sequences), and using the resulting data to determine a genetic call for a subject with greater confidence.

Petitioner Illumina, Inc., EX1004, p. 563

It should be appreciated that in some embodiments, aspects of the invention relate to multiplex diagnostic methods. In some embodiments, multiplex diagnostic methods comprise capturing a plurality of genetic loci in parallel (e.g., one or more genetic loci from Table 1). In some embodiments, the genetic loci possess one or more polymorphisms (e.g., one or more polymorphisms from Table 2) the genotypes of which correspond to disease causing alleles. Accordingly, in some embodiments, the disclosure provides methods for assessing multiple heritable disorders in parallel. In some embodiments, methods are provided for diagnosing multiple heritable disorders in parallel at a pre-implantation, prenatal, perinatal, or postnatal stage. In some embodiments, the disclosure provides methods for analyzing multiple genetic loci (e.g., a plurality of target nucleic acids selected from Table 1) from a patient sample, such as a blood, pre-implantation embryo, chorionic villus or amniotic fluid sample, or other sample (e.g., other biological fluid or tissue sample such as a biopsy sample) as aspects of the invention are not limited in this respect.

Other samples may include tumor tissue or circulating tumor cells. In some embodiments, a patient sample (*e.g.*, a tumor tissue or cell sample) is mosaic for one or more mutations of interest, and thus, may require higher sensitivity than is needed for a germline mutation analysis. In some embodiments, a sample comprises cells from a non-host organism (*e.g.*, bacterial or viral infections in a human subject) or a sample for environmental monitoring (*e.g.*, bacterial, viral, fungal composition of a soil, water, or air sample).

Accordingly, in some embodiments, aspects of the methods disclosed herein relate to genotyping a polymorphism of a target nucleic acid. In some embodiments, the genotyping may comprise determining that one or more alleles of the target nucleic acid are heterozygous or homozygous. In further embodiments, the genotyping may comprise determining the sequence of a polymorphism and comparing that sequence to a control sequence that is indicative of a disease risk. In some embodiments, the polymorphism is selected from a locus in Table 1 or Table 2. However, it should be appreciated that any locus associated with a disease or condition of interest may be used.

In some embodiments, a diagnosis, prognosis, or disease risk assessment is provided to a subject based on a genotype determined for that subject at one or more genetic loci (e.g., based on the analysis of a biological sample obtained from that subject). In some embodiments, an assessment is provided to a couple, based on their respective genotypes at one or more genetic

Petitioner Illumina, Inc., EX1004, p. 564

loci, of the risk of their having one or more children having a genotype associated with a disease or condition (e.g., a homozygous or heterozygous genotype associated with a disease or condition). In some embodiments, a subject or a couple may seek genetic or reproductive counseling in connection with a genotype determined according to embodiments of the invention. In some embodiments, genetic information from a tumor or circulating tumor cells is used to determine prognosis and guide selection of appropriate drugs/treatments.

It should be appreciated that any of the methods or compositions described herein may be used in combination with any of the medical evaluations associated with one or more genetic loci as described herein.

In some embodiments, aspects of the invention provide effective methods for overcoming challenges associated with systematic errors (bias) and/or stochastic effects in multiplex genomic capture and/or analysis (including sequencing analysis). In some embodiments, aspects of the invention are useful to avoid, reduce and/or account for variability in one or more sampling and/or analytical steps. For example, in some embodiments, variability in target nucleic acid representation and unequal sampling of heterozygous alleles in pools of captured target nucleic acids can be overcome.

Accordingly, in some embodiments, the disclosure provides methods that reduce variability in the detection of target nucleic acids in multiplex capture methods. In other embodiments, methods improve allelic representation in a capture pool and, thus, improve variant detection outcomes. In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve the use of different sets of multiple probes (e.g., molecular inversion probes MIPs) that capture overlapping regions of a target nucleic acid to achieve a more uniform representation of the target nucleic acids in a capture pool compared with methods of the prior art. In other embodiments, methods reduce bias, or the risk of bias, associated with large scale parallel capture of genetic loci, e.g., for diagnostic purposes. In other embodiments, methods are provided for increasing reproducibility (e.g., by reducing the effect of polymorphisms on target nucleic acid capture) in the detection of a plurality of genetic loci in parallel. In further embodiments, methods are provided for reducing the effect of probe synthesis and/or probe amplification variability on the analysis of a plurality of genetic loci in parallel.

Petitioner Illumina, Inc., EX1004, p. 565

According to some aspects, methods of analyzing a plurality of genetic loci are provided. In some embodiments, the methods comprise contacting each of a plurality of target nucleic acids with a probe set, wherein each probe set comprises a plurality of different probes, each probe having a central region flanked by a 5' region and a 3' region that are complementary to nucleic acids flanking the same strand of one of a plurality of subregions of the target nucleic acid, wherein the subregions of the target nucleic acid are different, and wherein each subregion overlaps with at least one other subregion, isolating a plurality of nucleic acids each having a nucleic acid sequence of a different subregion for each of the plurality of target nucleic acids, and analyzing the isolated nucleic acids.

In other embodiments, methods comprise contacting each of a plurality of target nucleic acids with a probe set, wherein each probe set comprises a plurality of different probes, each probe having a central region flanked by a 5' region and a 3' region that are complementary to nucleic acids flanking the same strand of one of a plurality of subregions of the target nucleic acid, wherein the subregions of the target nucleic acid are different, and wherein a portion of the 5' region and a portion of the 3' region of a probe have, respectively, the sequence of the 5' region and the sequence of the 3' region of a different probe, isolating a plurality of nucleic acids each having a nucleic acid sequence of a different subregion for each of the plurality of target nucleic acids, and analyzing the isolated nucleic acids.

Aspects of the disclosure are based, in part, on the discovery of methods for overcoming problems associated with systematic and random errors (bias) in genome capture, amplification and sequencing methods, namely high variability in the capture and amplification of nucleic acids and disproportionate representation of heterozygous alleles in sequencing libraries. Accordingly, in some embodiments, the disclosure provides methods that reduce errors associated with the variability in the capture and amplification of nucleic acids. In other embodiments, the methods improve allelic representation in sequencing libraries and, thus, improve variant detection outcomes. In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve the use of differentiator tag sequences to uniquely tag individual nucleic acid molecules. In some embodiments, the differentiator tag sequence permit the detection of bias based on the occurrence of combinations of differentiator tag and target sequences observed in a sequencing reaction. In other embodiments, the methods reduce errors caused by bias, or the risk of bias,

Petitioner Illumina, Inc., EX1004, p. 566

associated with the capture, amplification and sequencing of genetic loci, e.g., for diagnostic purposes.

Aspects of the invention relate to providing sequence tags (referred to as differentiator tags) that are useful to determine whether target nucleic acid sequences identified in an assay are from independently isolated target nucleic acids or from multiple copies of the same target nucleic acid molecule (e.g., due to bias in a preparative step, for example, amplification). This information can be used to help analyze a threshold number of independently isolated target nucleic acids from a biological sample in order to obtain sequence information that is reliable and can be used to make a genotype conclusion (e.g., call) with a desired degree of confidence. This information also can be used to detect bias in one or more nucleic acid preparative steps.

In some embodiments, the methods disclosed herein are useful for any application where reduction of bias, e.g., associated with genomic isolation, amplification, sequencing, is important. For example, detection of cancer mutations in a heterogeneous tissue sample, detection of mutations in maternally-circulating fetal DNA, and detection of mutations in cells isolated during a preimplantation genetic diagnostic procedure.

Accordingly, in some aspects, methods of genotyping a subject are provided. In some embodiments, the methods comprise determining the sequence of at least a threshold number of independently isolated nucleic acids, wherein the sequence of each isolated nucleic acid comprises a target nucleic acid sequence and a differentiator tag sequence, wherein the threshold number is a number of unique combinations of target nucleic acid and differentiator tag sequences, wherein the isolated nucleic acids are identified as independently isolated if they comprise unique combinations of target nucleic acid and differentiator tag sequences, and wherein the target nucleic acid sequence is the sequence of a genomic locus of a subject.

In some embodiments, the isolated nucleic acids are products of a circularization selection-based preparative method, e.g., molecular inversion probe capture products. In other embodiments, the isolated nucleic acids are products of an amplification-based preparative methods. In other embodiments, the isolated nucleic acids are products of hybridization-based preparative methods.

Circularization selection-based preparative methods selectively convert regions of interest (target nucleic acids) into a covalently-closed circular molecule which is then isolated typically by removal (usually enzymatic, e.g. with exonuclease) of any non-circularized linear

Petitioner Illumina, Inc., EX1004, p. 567

nucleic acid. Oligonucleotide probes (e.g., molecular inversion probes) are designed which have ends that flank the region of interest (target nucleic acid) and, optionally, primer sites, e.g., sequencing primer sites. The probes are allowed to hybridize to the genomic target, and enzymes are used to first (optionally) fill in any gap between probe ends and second ligate the probe closed. Following circularization, any remaining (non-target) linear nucleic acid is typically removed, resulting in isolation (capture) of target nucleic acid. Circularization selection-based preparative methods include molecular inversion probe capture reactions and 'selector' capture reactions. In some embodiments, molecular inversion probe capture of a target nucleic acid is indicative of the presence of a polymorphism in the target nucleic acid.

In amplification-based (e.g., PCR-based or LCR-based, etc.) preparative methods, genomic loci (target nucleic acids) are isolated directly by means of a polymerase chain reaction or ligase chain reaction (or other amplification method) that selectively amplifies each locus using one or more oligonucleotide primers. It is to be understood that primers will be sufficiently complementary to the target sequence to hybridize with and prime amplification of the target nucleic acid. Any one of a variety of art known methods may be utilized for primer design and synthesis. One or more of the primers may be perfectly complementary to the target sequence. Degenerate primers may also be used. Primers may also include additional nucleic acids that are not complementary to target sequences but that facilitate downstream applications, including for example restriction sites and differentiator tag sequences. Amplification-based methods include amplification of a single target nucleic acid and multiplex amplification (amplification of multiple target nucleic acids in parallel).

Hybridization-based preparative methods involve selectively immobilizing target nucleic acids for further manipulation. It is to be understood that one or more oligonucleotides (immobilization oligonucleotides), which comprise differentiator tag sequences, and which may be from 15 to 170 nucleotides in length, are used which hybridize along the length of a target region of a genetic locus to immobilize it. In some embodiments, immobilization oligonucleotides, are either immobilized before hybridization is performed (e.g., Roche/Nimblegen 'sequence capture'), or are prepared such that they include a moiety (e.g. biotin) which can be used to selectively immobilize the target nucleic acid after hybridization by binding to e.g., streptavidin-coated microbeads (e.g. Agilent 'SureSelect').

Petitioner Illumina, Inc., EX1004, p. 568

It should be appreciated that any of the circularization, amplification, and/or hybridization based methods described herein may be used in connection with one or more of the tiling/staggering, tagging, size-detection, and/or sensitivity enhancing algorithms described herein.

In some embodiments, the methods disclosed herein comprise determining the sequence of molecular inversion probe capture products, each comprising a molecular inversion probe and a target nucleic acid, wherein the sequence of the molecular inversion probe comprises a differentiator tag sequence and, optionally, a primer sequence, and wherein the target nucleic acid is a captured genomic locus of a subject, and genotyping the subject at the captured genomic locus based on the sequence of at least a threshold number of unique combinations of target nucleic acid and differentiator tag sequences of molecular inversion probe capture products.

In some embodiments, the methods disclosed herein comprise obtaining molecular inversion probe capture products, each comprising a molecular inversion probe and a target nucleic acid, wherein the sequence of the molecular inversion probe comprises a differentiator tag sequence and, optionally, a primer sequence, wherein the target nucleic acid is a captured genomic locus of the subject, amplifying the molecular inversion probe capture products, and genotyping the subject by determining, for each target nucleic acid, the sequence of at least a threshold number of unique combinations of target nucleic acid and differentiator tag sequence of molecular inversion probe capture products. In certain embodiments, obtaining comprises capturing target nucleic acids from a genomic sample of the subject with molecular inversion probes, each comprising a unique differentiator tag sequence. In specific embodiments, capturing is performed under conditions wherein the likelihood of obtaining two or more molecular inversion probe capture products with identical combinations of target and differentiator tag sequences is equal to or less than a predetermined value, optionally wherein the predetermined value is about 0.05.

In one embodiment, the threshold number for a specific target nucleic acid sequence is selected based on a desired statistical confidence for the genotype. In some embodiments, the methods further comprising determining a statistical confidence for the genotype based on the number of unique combinations of target nucleic acid and differentiator tag sequences.

According to some aspects, methods of analyzing a plurality of genetic loci are provided. In some embodiments, the methods comprise obtaining a plurality of molecular inversion probe

Petitioner Illumina, Inc., EX1004, p. 569

capture products each comprising a molecular inversion probe and a target nucleic acid, wherein the sequence of the molecular inversion probe comprises a differentiator tag sequence and, optionally, a primer sequence (e.g., a sequence that is complementary to the sequence of a nucleic acid that is used as a primer for sequencing or other extension reaction), amplifying the plurality of molecular inversion probe capture products, determining numbers of occurrence of combinations of target nucleic acid and differentiator tag sequence of molecular inversion probe capture products in the amplified plurality, and if the number of occurrence of a specific combination of target nucleic acid sequence and differentiator tag sequence exceeds a predetermined value, detecting bias in the amplification of the molecular inversion probe comprising the specific combination. In some embodiments, the methods further comprise genotyping target sequences in the plurality, wherein the genotyping comprises correcting for bias, if detected.

In some embodiments, the target nucleic acid is a gene (or portion thereof) selected from Table 1. In some embodiments, the genotyping comprises determining the sequence of a target nucleic acid (e.g., a polymorphic sequence) at one or more (both) alleles of a genome (a diploid genome) of a subject. In certain embodiments, the genotyping comprises determining the sequence of a target nucleic acid at both alleles of a diploid genome of a subject, wherein in the target nucleic acid comprises, or consists of, a sequence of Table 1, Table 2, or other locus of interest.

In some embodiments, aspects of the invention provide methods and compositions for identifying nucleic acid insertions or deletions in genomic regions of interest without determining the nucleotide sequences of these regions. Aspects of the invention are particularly useful for detecting nucleic acid insertions or deletions in genomic regions containing nucleic acid sequence repeats (e.g., di- or tri-nucleotide repeats). However, the invention is not limited to analyzing nucleic acid repeats and may be used to detect insertions or deletions in any target nucleic acid of interest. Aspects of the invention are particularly useful for analyzing multiple loci in a multiplex assay.

In some embodiments, aspects of the invention relate to determining whether an amount of target nucleic acid that is captured in a genomic capture assay is higher or lower than expected. In some embodiments, a statistically significant deviation from an expected amount (e.g., higher or lower) is indicative of the presence of a nucleic acid insertion or deletion in the

Petitioner Illumina, Inc., EX1004, p. 570

genomic region of interest. In some embodiments, the amount is a number of nucleic acid molecules that are captured. In some embodiments, the amount is a number of independently captured nucleic acid molecules in a sample. It should be appreciated that the captured nucleic acids may be literally captured from a sample, or their sequences may be captured without actually capturing the original nucleic acids in the sample. For example, nucleic acid sequences may be captured in an assay that involves a template-based extension of nucleic acids having the region of interest, in the sample.

Aspects of the invention are based on the recognition that the efficiency of certain capture techniques is affected by the length of the nucleic acid being captured.

Accordingly, an increase or decrease in the length of a target nucleic acid (e.g., due to an insertion or deletion of a repeated sequence) can alter the capture efficiency of that nucleic acid. In some embodiments, a difference in the capture efficiency (e.g., a statistically significant difference in the capture efficiency) of a target nucleic acid is indicative of an insertion or deletion in the target nucleic acid. It should be appreciated that the capture efficiency for a target nucleic acid may be evaluated based on an amount of captured nucleic acid (e.g., number of captured nucleic acid molecules) relative to a control amount (e.g., based on an amount of control nucleic acid that is captured). However, the invention is not limited in this respect and other techniques for evaluating capture efficiency also may be used.

According to aspects of the invention, evaluating the capture efficiency as opposed to determining the sequence of the entire repeat region reduces errors associated with sequencing through repeat regions. Repeat sequences often give rise to stutters or skips in sequencing reactions that make it very difficult to accurately determine the number of repeats in a target region without running multiple sequencing reactions under different conditions and carefully analyzing the results. Such procedures are cumbersome and not readily scalable in a manner that is consistent with high throughput analyses of target nucleic acids. In some embodiments, repeat regions may be longer than the length of the individual sequence read, making length determination on the basis of a single read impossible. For example, when using next-generation sequencing the repeat regions may be longer than the length of the individual sequence read, making length determination on the basis of a single read impossible. Accordingly, aspects of the invention are useful to increase the sensitivity of detecting insertions or deletions in target regions, particularly target regions containing repeated sequences.

Petitioner Illumina, Inc., EX1004, p. 571

In some embodiments, aspects of the invention relate to capturing genomic nucleic acid sequences using a molecular inversion probe (e.g., MIP or Padlock probe) technique, and determining whether the amount (e.g., number) of captured sequences is higher or lower than expected. In some embodiments, the amount (e.g., number) of captured sequences is compared to an amount (e.g., number) of sequences captured in a control assay. The control assay may involve analyzing a control sample that contains a nucleic acid from the same genetic locus having a known sequence length (e.g., a known number of nucleic acid repeats). However, a control may involve analyzing a second (e.g., different) genetic locus that is not expected to contain any insertions or deletions. The second genetic locus may be analyzed in the same sample as the locus being interrogated or in a different sample where its length has been previously determined. The second genetic locus may be a locus that is not characterized by the presence of nucleic acid repeats (and thus not expected to contain insertions or deletions of the repeat sequence).

In some embodiments, a target nucleic acid region that is being evaluated may be determined by the identity of the targeting arms of a probe that is designed to capture the target region (or sequence thereof). For example, the targeting arms of a MIP probe may be designed to be complementary (e.g., sufficiently complementary for selective hybridization and//or polymerase extension and/or ligation) to genomic regions flanking a target region suspected of containing an insertion or deletion. It should be appreciated that two targeting arms may be designed to be complementary (e.g., sufficiently complementary for selective hybridization and/or polymerase extension and/or ligation) to the two flanking regions that are immediately adjacent (e.g., immediately 5' and 3', respectively) to a region of a sequence repeat on one strand of a genomic nucleic acid. However, one or both targeting arms may be designed to hybridize several bases (e.g., 1-5, 5-10, 10-25, 25-50, or more) upstream or downstream from the repeat region in such a way that the captured sequence includes a region of unique genomic sequence that on one or both sides of the repeat region. This unique region can then be used to identify the captured target (e.g., based on sequence or hybridization information).

In some embodiments, two or more (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10 or more) different loci may be interrogated in parallel in a single assay (e.g., in a multiplex assay). In some embodiments, the ratio of captured nucleic acids for each locus may be used to determine whether a nucleic acid insertion or deletion is present in one locus relative to the other. For

Petitioner Illumina, Inc., EX1004, p. 572

example, the ratio may be compared to a control ratio that is representative of the two loci when neither one has an insertion or deletion relative to control sequences (e.g., sequences that are normal or known to be associated with healthy phenotypes for those loci). However, the amount of captured nucleic acids may be compared to any suitable control as discussed herein.

The locus of a captured sequence may be identified by determining a portion of unique sequence 5' and/or 3' to the repeat region in the target nucleic acid suspected of containing a deletion or insertion. This does not require sequencing the captured repeat region itself. However, some or all of the repeat region also could be sequenced as aspects of the invention are not limited in this respect.

Aspects of the invention may be combined with one or more sequence-based assays (e.g., SNP detection assays), for example in a multiplex format, to determine the genotype of one or more regions of a subject.

In some embodiments, methods of detecting a polymorphism in a nucleic acid in a biological sample are provided. In some embodiments, the methods comprise evaluating the efficiency of capture at one or more loci and determining whether one or both alleles at that locus contain an insertion or deletion relative to a control locus (e.g., a locus indicative of a length of repeat sequence that is associated with a healthy phenotype).

Accordingly, aspects of the invention relate to methods for determining whether a target nucleic acid has an abnormal length by evaluating the capture efficiency of a target nucleic acid in a biological sample from a subject, wherein a capture efficiency that is different from a reference capture efficiency is indicative of the presence, in the biological sample, of a target nucleic acid having an abnormal length. It should be appreciated that the term "abnormal" is a relative term based on a comparison to a "normal" length. In some embodiments, a normal length is a length that is associated with a normal (e.g., healthy or non-carrier phenotype). Accordingly, an abnormal length is a length that is either shorter or longer than the normal length. In some embodiments, the presence of an abnormal length is indicative of an increased risk that the locus is associated with a disease or a disease carrier phenotype. In some embodiments, the abnormal length is indicative that the subject is either has a disease or condition or is a carrier of a disease or condition (e.g., associated with the locus). However, it should be appreciated that the description of embodiments relating to detecting the presence of

Petitioner Illumina, Inc., EX1004, p. 573

an abnormal length also support detecting the presence of a length that is different from an expected or control length.

In some embodiments, aspects of the invention relate to estimating the length of a target nucleic acid (e.g., of a sub-target region within a target nucleic acid). In some embodiments, aspects of the invention relate to methods for estimating the length of a target nucleic acid by contacting the target nucleic acid with a plurality of detection probes under conditions that permit hybridization of the detection probes to the target nucleic acid, wherein each detection probe is a polynucleotide that comprises a first arm that hybridizes to a first region of the target nucleic acid and a second arm that hybridizes to a second region of the target nucleic acid, wherein the first and second regions are on a common strand of the target nucleic acid, and wherein the nucleotide sequence of the target between the 5' end of the first region and the 3' end of the second region is the nucleotide sequence of a sub-target nucleic acid; and capturing a plurality of sub-target nucleic acids that are hybridized with the plurality of detection probes; and measuring the frequency of occurrence of a sub-target nucleic acid in the plurality of sub-target nucleic acids, wherein the frequency of occurrence of the sub-target nucleic acid in the plurality of sub-target nucleic acids is indicative of the length of the sub-target nucleic acid. It should be appreciated that methods for estimating a nucleic acid length may involve comparing a capture efficiency for a target nucleic acid region to two or more reference efficiencies for known nucleic acid lengths in order to determine whether the target nucleic acid region is smaller, intermediate, or larger in size than the known control lengths. In some embodiments, a series of nucleic acids of known different lengths may be used to provide a calibration curve for evaluating the length of a target nucleic acid region of interest.

In some embodiments, the capture efficiency of a target region suspected of having a deletion or insertion is determined by comparing the capture efficiency to a reference indicative of a normal capture efficiency. In some embodiments, the capture efficiency is lower than the reference capture efficiency. In some embodiments, the subject is identified as having an insertion in the target region. In some embodiments, the capture efficiency is higher than the reference capture efficiency. In some embodiments, the subject is identified as having a deletion in the target region. In some embodiments, the subject is identified as being heterozygous for the insertion. In some embodiments, the subject is identified as being heterozygous for the deletion.

Petitioner Illumina, Inc., EX1004, p. 574

In some embodiments of any of the methods described herein (e.g., tiling/staggering, tagging, size-detection, and/or sensitivity enhancement) aspects of the invention relate to capturing a sub-target nucleic acid (or a sequence of a sub-target nucleic acid). In some embodiments, a molecular inversion probe technique is used. In some embodiments, a molecular inversion probe is a single linear strand of nucleic acid that comprises a first targeting arm at its 5' end and a second targeting arm at its 3' end, wherein the first targeting arm is capable of specifically hybridizing to a first region flanking one end of the sub-target nucleic acid, and wherein the second targeting arm is capable of specifically hybridizing to a second region flanking the other end of the sub target nucleic acid on the same strand of the target nucleic acid. In some embodiments, the first and second targeting arms are between about 10 and about 100 nucleotides long. In some embodiments, the first and second targeting arms are about 10-20, 20-30, 30-40, or 40-50 nucleotides long. In some embodiments, the first and second targeting arms are about 20 nucleotides long. In some embodiments, the first and second targeting arms have the same length. In some embodiments, the first and second targeting arms have different lengths. In some embodiments, each pair of first and second targeting arms in a set of probes has the same length. Accordingly, if one of the targeting arms is longer, the other one is correspondingly shorter. This allows for a quality control step in some embodiments to confirm that all captured probe/target sequence products have the same length after a multiplexed plurality of capture reactions. In some embodiments, a set of probes may be designed to have the same length if the intervening region is varied to accommodate any differences in the length of either one or both of the first and second targeting arms.

In some embodiments, the hybridization Tms of the first and second targeting arms are similar. In some embodiments, the hybridization Tms of the first and second targeting arms are within 2-5 °C of each other. In some embodiments, the hybridization Tms of the first and second targeting arms are identical. In some embodiments, the hybridization Tms of the first and second targeting arms are close to empirically-determined optima but not necessarily identical.

In some embodiments, the first and second targeting arms of a molecular inversion probe have different Tms. For example, the Tm of the first targeting arm (at the 5' end of the molecular inversion probe) may be higher than the Tm of the second targeting arm (at the 3' end of the molecular inversion probe). According to aspects of the invention, and without wishing to be bound by theory, a relatively high Tm for the first targeting arm may help avoid or prevent the

Petitioner Illumina, Inc., EX1004, p. 575

first targeting arm from being displaced after hybridization by the extension product of the 3' end of the second targeting arm. It should be appreciated that a reference to the Tm of a targeting arm as used herein relates to the Tm of hybridization of the targeting arm to a nucleic acid having the complementary sequence (e.g., the region of the target nucleic acid that has a sequence that is complementary to the sequence of the targeting arm). It also should be appreciated that the Tms of the targeting arms described herein may be calculated using any appropriate method. For example, in some embodiments an experimental method (e.g., a gel shift assay, a hybridization assay, a melting curve analysis, for example in a PCR machine with a SYBR dye by stepping through a temperature ramp while monitoring signal level from an intercalating dye, for example, bound to a double-stranded DNA, etc.) may be used to determine one or more Tms empirically. In some embodiments, an optimal Tm may be determined by evaluating the number of products formed (e.g., for each of a plurality of MIP probes), and determining the optimal Tm as the center point in a histogram of Tm for all targeting arms. In some embodiments, a predictive algorithm may be used to determine a Tm theoretically. In some embodiments, a relatively simple predictive algorithm may be used based on the number of G/C and A/T base pairs when the sequence is hybridized to its target and/or the length of the hybridized product (e.g., for example, 64.9 + 41 * ([G+C] - 16.4) / (A + T + G + C), see for example, Wallace,R.B., Shaffer,J., Murphy,R.F., Bonner,J., Hirose,T., and Itakura,K. (1979) Nucleic Acids Res 6:3543-3557). In some embodiments, a more complex algorithm may be used to account for the effects of base stacking entropy and enthalpy, ion concentration, and primer concentration (see, for example, SantaLucia J (1998), Proc Natl Acad Sci USA, 95:1460-5). In some embodiments an algorithm may use modified parameters (e.g., nearest-neighbor parameters for basepair entropy/enthalpy values). It should be appreciated that any suitable algorithm may be used as aspects of the invention are not limited in this respect. However, it also should be appreciated that different methodologies may results in different calculated or predicted Tms for the same sequences. Accordingly, in some embodiments, the same empirical and/or theoretical method is used to determine the Tms of different sequences for a set of probes to avoid a negative impact of any systematic difference in the Tm determination or prediction when designing a set of probes with predetermined similarities or differences for different Tms.

In some embodiments, the Tm of the first targeting arm may be about 1 °C, about 2 °C, about 3 °C, about 4 °C, about 5 °C, or more than about 5 °C higher than the Tm of the second

Petitioner Illumina, Inc., EX1004, p. 576

targeting arm. In some embodiments, each probe in a plurality of probes (e.g., each probe in a set of 5-10, each probe in a set of at least 10, each probe in a set of 10-50, each probe in a set of 50-100, each probe in a set of 100-500, each probe in a set of 500-1,000, each probe in a set of 1,000-1,500, each probe in a set of 1,500-2,000, each probe in a set of 2,000-3,000, 3,000-5,000, 5,000-10,000 or each probe in a set of at least 5,000 different probes) has a unique first targeting arm (e.g., they all have different sequences) and a unique second targeting arm (e.g., they all have different sequences). In some embodiments, for at least 10% of the probes (e.g., at least 25%, 25%-50%, 50%- 75%, 75%-90%, 90%-95% or over 95%, or all of the probes) the first targeting arm has a Tm for its complementary sequence that is higher (e.g., about 1 °C, about 2 °C, about 3 °C, about 4 °C, about 5 °C, or more than about 5 °C higher) than the Tm of the second targeting arm for its complementary sequence. In some embodiments, each of the first targeting arms have similar or identical Tms for their respective complementary sequences and each of the second targeting arms have similar or identical Tms for their respective complementary sequences (and the first targeting arms have higher Tms than the second targeting arms). For example, in some embodiments, the Tm of the first arm(s) may be about 58 °C and the Tm of the second arm(s) may be about 56 °C. In some embodiments, the Tm of the first arm(s) may be about 68 °C, and the Tm of the second arm(s) may be about 65 °C. It should be appreciated that in some embodiments the similarity (e.g., within a range of 1 °C, 2 °C, 3 °C, 4 °C, 5 °C) or identity of the Tms for the different targeting arms should be based either on empirical data for each arm or based on the same predictive algorithm for each arm (e.g., Wallace,R.B., Shaffer,J., Murphy,R.F., Bonner,J., Hirose,T., and Itakura,K. (1979) Nucleic Acids Res 6:3543-3557, SantaLucia J (1998), Proc Natl Acad Sci USA, 95:1460-5, or other algorithm).

In some embodiments, the Tm of the first targeting arm of a molecular inversion probe (at the 5' end of the molecular inversion probe) is selected to be sufficiently stable to prevent displacement of the first targeting arm from its complementary sequence on a target nucleic acid. In some embodiments, the Tm of the first targeting arm is 50-55 °C, at least 55 °C, 55-60 °C, at least 60 °C, 60-65 °C, at least 65 °C, at least 70 °C, at least 75 °C, or at least 80 °C. As discussed above, it should be appreciated that the for a particular targeting arm may be determined empirically or theoretically. Different theoretical models may be used to determine a Tm and it should be appreciated that the predicted Tm for a particular sequence may be different

Petitioner Illumina, Inc., EX1004, p. 577

depending on the algorithm used for the prediction. In some embodiments, each probe in a plurality of probes (e.g., each probe in a set of 5-10, each probe in a set of at least 10, each probe in a set of 10-50, each probe in a set of 50-100, each probe in a set of 100-500, or each probe in a set of at least 500 different probes) has a different first targeting arm (e.g., different sequences) but each different first targeting arm has a similar or identical Tm for its complementary sequence on a target nucleic acid. It should be appreciated that in some embodiments the similarity (e.g., within a range of I C, 2 C, 3 C, 4 C, 5 C) or identity of the Tms for the different targeting arms should be based either on empirical data for each arm or based on the same predictive algorithm for each arm (e.g., Wallace,R.B., Shaffer,J., Murphy,R.F., Bonner,J., Hirose,T., and Itakura,K. (1979) Nucleic Acids Res 6:3543- 3557, SantaLucia J (1998), Proc Natl Acad Sci USA, 95:1460-5, or other algorithm).

In some embodiments, the sub-target nucleic acid contains a nucleic acid repeat. In some embodiments, the nucleic acid repeat is a dinucleotide or trinucleotide repeat. In some embodiments, the sub-target nucleic acid contains 10-100 copies of the nucleic acid repeat in the absence of an abnormal increase or decrease in nucleic acid repeats. In some embodiments, the sub-target nucleic acid is a region of the Fragile-X locus that contains a nucleic acid repeat. In some embodiments, one or both targeting arms hybridize to a region on the target nucleic acid that is immediately adjacent to a region of nucleic acid repeats. In some embodiments, one or both targeting arms hybridize to a region on the target nucleic acid that is separated from a region of nucleic acid repeats by a region that does not contain any nucleic acid repeats. In some embodiments, the molecular inversion probe further comprises a primer-binding region that can be used to sequence the captured sub-target nucleic acid and optionally the first and/or second targeting arm.

In some embodiments, aspects of the invention relate to evaluating the length of a plurality of different target nucleic acids in a biological sample. In some embodiments, the plurality of target nucleic acids are analyzed using a plurality of different molecular inversion probes. In some embodiments, each different molecular inversion probe comprises a different pair of first and second targeting arms at each of the 3' and 5' ends. In some embodiments, each different molecular inversion probe comprises the same primer-binding sequence.

In some embodiments, aspects of the invention relate to analyzing nucleic acid from a biological sample obtained from a subject. In some embodiments, the biological sample is a

Petitioner Illumina, Inc., EX1004, p. 578

blood sample. In some embodiments, the biological sample is a tissue sample, specific cell population, tumor sample, circulating tumor cells, or environmental sample. In some embodiments, the biological sample is a single cell. In some embodiments, nucleic acids are analyzed in biological samples obtained from a plurality of different subjects. In some embodiments, nucleic acids from a biological sample are analyzed in multiplex reactions. It should be appreciated that a biological sample contains a plurality of copies of a genome derived from a plurality of cells in the sample. Accordingly, a sample may contain a plurality of independent copies of a target nucleic acid region of interest, the capture efficiency of which can be used to evaluate its size as described herein.

In some embodiments, aspects of the invention relate to evaluating a nucleic acid capture efficiency by determining an amount of target nucleic acid that is captured (e.g., an amount of sub-target nucleic acid sequences that are captured). In some embodiments, the amount of target nucleic acid that is captured is determined by determining a number of independently captured target nucleic acid molecules (e.g., the amount of independently captured molecules that have the sequence of the sub-target region). In some embodiments, the amount of target nucleic acid that is captured is compared to a reference amount of captured nucleic acid. In some embodiments, the reference amount is determined by determining a number of independently captured molecules of a reference nucleic acid. In some embodiments, the reference nucleic acid is a nucleic acid of a different locus in the biological sample that is not suspected of containing a deletion or insertion. In some embodiments, the reference nucleic acid is a nucleic acid of known size and amount that is added to the capture reaction. As described herein, a number of independently captured nucleic acid sequences can be determined by contacting a nucleic acid sample with a preparation of a probe (e.g., a MIP probe as described herein). It should be appreciated that the preparation may comprise a plurality of copies of the same probe and accordingly a plurality of independent copies of the target region may be captured by different probe molecules. The number of probe molecules that actually capture a sequence can be evaluated by determining an amount or number of captured molecules using any suitable technique. This number is a reflection of both the number of target molecules in the sample and the efficiency of capture of those target molecules, which in turn is related to the size of the target molecules as described herein. Accordingly, the capture efficiency can be evaluated by controlling for the abundance of the target nucleic acid, for example by comparing the number or

Petitioner Illumina, Inc., EX1004, p. 579

amount of captured target molecules to an appropriate control (e.g., a known size and amount of control nucleic acid, or a different locus that should be present in the same amount in the biological sample and is not expected to contain any insertions or deletions). It should be appreciated that other factors may affect the capture efficiency of a particular target nucleic acid region (e.g., the sequence of the region, the GC content, the presence of secondary structures, etc.). However, these factors also can be accounted for by using appropriate controls (e.g., known sequences having similar properties, the same sequences, other genomic sequences expected to be present in the biological sample at the same frequency, etc., or any combination thereof).

In some embodiments, aspects of the invention relate to identifying a subject as having an insertion or deletion in one or more alleles of a genetic locus if the capture efficiency for that genetic locus is statistically significantly different than a reference capture efficiency.

It should be appreciated that hybridization conditions used for any of the capture techniques described herein (e.g., MIP capture techniques) can be based on known hybridization buffers and conditions.

In some embodiments, the methods disclosed herein are useful for any application where the detection of deletions or insertions is important.

In some embodiments, aspects of the invention relate to basing a nucleic acid sequence analysis on results from two or more different nucleic acid preparatory techniques that have different systematic biases in the types of nucleic acids that they sample. According to the invention, different techniques have different sequence biases that are systematic and not simply due to stochastic effects during nucleic acid capture or amplification. Accordingly, the degree of oversampling required to overcome variations in nucleic acid preparation needs to be sufficient to overcome the biases (e.g., an oversampling of 2-5 fold, 5-10 fold, 5-15 fold, 15-20 fold, 20-30 fold, 30-50 fold, or intermediate to higher fold).

According to some embodiments, different techniques have different characteristic or systematic biases. For example, one technique may bias a sample analysis towards one particular allele at a genetic locus of interest, whereas a different technique would bias the sample analysis towards a different allele at the same locus. Accordingly, the same sample may be identified as being different depending on the type of technique that is used to prepare nucleic

Petitioner Illumina, Inc., EX1004, p. 580

acid for sequence analysis. This effectively represents a sensitivity limitation, because each technique has different relative sensitivities for polymorphic sequences of interest.

According to aspects of the invention, the sensitivity of a nucleic acid analysis can be increased by combining the sequences from different nucleic acid preparative steps and using the combined sequence information for a diagnostic assay (e.g., for a making a call as to whether a subject is homozygous or heterozygous at a genetic locus of interest).

In some embodiments, the invention provides a method of increasing the sensitivity of a nucleic acid detection assay by obtaining a first preparation of a target nucleic acid using a first preparative method on a biological sample, obtaining a second preparation of a target nucleic acid using a second preparative method on the biological sample, assaying the sequences obtained in both first and second nucleic acid preparations, and using the sequence information from both first and second nucleic acid preparations to determine the genotype of the target nucleic acid in the biological sample, wherein the first and second preparative methods have different systematic sequence biases. In some embodiments, the first and second nucleic acid preparations are combined prior to performing a sequence assay. In some embodiments, separate sequence assays are performed on the first and second nucleic acid preparations and the sequence information from both assays are combined to determine the genotype of the target nucleic acid in the biological sample. In some embodiments, the first preparative method is an amplification-based, a hybridization-based, or a circular probe-based preparative method. In some embodiments, the second method is an amplification-based, a hybridization-based, or a circular probe-based preparative method. In some embodiments, the first and second methods are of different types (e.g., only one of them is an amplification-based, a hybridization-based, or a circular probe-based preparative method, and the other one is one or the other two types of method). Accordingly, in some embodiments the second preparative method is an amplification-based, a hybridization-based, or a circular probe-based preparative method, provided that the second method is different from the first method. However, in some embodiments, both methods may be of the same type, provided they are different methods (e.g., both are amplification based or hybridization-based, but are different types of amplification or hybridization methods, e.g., with different relative biases).

In amplification-based (e.g., PCR-based or LCR-based, etc.) preparative methods, genomic loci (target nucleic acids) are isolated directly by means of a polymerase chain reaction

Petitioner Illumina, Inc., EX1004, p. 581

or ligase chain reaction (or other amplification method) that selectively amplifies each locus using a pair of oligonucleotide primers. It is to be understood that primers will be sufficiently complementary to the target sequence to hybridize with and prime amplification of the target nucleic acid. Any one of a variety of art known methods may be utilized for primer design and synthesis. One or both of the primers may be perfectly complementary to the target sequence. Degenerate primers may also be used. Primers may also include additional nucleic acids that are not complementary to target sequences but that facilitate downstream applications, including for example restriction sites and identifier sequences (e.g., source sequences). PCR based methods may include amplification of a single target nucleic acid and multiplex amplification (amplification of multiple target nucleic acids in parallel).

Hybridization-based preparative may methods involve selectively immobilizing target nucleic acids for further manipulation. It is to be understood that one or more oligonucleotides (immobilization oligonucleotides), which in some embodiments may be from 10 to 200 nucleotides in length, are used which hybridize along the length of a target region of a genetic locus to immobilize it. In some embodiments, immobilization oligonucleotides are either immobilized before hybridization is performed (e.g., Roche/Nimblegen 'sequence capture'), or are prepared such that they include a moiety (e.g., biotin) which can be used to selectively immobilize the target nucleic acid after hybridization by binding to e.g., streptavidin-coated microbeads (e.g., Agilent 'SureSelect').

Circularization selection-based preparative methods selectively convert each region of interest into a covalently-closed circular molecule which is then isolated by removal (usually enzymatic, e.g., with exonuclease) of any non-circularized linear nucleic acid. Oligonucleotide probes are designed which have ends that flank the region of interest. The probes are allowed to hybridize to the genomic target, and enzymes are used to first (optionally) fill in any gap between probe ends and second ligate the probe closed. In some embodiments, following circularization, any remaining (non-target) linear nucleic acid can be removed, resulting in isolation (capture) of target nucleic acid. Circularization selection-based preparative methods include molecular inversion probe capture reactions and 'selector' capture reactions. However, other techniques may be used as aspects of the invention are not limited in this respect. In some embodiments, molecular inversion probe capture of a target nucleic acid is indicative of the presence of a polymorphism in the target nucleic acid.

Petitioner Illumina, Inc., EX1004, p. 582

A variety of methods may be used to evaluate and compare bias profiles of each preparative technique. Next-generation sequencing may be used to quantitatively measure the abundance of each isolated target nucleic acid obtained from a certain preparative method. This abundance may be compared to a control abundance value (e.g., a known starting abundance of the target nucleic acid) and/or with an abundance determined through the use of an alternative preparative method. For example, a set of target nucleic acids may be isolated by one or more of the three preparative methods; the target nucleic acid may be observed x times using the amplification technique, y times using the hybridization enrichment technique, and z times using the circularization selection technique. A pairwise correlation coefficient may be computed between each abundance value (e.g., x and y, x and z, and y and z) to assess bias in nucleic acid isolation between pairs of preparative methods. Since the mechanisms of isolation are different in each approach, the abundances will usually be different and largely uncorrelated with each other.

In some embodiments, the invention provides a method of obtaining a nucleic acid preparation that is representative of a target nucleic acid in a biological sample by obtaining a first preparation of a target nucleic acid using a first preparative method on a biological sample, obtaining a second preparation of a target nucleic acid using a second preparative method on the biological sample, and combining the first and second nucleic acid preparations to obtain a combined preparation that is representative of the target nucleic acid in the biological sample.

In some embodiments of any of the methods described herein, a third preparation of the target nucleic acid is obtained using a third preparative method that is different from the first and second preparative methods, wherein the first, second, and third preparative methods all have different systematic sequence biases. In some embodiments of any of the methods described herein, the different preparative methods are used for a plurality of different loci in the biological sample to increase the sensitivity of a multiplex nucleic acid analysis. In some embodiments, the target nucleic acid has a sequence of a gene selected from Table 1.

However, it should be appreciated that a genotyping method of the invention may include several steps, each of which independently may involve one or more different preparative techniques described herein. In some embodiments, a nucleic acid preparation may be obtained using one or more (e.g., 2, 3, 4, 5, or more) different techniques described herein (e.g., amplification, hybridization capture, circular probe capture, etc., or any combination thereof) and

Petitioner Illumina, Inc., EX1004, p. 583

the nucleic acid preparation may be analyzed using one or more different techniques (e.g., amplification, hybridization capture, circular probe capture, etc., or any combination thereof) that are selected independently of the techniques used for the initial preparation.

In some embodiments, aspects of the invention also provide compositions, kits, devices, and analytical methods for increasing the sensitivity of nucleic acid assays. Aspects of the invention are particularly useful for increasing the confidence level of genotyping analyses. However, aspects of the invention may be used in the context of any suitable nucleic acid analysis, for example, but not limited to, a nucleic acid analysis that is designed to determine whether more than one sequence variant is present in a sample.

In some embodiments, aspects of the invention relate to a plurality of nucleic acid probes (e.g., 10-50, 50-100, 100-250, 250-500, 500-1,000, 1,000-2,000, 2,000-5,000, 5,000-7,500, 7,500-10,000, or lower, higher, or intermediate number of different probes).

In some embodiments, each probe or each of a subset of probes (e.g., 10-25%, 25-50%, 50-75%, 75-90%, or 90-99%) has a different first targeting arm. In some embodiments, each probe or each probe of a subset of probes (e.g., 10-25%, 25-50%, 50-75%, 75-90%, or 90-99%) has a different second targeting arm. In some embodiments, the first and second targeting arms are separated by the same intervening sequence. In some embodiments, the first and second targeting arms are complementary to target nucleic acid sequences that are separated by the same or a similar length (e.g., number of nucleic acids, for example, 0-25, 25-50, 50-100, 100-250, 250-500, 500-1,000, 1,000-2,500 or longer or intermediate number of nucleotides) on their respective target nucleic acids (e.g., genomic loci). In some embodiments, each probe or a subset of probes (e.g., 10-30 25%, 25-50%, 50-75%, 75-90%, or 90-99%) includes a first primer binding sequence. In some embodiments, the primer binding sequence is the same (e.g., it can be used to prime sequencing or other extension reaction). In some embodiments, each probe or a subset of probes (e.g., 10-25%, 25-50%, 50-75%, 75-90%, or 90-99%) includes a unique identifier sequence tag (e.g., that is predetermined and can be used to distinguish each probe).

In some embodiments, the methods disclosed herein are useful for any application where sensitivity is important. For example, detection of cancer mutations in a heterogenous tissue sample, detection of mutations in maternally-circulating fetal DNA, and detection of mutations in cells isolated during a preimplantation genetic diagnostic procedure.

Petitioner Illumina, Inc., EX1004, p. 584

According to some aspects of the invention, methods of detecting a polymorphism in a nucleic acid in a biological sample are provided. In some embodiments, the methods comprise obtaining a nucleic acid preparation using a preparative method (e.g., any of the preparative methods disclosed herein) on a biological sample, and performing a molecular inversion probe capture reaction on the nucleic acid preparation, wherein a molecular inversion probe capture (e.g., using a mutation-detection MIP) of a target nucleic acid of the nucleic acid preparation is indicative of the presence of a mutation (polymorphism) in the target nucleic acid, optionally wherein the polymorphism is selected from Table 2.

According to some aspects of the invention, methods of genotyping a nucleic acid in a biological sample are provided. In some embodiments, the methods comprise obtaining a nucleic acid preparation using a preparative method on a biological sample, sequencing a target nucleic acid of the nucleic acid preparation, and performing a molecular inversion probe capture reaction on the biological sample, wherein a molecular inversion probe capture of the target nucleic acid in the biological sample is indicative of the presence of a polymorphism in the target nucleic acid, genotyping the target nucleic acid based on the results of the sequencing and the capture reaction.

In some embodiments of the methods disclosed herein, the target nucleic acid has a sequence of a gene selected from Table 1.

It should be appreciated that any one or more embodiments described herein may be used for evaluating multiple genetic markers in parallel. Accordingly, in some embodiments, aspects of the invention relate to determining the presence of one or more markers (e.g., one or more alleles) at multiple different genetic loci in parallel.

Accordingly, the risk or presence of multiple heritable disorders may be evaluated in parallel. In some embodiments, the risk of having offspring with one or more heritable disorders may be evaluated. In some embodiments, an evaluation may be performed on a biological sample of a parent or a child (e.g., at a pre-implantation, prenatal, perinatal, or postnatal stage). In some embodiments, the disclosure provides methods for analyzing multiple genetic loci (e.g., a plurality of target nucleic acids selected from Table 1 or 2) from a patient sample, such as a blood, pre-implantation embryo, chorionic villus or amniotic fluid sample. A patient or subject may be a human. However, aspects of the invention are not limited to humans and may be applied to other species (e.g., mammals, birds, reptiles, other vertebrates or invertebrates) as

Petitioner Illumina, Inc., EX1004, p. 585

aspects of the invention are not limited in this respect. A subject or patient may be male or female. In some embodiments, in connection with reproductive genetic counseling, samples from a male and female member of a couple may be analyzed. In some embodiments, for example, in connection with an animal breeding program, samples from a plurality of male and female subjects may be analyzed to determine compatible or optimal breeding partners or strategies for particular traits or to avoid one or more diseases or conditions. Accordingly, reproductive risks may be determined and/or reproductive recommendations may be provided based on information derived from one or more embodiments of the invention.

However, it should be appreciated that aspects of the invention may be used in connection with any medical evaluation where the presence of one or more alleles at a genetic locus of interest is relevant to a medical determination (e.g., risk or detection of disease, disease prognosis, therapy selection, therapy monitoring, etc.). Further aspects of the invention may be used in connection with detection, in tumor tissue or circulating tumor cells, of mutations in cellular pathways that cause cancer or predict efficacy of treatment regimens, or with detection and identification of pathogenic organisms in the environment or a sample obtained from a subject, e.g., a human subject.

These and other aspects of the invention are described in more detail in the following description and non-limiting examples and drawings.

BRIEF DESCRIPTION OF DRAWINGS

FIG. 1 illustrates a non-limiting embodiment of a tiled probe layout;

FIG. 2 illustrates a non-limiting embodiment of a staggered probe layout;

FIG. 3 illustrates a non-limiting embodiment of an alternating staggered probe layout;

~~FIGs. 4A, B, and C~~ FIGS. 4A-C depict various non-limiting methods for combining differentiator tag sequence and target sequences (NNNN depicts a differentiator tag sequence);

FIG. 5 depicts a non-limiting method for genotyping based on target and 5 differentiator tag sequences;

FIG. 6 depicts non-limiting results of a simulation of a MIP capture reaction;

FIG. 7 depicts a non-limiting graph of sequencing coverage;

FIG. 8 illustrates that shorter sequences are captured with higher efficiency that longer sequences using MIPs;

Petitioner Illumina, Inc., EX1004, p. 586

FIG. 9 illustrates a non-limiting scheme of padlock (MIP) capture of a region that includes both repetitive regions (thick wavy line) and the adjacent unique sequence (thick strait line);

FIG. 10 illustrates a non-limiting hypothetical relationship between target gap size and the relative number of reads of the repetitive region;

FIG. 11A depicts MIP capture of FMR1 repeat regions from a diploid genome;

FIG. 11B depicts preparative methods for biallelic resolution of FMR I repeat region lengths in a diploid genome using MIP capture probes and unique differentiator tags;

FIG. 11C depicts an analysis of FMR1 repeat region lengths in a diploid genome;

FIG. 12 is a schematic of an embodiment of an algorithm of the invention;

FIG. 13 illustrates a non-limiting example of a graph of per-target abundance with MIP capture; and,

FIG. 14 shows a non-limiting a graph of correlation between two MIP capture reactions.

DETAILED DESCRIPTION

Aspects of the invention relate to preparative and analytical methods and compositions for evaluating genotypes, and in particular, for determining the allelic identity (or identities in a diploid organism) of one or more genetic loci in a subject.

Aspects of the invention are based, in part, on the identification of different sources of ambiguity and error in genetic analyses, and, in part, on the identification of one or more approaches to avoid, reduce, recognize, and/or resolve these errors and ambiguities at different stages in a genetic analysis. Aspects of the invention relate to methods and compositions for addressing bias and/or stochastic variation associated with one or more preparative and/or analytical steps of a nucleic acid evaluation technology. In some embodiments, preparative methods can be adapted to avoid or reduce the risk of bias skewing the results of a genetic analysis. In some embodiments, analytical methods can be adapted to recognize and correct for data variations that may give rise to misinterpretation (e.g., incorrect calls such as homozygous when the subject is actually heterozygous or heterozygous when the subject is actually homozygous). Methods of the invention may be used for any type of mutation, for example a single base change (e.g., insertion, deletion, transversion or transition, etc.), a multiple base insertion, deletion, duplication, inversion, and/or any other change or combination thereof.

Petitioner Illumina, Inc., EX1004, p. 587

In some embodiments, additional or alternative techniques may be used to address loci characterized by multiple repeats of a core sequence where the length of the repeat is longer than a typical sequencing read thereby making it difficult to determine whether a deletion or duplication of one or more core sequence units has occurred based solely on a sequence read.

In some embodiments, increased confidence in an assay result may be obtained by i) selecting two or more different preparative and/or analytical techniques that have different biases (e.g., known to have different biases), ii) evaluating a patient sample using the two or more different techniques, iii) comparing the results from the two or more different techniques, and/or iv) determining whether the results are consistent for the two or more different techniques. In some embodiments, if determining in step (iv) indicates that the results are consistent (e.g., the same) then increased confidence in the assay result is obtained. In other embodiments, if determining in step (iv) indicates that the results are inconsistent (e.g., that the results are ambiguous) then one or more additional preparative and/or analytical techniques, which have a different bias (e.g., known to have a different bias) compared with the two or more different preparative and/or analytical techniques selected in step (i), are used to evaluate the patient sample, and the results of the one or more additional preparative and/or analytical techniques are compared with the results from step (ii) to resolve the inconsistency.

In some embodiments, two or more independent samples may be obtained from a subject and independently analyzed. In some embodiments, two or more independent samples are obtained at approximately the same time point. In some embodiments, two or more independent samples are obtained at multiple different time points. In some embodiments, the use of two or more independent sample facilitates the elimination, normalization, and/or quantification of stochastic measurement noise. It is to be appreciated that two or more independent samples may be obtained in connection with any of the methods disclosed herein, including, for example, methods for pathogen profiling in a human or other animal subjects, monitoring tumor progression/regression, analyzing circulating tumor cells, analyzing fetal cells in maternal circulation, and analyzing/monitoring/profiling of environmental pathogens.

In some embodiments, one or more of the techniques described herein may be combined in a single assay protocol for evaluating multiple patient samples in parallel.

It should be appreciated that aspects of the invention may be useful for high throughput, cost-effective, yet reliable, genotyping of multiple patient samples (e.g., in parallel, for example

Petitioner Illumina, Inc., EX1004, p. 588

in multiplex reactions). In some embodiments, aspects of the invention are useful to reduce the error frequency in a multiplex analysis. Certain embodiments may be particularly useful where multiple reactions (e.g., multiple loci and/or multiple patient samples) are being processed. For example, 10-25, 25-50, 50-75, 75-100 or more loci may be evaluated for each subject out of any number of subject samples that may be processed in parallel (e.g., 1-25, 25-50, 50-100, 100-500, 500-1,000, 1,000-2,500, 2,500-5,000 or more or intermediate numbers of patient samples). It should be appreciated that different embodiments of the invention may involve conducting two or more target capture reactions and/or two or more patient sample analyses in parallel in a single multiplex reaction. For example, in some embodiments a plurality of capture reactions (e.g., using different capture probes for different target loci) may be performed in a single multiplex reaction on a single patient sample. In some embodiments, a plurality of captured nucleic acids from each one of a plurality of patient samples may be combined in a single multiplex analysis reaction. In some embodiments, samples from different subjects are tagged with subject-specific (e.g., patient-specific) tags (e.g., unique sequence tags) so that the information from each product can be assigned to an identified subject. In some embodiments, each of the different capture probes used for each patient sample have a common patient-specific tag. In some embodiments, the capture probes do not have patient-specific tags, but the captured products from each subject may be amplified using one or a pair of amplification primers that are labeled with a patient-specific tag. Other techniques for associating a patient-specific tag with the captured product from a single patient sample may be used as aspects of the invention are not limited in this respect. It should be appreciated that patient-specific tags as used herein may refer to unique tags that are assigned to identified patients in a particular assay. The same tags may be used in a separate multiplex analysis with a different set of patient samples (e.g., from different patients) each of which is assigned one of the tags. In some embodiments, different sets of unique tags may be used in sequential (e.g., alternating) multiplex reactions in order to reduce the risk of contamination from one assay to the next and allow contamination to be detected on the basis of the presence of tags that are not expected to be present in a particular assay.

Embodiments of the invention may be used for any of a number of different settings: reproductive settings, disease screening, identifying subjects having cancer, identifying subjects having increased risk for a disease, stratifying a population of subjects according to one or more of a number of factors, for example responsiveness to a particular drug, lack or not of an adverse

Petitioner Illumina, Inc., EX1004, p. 589

reaction (or risk therefore) to a particular drug, and/or providing information for medical records (e.g., homozygosity, heterozygosity at one or more loci). It should be appreciated that the invention is not limited to genomic analysis of patient samples. For example, aspects of the invention may be useful for high throughput genetic analysis of environment samples to detect pathogens.

In some embodiments, the methods disclosed herein are useful for diagnosis of one or more heritable disorders. In some embodiments, a heritable disorder that may be diagnosed with the methods disclosed herein is a genetic disorder that is prevalent in the Ashkenazi Jewish population. In some embodiments, the heritable disorders are selected from: 21-Hydroxylase-Defiocient Congenital Adrenal Hyperplasia; ABCC8-Related Hyperinsulinism; Alpha-Thalassemia, includes Constant Spring, & MR associated; Arylsulfatase A Deficiency-Metyachromatic Leukodystrophy; Biotinidase Deficiency Holocarboxylase Synthetase Deficiency; Bloom's Syndrome; Canavan Disease; CFTR-Related Disorders-cystic fibrosis; Citrullinemia Type 1; Combined MMA & Homocystinuria-dblC; Dystrophinopathies (DMD & BMD); Familial Dysautonomia; Fanconi Anemia-FANCC; Galactosemia-Classical: Galactokinase Defiency & Galactose Epimerase Deficiency; Gaucher Disease; GJB2-Related DFNB 1 Nonsyndromic Hearing Loss and Deafness; Glutaric acidemia Type 1; Hemoglobinopathies beta-chain disorders; Glycogen Storage Disease Type 1A; Maple Syrup Urine Disease; Types 1A, 1B, 2, 3; Medium Chain Acyl-Coenzyme A; Dehydrogenase Deficiency-MCADD; Methylmalonic Acidemia; Mucolipidosis IV; Nemaline Myopathy; Nieman-Pick Type A-Acid Sphingomyelinase Deficiency; Non-Ketotic Hyperglycinemia-Glycine Encephalopathy; Ornithine Transcarbamylase Deficiency; PKU Phenylalanine Hydroxylase Deficiency; Propionic Acidemia; Short Chain Acyl-CoA Dehydrogenase Deficiency-SCADD; Smith-Lemli-Opitz Syndrome; Spinal Muscular Atrophy (SMNI)-SMA; Tay Sachs-HexA Deficiency; Usher Synbdrome-Type I (Type IB, Type IC, Type ID, Type IF, Type IG) ; X-Linked Mental Retardation ARX-Related Disorders; X-Linked Mental Retardation with Cerebellar Cypoplasia and sistinctive Facial Appearance; X-Linked Mental Retardation; includes 9, 21, 30, 46, 58, 63, 88, 89; X linked mental retardation: FM1-Related Disorders-FRXA, Fragile X MR; X-linked SMR: Renpenning Syndrome 1; Zellweger Spectrum disorders - Peroxisomal Bifunctional Enzyme Deficiencies including Zellweger, NALD, and/or infantile

Petitioner Illumina, Inc., EX1004, p. 590

Refsums. However, all of these, subsets of these, other genes, or combinations thereof may be used.

According to some aspects, the disclosure relates to multiplex diagnostic methods. In some embodiments, multiplex diagnostic methods comprise capturing a plurality of genetic loci in parallel (e.g., a genetic locus of Table 1). In some embodiments, genetic loci possess one or more polymorphisms (e.g., a polymorphism of Table 2) the genotypes of which correspond to disease causing alleles. Accordingly, in some embodiments, the disclosure provides methods for assessing multiple heritable disorders in parallel.

In some embodiments, methods are provided for diagnosing multiple heritable disorders in parallel at a pre-implantation, prenatal, perinatal, or postnatal stage. In some embodiments, the disclosure provides methods for analyzing multiple genetic loci (e.g., a plurality of target nucleic acids selected from Table 1) from a patient sample, such as a blood, pre-implantation embryo, chorionic villus or amniotic fluid sample. A patient or subject may be a human. However, aspects of the invention are not limited to humans and may be applied to other species (e.g., mammals, birds, reptiles, other vertebrates or invertebrates) as aspects of the invention are not limited in this respect. A subject or patient may be male or female. In some embodiments, in connection with reproductive genetic counseling, samples from a male and female member of a couple may be analyzed. In some embodiments, for example, in connection with an animal breeding program, samples from a plurality of male and female subjects may be analyzed to determine compatible or optimal breeding partners or strategies for particular traits or to avoid one or more diseases or conditions.

However, it should be appreciated that any other diseases may be studied and/or risk factors for diseases or disorders including, but not limited to allergies, responsiveness to treatment, cancer tumor profiling for treatment and prognosis, monitoring and identification of patient infections, and monitoring of environmental pathogens.

1. Reducing representational bias in multiplex amplification reactions:

In some embodiments, aspects of the invention relate to methods that reduce bias and increase reproducibility in multiplex detection of genetic loci, e.g., for diagnostic purposes.

Molecular inversion probe technology is used to detect or amplify particular nucleic acid sequences in potentially complex mixtures. Use of molecular inversion probes has been demonstrated for detection of single nucleotide polymorphisms (Hardenbol et al. 2005 Genome

Petitioner Illumina, Inc., EX1004, p. 591

Res 15:269-75) and for preparative amplification of large sets of exons (Porreca et al. 2007 Nat Methods 4:931-6, Krishnakumar et al. 2008 Proc Natl Acad Sci USA 105:9296-301). One of the main benefits of the method is in its capacity for a high degree of multiplexing, because generally thousands of targets may be captured in a single reaction containing thousands of probes. However, challenges associated with, for example, amplification efficiency (See, e.g., Turner EH, et al., Nat Methods. 2009 Apr 6:1-2.) have limited the practical utility of the method in research and diagnostic settings.

Aspects of the disclosure are based, in part, on the discovery of effective methods for overcoming challenges associated with systematic errors (bias) in multiplex genomic capture and sequencing methods, namely high variability in target nucleic acid representation and unequal sampling of heterozygous alleles in pools of captured target nucleic acids (e.g., isolated from a biological sample). Accordingly, in some embodiments, the disclosure provides methods that reduce variability in the detection of target nucleic acids in multiplex capture methods. In other embodiments, methods improve allelic representation in a capture pool and, thus, improve variant detection outcomes. In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve the use of different sets of multiple probes (e.g., molecular inversion probes MIPs) that capture overlapping regions of a target nucleic acid to achieve a more uniform representation of the target nucleic acids in a capture pool compared with methods of the prior art. In other embodiments, methods reduce bias, or the risk of bias, associated with large scale parallel capture of genetic loci, e.g., for diagnostic purposes. In other embodiments, methods are provided for increasing reproducibility (e.g., by reducing the effect of polymorphisms on target nucleic acid capture) in the detection of a plurality of genetic loci in parallel. In further embodiments, methods are provided for reducing the effect of probe synthesis and/or probe amplification variability on the analysis of a plurality of genetic loci in parallel.

In some aspects, the disclosure provides probe sets that comprise a plurality of different probes. As used herein, a 'probe' is a nucleic acid having a central region flanked by a 5' region and a 3' region that are complementary to nucleic acids flanking the same strand of a target nucleic acid or subregion thereof. An exemplary probe is a molecular inversion probe (MIP). A 'target nucleic acid' may be a genetic locus.

Exemplary genetic loci are disclosed herein in Table 1 (RefSeqGene Column).

Petitioner Illumina, Inc., EX1004, p. 592

While probes have been typically designed to meet certain constraints (e.g. melting temperature, G/C content, etc.) known to partially affect capture/amplification efficiency (Ball et al (2009) Nat Biotech 27:361-8 AND Deng et al (2009) Nat Biotech 27:353-60), a set of constraints which is sufficient to ensure either largely uniform or highly reproducible capture/amplification efficiency has not previously been achieved.

As disclosed herein, uniformity and reproducibility can be increased by designing multiple probes per target, such that each base in the target is captured by more than one probe. In some embodiments, the disclosure provides multiple MIPs per target to be captured, where each MIP in a set designed for a given target nucleic acid has a central region and a 5' region and 3' region ('targeting arms') which hybridize to (at least partially) different nucleic acids in the target nucleic acid (immediately flanking a subregion of the target nucleic acid). Thus, differences in efficiency between different targeting arms and fill-in sequences may be averaged across multiple MIPs for a single target, which results in more uniform and reproducible capture efficiency.

In some embodiments, the methods involve designing a single probe for each target (a target can be as small as a single base or as large as a kilobase or more of contiguous sequence).

It may be preferable, in some cases, to design probes to capture molecules (e.g., target nucleic acids or subregions thereof) having lengths in the range of 1 - 200 bp (as used herein, a bp refers to a base pair on a double-stranded nucleic acid - however, where lengths are indicated in bps, it should be appreciated that single-stranded nucleic acids having the same number of bases, as opposed to base pairs, in length also are contemplated by the invention). However, probe design is not so limited. For example, probes can be designed to capture targets having lengths in the range of up to 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 1000, or more bps, in some cases.

It is to be appreciated that the length of a capture molecule (e.g., a target nucleic acid or subregion thereof) is selected based upon multiple considerations. For example, where analysis of a target involves sequencing, e.g., with a next-generation sequencer, the target length should typically match the sequencing read-length so that shotgun library construction is not necessary. However, it should be appreciated that captured nucleic acids may be sequenced using any suitable sequencing technique as aspects of the invention are not limited in this respect.

Petitioner Illumina, Inc., EX1004, p. 593

It is also to be appreciated that some target nucleic acids are too large to be captured with one probe. Consequently, it may be necessary to capture multiple subregions of a target nucleic acid in order to analyze the full target.

In some embodiments, a subregion of a target nucleic acid is at least 1 bp. In other embodiments, a subregion of a target nucleic acid is at least 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 bp or more. In other embodiments, a subregion of a target nucleic acid has a length that is up to 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or more percent of a target nucleic acid length.

The skilled artisan will also appreciate that consideration is made, in the design of MIPs, for the relationship between probe length and target length. In some embodiments, MIPs are designed such that they are several hundred basepairs (e.g., up to 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 bp or more) longer than corresponding target (e.g., subregion of a target nucleic acid, target nucleic acid).

In some embodiments, lengths of subregions of a target nucleic acid may differ.

For example, if a target nucleic acid contains regions for which probe hybridization is not possible or inefficient, it may be necessary to use probes that capture subregions of one or more different lengths in order to avoid hybridization with problematic nucleic acids and capture nucleic acids that encompass a complete target nucleic acid.

Aspects of the invention involve using multiple probes, e.g., MIPs, to amplify each target nucleic acid. In some embodiments, the set of probes for a given target can be designed to 'tile' across the target, capturing the target as a series of shorter sub targets. In some embodiments, where a set of probes for a given target is designed to 'tile' across the target, some probes in the set capture flanking non-target sequence). Alternately, the set can be designed to 'stagger' the exact positions of the hybridization regions flanking the target, capturing the full target (and in some cases capturing flanking non-target sequence) with multiple probes having different targeting arms, obviating the need for tiling. The particular approach chosen will depend on the nature of the target set. For example, if small regions are to be captured, a staggered-end approach might be appropriate, whereas if longer regions are desired, tiling might be chosen. In all cases, the amount of bias-tolerance for probes targeting pathological loci can be adjusted ('dialed in') by changing the number of different MIPs used to capture a given molecule.

Petitioner Illumina, Inc., EX1004, p. 594

In some embodiments, the 'coverage factor', or number of probes used to capture a basepair in a molecule, is an important parameter to specify. Different numbers of probes per target are indicated depending on whether one is using the tiling approach (see, e.g., FIG. 1) or one of the staggered approaches (see, e.g., FIGs. 2 or 3).

FIG. 1 illustrates a non-limiting embodiment of a tiled probe layout showing ten captured sub-targets tiled across a single target. Each position in the target is covered by three sub-targets such that MIP performance per base pair is averaged across three probes.

FIG. 2 illustrates a non-limiting embodiment of a staggered probe layout showing the targets captured by a set of three MIPs. Each MIP captures the full target, shown in black, plus (in some cases) additional extra-target sequence, shown in gray, such that the targeting arms of each MIP fall on different sequence. Each position in the target is covered by three sub-targets such that MIP performance per basepair is averaged across three probes. Targeting arms land immediately adjacent to the black or gray regions shown. It should be appreciated that in some embodiments, the targeting arms (not shown) can be designed so that they do not overlap with each other.

FIG. 3 illustrates a non-limiting embodiment of an alternating staggered probe layout showing the targets captured by a set of three MIPs. Each MIP captures the full target, shown in black, plus (in some cases) additional extra-target sequence, shown in gray, such that the targeting arms of each MIP fall on different sequence. Each position in the target is covered by three sub-targets such that MIP performance per basepair is averaged across three probes. Targeting arms land immediately adjacent to the black or gray regions shown.

It should be appreciated that for any of the layouts, the targeting arms on adjacent tiled or staggered probes may be designed to either overlap, not overlap, or overlap for only a subset of the probes.

In certain embodiments for any of the layouts, a coverage factor of about 3 to about 10 is used. However, the methods are not so limited and coverage factors of up to 2, 3, 4, 5, 6, 7, 8, 9, 10, 20 or more may be used. It is to be appreciated that the coverage factor selected may depend the probe layout being employed. For example, in the tiling approach, for a desired coverage factor, the number of probes per target is typically a function of target length, sub-target length, and spacing between adjacent sub-target start locations (step size). For example, for a desired coverage factor of 3, a 200 bp target with a start-site separation of 20 bp and sub-target length of

Petitioner Illumina, Inc., EX1004, p. 595

60 bp may be encompassed with 12 MIPs (FIG. 1). Thus, a specific coverage factor may be achieved by varying the number of probes per target nucleic acid and the length of the molecules captured. In the staggered approach, a fixed-length target nucleic acid is captured as several subregions or as 'super-targets', which are molecules comprising the target nucleic acid and additional flanking nucleic acids, which may be of varying lengths. For example, a target of 50 bp can be captured at a coverage factor of 3 with 3 probes in either a 'staggered' (FIG. 2) or 'alternating staggered' configuration (FIG. 3).

The coverage factor will be driven by the extent to which detection bias is tolerable. In some cases, where the bias tolerance is small, it may be desirable to target more subregions of target nucleic acid with, perhaps, higher coverage factors. In some embodiments, the coverage factor is up to 2, 3, 4, 5, 6, 7, 8, 9, 10 or more.

In some embodiments, when a tiled probe layout is used, when the target length is greater than 1 bp and when a step size (distance between the 5'-end of a target and the 5' end of its adjacent target) is less than the length of a target or subregion thereof, it is possible to compute probe number for a particular target based on target length (T), sub target length (S), and coverage factor (C), such that probe number = T/(S/C) + (C-1).

In some aspects, the disclosure provides methods to increase the uniformity of amplification efficiency when multiple molecules are amplified in parallel; methods to increase the reproducibility of amplification efficiency; methods to reduce the contribution of targeting probe variability to amplification efficiency; methods to reduce the effect on a given target nucleic acid of polymorphisms in probe hybridization regions; and/or methods to simplify downstream workflows when multiplex amplification by MIPs is used as a preparative step for analysis by nucleic acid sequencing.

Polymorphisms in the target nucleic acid under the regions flanking a target can interfere with hybridization, polymerase fill-in, and/or ligation. Furthermore, this may occur for only one allele, resulting in allelic drop-out, which ultimately decreases downstream sequencing accuracy. In some embodiments, using a set of MIPs having multiple hybridization sites for the capture of any given target, the probability of loss from polymorphism is substantially decreased because not all targeting arms in the set of MIPs will cover the location of the mutation.

Probes for MIP capture reactions may be synthesized on programmable microarrays because of the large number of sequences required. Because of the low synthesis yields of these

Petitioner Illumina, Inc., EX1004, p. 596

methods, a subsequent amplification step is required to produce sufficient probe for the MIP amplification reaction. The combination of multiplex oligonucleotide synthesis and pooled amplification results in uneven synthesis error rates and representational biases. By synthesizing multiple probes for each target, variation from these sources may be averaged out because not all probes for a given target will have the same error rates and biases.

Multiplex amplification strategies disclosed herein may be used analytically, as in detection of SNPs, or preparatively, often for next-generation sequencing or other sequencing techniques. In the preparative setting, the output of an amplification reaction is generally the input to a shotgun library protocol, which then becomes the input to the sequencing platform. The shotgun library is necessary in part because next-generation sequencing yields reads significantly shorter than amplicons such as exons. In addition to the bias-reduction afforded by the multi-tiled approach described here, tiling also obviates the need for shotgun library preparation. Since the length of the capture molecule can be specified when the probes, e.g., MIPs, are designed, it can be chosen to match the readlength of the sequencer. In this way, reads can 'walk' across an exon by virtue of the start position of each capture molecule in the probe set for that exon. Reducing analytical errors associated with bias in nucleic acid preparations:

In some embodiments, aspects of the invention relate to preparative steps in DNA sequencing-related technologies that reduce bias and increase the reliability and accuracy of downstream quantitative applications.

There are currently many genomics assays that utilize next-generation (polony-based) sequencing to generate data, including genome resequencing, RNA-seq for gene expression, bisulphite sequencing for methylation, and Immune-seq, among others. In order to make quantitative measurements (including genotype calling), these methods utilize the counts of sequencing reads of a given genomic locus as a proxy for the representation of that sequence in the original sample of nucleic acids. The majority of these techniques require a preparative step to construct a high-complexity library of DNA molecules that is representative of a sample of interest. This may include chemical or biochemical treatment of the DNA (e.g., bisulphite treatment), capture of a specific subset of the genome (e.g., padlock probe capture, solution hybridization), and a variety of amplification techniques (e.g., polymerase chain reaction, whole genome amplification, rolling circle amplification).

Petitioner Illumina, Inc., EX1004, p. 597

Systematic and random errors are common problems associated with genome amplification and sequencing library construction techniques. For example, genomic sequencing library may contain an over- or under-representation of particular sequences from a source genome as a result of errors (bias) in the library construction process. Such bias can be particularly problematic when it results in target sequences from a genome being absent or undetectable in the sequencing libraries. For example, an under representation of particular allelic sequences (e.g., heterozygotic alleles) from a genome in a sequencing library can result in an apparent homozygous representation in a sequencing library. As most downstream sequencing library quantification techniques depend on stochastic counting processes, these problems have typically been addressed by sampling enough (over-sampling) to obtain a minimum number of observations necessary to make statistically significant decisions. However, the strategy of oversampling is generally limited to elimination of low-count Poisson noise, and the approach wastes resources and increases the expense required to perform such experiments. Moreover, oversampling can result in a reduced statistical confidence in certain conclusions (e.g., diagnostic calls) based on the data. Accordingly, new approaches are needed for overcoming bias in sequencing library preparatory methods.

Aspects of the disclosure are based, in part, on the discovery of methods for overcoming problems associated with systematic and random errors (bias) in genome capture, amplification and sequencing methods, namely high variability in the capture and amplification of nucleic acids and disproportionate representation of heterozygous alleles in sequencing libraries. Accordingly, in some embodiments, the disclosure provides methods that reduce variability in the capture and amplification of nucleic acids. In other embodiments, the methods improve allelic representation in sequencing libraries and, thus, improve variant detection outcomes. In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve the use of differentiator tag sequences to uniquely tag individual nucleic acid molecules. In some embodiments, the differentiator tag sequence permits the detection of bias based on the frequency with which pairs of differentiator tag and target sequences are observed in a sequencing reaction. In other embodiments, the methods reduce errors caused by bias, or the risk of bias, associated with the capture, amplification and sequencing of genetic loci, e.g., for diagnostic purposes.

Petitioner Illumina, Inc., EX1004, p. 598

Aspects of the invention relate to associating unique sequence tags (referred to as differentiator tag sequences) with individual target molecules that are independently captured and/or analyzed (e.g., prior to amplification or other process that may introduce bias). These tags are useful to distinguish independent target molecules from each other thereby allowing an analysis to be based on a known number of individual target molecules. For example, if each of a plurality of target molecule sequences obtained in an assay is associated with a different differentiator tag, then the target sequences can be considered to be independent of each other and a genotype likelihood can be determined based on this information. In contrast, if each of the plurality of target molecule sequences obtained in the assay is associated with the same differentiator tag, then they probably all originated from the same target molecule due to over-representation (e.g., due to biased amplification) of this target molecule in the assay. This provides less information than the situation where each nucleic acid was associated with a different differentiator tag. In some embodiments, a threshold number of independently isolated molecules (e.g., unique combinations of differentiator tag and target sequences) is analyzed to determine the genotype of a subject.

In some embodiments, the invention relates to compositions comprising pools (libraries) of preparative nucleic acids that each comprise "differentiator tag sequences" for detecting and reducing the effects of bias, and for genotyping target nucleic acid sequences. As used herein, a "differentiator tag sequence" is a sequence of a nucleic acid (a preparative nucleic acid), which in the context of a plurality of different isolated nucleic acids, identifies a unique, independently isolated nucleic acid. Typically, differentiator tag sequences are used to identify the origin of a target nucleic acid at one or more stages of a nucleic acid preparative method. For example, in the context of a multiplex nucleic acid capture reaction, differentiator tag sequences provide a basis for differentiating between multiple independent, target nucleic acid capture events. Also, in the context of a multiplex nucleic acid amplification reaction, differentiator tag sequences provide a basis for differentiating between multiple independent, primary amplicons of a target nucleic acid, for example. Thus, combinations of target nucleic acid and differentiator tag sequence (target:differentiator tag sequences) of an isolated nucleic acid of a preparative method provide a basis for identifying unique, independently isolated target nucleic acids. FIG. 4A-C depict various non-limiting examples of methods for combining differentiator tag sequence and target sequences.

Petitioner Illumina, Inc., EX1004, p. 599

It will be apparent to the skilled artisan that differentiator tags may be synthesized using any one of a number of different methods known in the art. For example, differentiator tags may be synthesized by random nucleotide addition.

Differentiator tag sequences are typically of a predefined length, which is selected to control the likelihood of producing unique target:differentiator tag sequences in a preparative reaction (e.g., amplification-based reaction, a circularization selection-based reaction, e.g., a MIP reaction). Differentiator tag sequences may be, up to 5, up to 6, up to 7 up to 8, up to 9, up to 10, up to 11, up to 12, up to 13, up to 14, up to 15, up to 16, up to 17, up to 18, up to 19, up to 20, up to 21, up to 22, up to 23, up to 24, up to 25, or more nucleotides in length. For purposes of genotyping, isolated nucleic acids are identified as independently isolated if they comprise unique combinations of target nucleic acid and differentiator tag sequences, and observance of threshold numbers of unique combinations of target nucleic acid and differentiator tag sequences provide a certain statistical confidence in the genotype.

During a library preparation process, each nucleic acid molecule may be tagged with a unique differentiator tag sequence in a configuration that permits the differentiator tag sequence to be sequenced along with the target nucleic acid sequence of interest (the nucleic acid sequence for which the library is being prepared, e.g., a polymorphic sequence). The incorporation of the nucleic acid comprising a differentiator tag sequence at a particular step allows the detection and correction of biases in subsequent steps of the protocol.

A large library of unique differentiator tag sequences may be created by using degenerate, random-sequence polynucleotides of defined length. The differentiator tag sequences of the polynucleotides may be read at the final stage of the sequencing. The observations of the differentiator tag sequences may be used to detect and correct biases in the final sequencing read-out of the library. For example, the total possible number of differentiator tag sequences, which may be produced, e.g., randomly, is 4^N, where N is the length of the differentiator tag sequence. Thus, it is to be understood that the length of the differentiator tag sequence may be adjusted such that the size of the population of MIPs having unique differentiator tag sequences is sufficient to produce a library of MIP capture products in which identical independent combinations of target nucleic acid and differentiator tag sequence are rare. As used herein combinations of target nucleic acid and differentiator tag sequences, may also be referred to as "target:differentiator tag sequences".

Petitioner Illumina, Inc., EX1004, p. 600

In the final readout of a sequencing process, each read may have an additional unique differentiator tag sequence. In some embodiments, when differentiator tag sequences are distributed randomly in a library, all the unique differentiator tag sequences will be observed about an equal number of times. Accordingly, the number of occurrences of a differentiator tag sequence may follow a Poisson distribution.

In some embodiments, overrepresentation of target:differentiator tag sequences in a pool of preparative nucleic acids (e.g., amplified MIP capture products) is indicative of bias in the preparative process (e.g., bias in the amplification process). For example, target:differentiator tag sequence combinations that are statistically overrepresented are indicative of bias in the protocol at one or more steps between the incorporation of the differentiator tag sequences into MIPs and the actual sequencing of the MIP capture products.

The number of reads of a given target:differentiator tag sequence may be indicative (may serve as a proxy) of the amount of that target sequence present in the originating sample. In some embodiments, the numbers of occurrence of sequences in the originating sample is the quantity of interest. For example, using the methods disclosed herein, the occurrence of differentiator tag sequences in a pool of MIPs may be predetermined (e.g., may be the same for all differentiator tag sequences). Accordingly, changes in the occurrence of differentiator tag sequences after amplification and sequencing may be indicative of bias in the protocol. Bias may be corrected to provide an accurate representation of the composition of the original MIP pool, e.g., for diagnostic purposes.

According to some aspects, a library of preparative nucleic acid molecules (e.g., MIPs, each nucleic acid in the library having a unique differentiator tag sequence, may be constructed such that the number of nucleic acid molecules in the library is significantly larger than the number prospective target nucleic acid molecules to be captured using the library. This ensures that products of the preparative methods include only unique target:differentiator tag sequence; e.g., in a MIP reaction the capture step would undersample the total population of unique differentiator tag sequences in the MIP library. For example, an experiment utilizing 1 ug of genomic DNA will contain about ~150,000 copies of a diploid genome. For a MIP library, each MIP in the library comprising a randomly produced 12-mer differentiator tag sequence (~1.6 million possible unique differentiator tag sequences), there would be more than 100 unique

Petitioner Illumina, Inc., EX1004, p. 601

differentiator tag sequences per genomic copy. For a MIP library, each MIP in the library comprising a randomly produced 15-mer differentiator tag sequence (~1 billion possible unique differentiator tag sequences), there would be more than 7000 unique differentiator tag sequences per genomic copy. Therefore, the probability of the same differentiator tag sequence being incorporated multiple times is incredibly small. Thus, it is to be appreciated that the length of the differentiator tag sequence is to be selected based on the amount of target sequence in a MIP capture reaction and the desired probability for having multiple, independent occurrences of target:differentiator tag sequence combinations.

FIG. 5 depicts a non-limiting method for genotyping based on target and differentiator tag sequences. Sequencing reads of target and differentiator tags sequences are collapsed to make diploid genotype calls. FIG. 6 depicts non-limiting results of a simulation of a MIP capture reaction in which MIP probes, each having a differentiator tag sequence of 15 nucleotides, are combined with 10000 target sequence copies (e.g., genome equivalents). In this simulated reaction, the probability of capturing one or more copies of a target sequence having the same differentiator tag sequence is 0.05. The Y axis reflects the number of observations. The X axis reflects the number of independent occurrences of target:differentiator tag combinations. FIG. 7 depicts a non-limiting graph of sequencing coverage, which can help ensure that alleles are sampled to sufficient depth (e.g., either 10x or 20x minimum sampling per allele, assuming 1000 targets). In this non-limiting example, the X axis is total per-target coverage required, and the Y axis is the probability that a given total coverage will result in at least 10x or 20x coverage for each allele.

The skilled artisan will appreciate that as part of a MIP library preparation process, adapters may be ligated onto the ends of the molecules of interest. Adapters often contain PCR primer sites (for amplification or emulsion PCR) and/or sequencing primer sites. In addition, barcodes may be included, for example, to uniquely identify individual samples (e.g., patient samples) that may be mixed together. (See, e.g., USPTO Publication Number US 2007/0020640 Al (McCloskey et al.)

The actual incorporation of the random differentiator tag sequences can be performed through various methods known in the art. For example, nucleic acids comprising differentiator tag sequences may be incorporated by ligation. This is a flexible method, because molecules having differentiator tag sequence can be ligated to any blunt-ended nucleic acids. The

Petitioner Illumina, Inc., EX1004, p. 602

sequencing primers must be incorporated subsequently such that they sequence both the differentiator tag sequence and the target sequence. Alternatively, the sequencing adaptors can be synthesized with the random differentiator tag sequences at their 3' end (as degenerate bases), so that only one ligation must be performed. Another method is to incorporate the differentiator tag sequence into a PCR primer, such that the primer structure is arranged with the common adaptor sequence followed by the random differentiator tag sequence followed by the PCR priming sequence (in 5' to 3' order). A differentiator tag sequence and adaptor sequence (which may contain the sequencing primer site) are incorporated as tags. Another method to incorporate the differentiator tag sequences is to synthesize them into a padlock probe prior to performing a gene capture reaction. The differentiator tag sequence is incorporated 3' to the targeting arm but 5' to the amplification primer that will be used downstream in the protocol. Another method to incorporate the differentiator tag sequences is as a tag on a gene-specific or poly-dT reverse-transcription primer. This allows the differentiator tag sequence to be incorporated directly at the cDNA level.

In some embodiments, at the incorporation step, the distribution of differentiator tag sequences can be assumed to be uniform. In this case, bias in any part of the protocol would change the uniformity of this distribution, which can be observed after sequencing. This allows the differentiator tag sequence to be used in any preparative process where the ultimate output is sequencing of many molecules in parallel.

Differentiator tag sequences may be incorporated into probes (e.g., MIPs) of a plurality when they are synthesized on-chip in parallel, such that degeneracy of the incorporated nucleotides is sufficient to ensure near-uniform distribution in the plurality of probes. It is to be appreciated that amplification of a pool of unique differentiator tag sequences may itself introduce bias in the initial pool. However, in most practical cases, the scale of synthesis (e.g., by column synthesis, chip based synthesis, etc.) is large enough that amplification of an initial pool of differentiator tag sequences is not necessary. By avoiding amplification or selection steps on the pool of unique differentiator tag sequences, potential bias may be minimized.

One example of the use of the differentiator tag sequences is in genome re-sequencing. Considering that the raw accuracy of most next-generation sequencing instruments is relatively low, it is crucial to oversample the genomic loci of interest.

Petitioner Illumina, Inc., EX1004, p. 603

Furthermore, since there are two alleles at every locus, it is important to sample enough to ensure that both alleles have been observed a sufficient number of times to determine with a sufficient degree of statistical confidence whether the sample is homozygous or heterozygous. Indeed, the sequencing is performed to sample the composition of molecules in the originating sample. However, after multiple reads have been collected for a given locus, it is possible that due to bias (e.g., caused by PCR amplification steps), a large fraction of the reads are derived from a single originating molecule. This would skew the population of target sequences observed, and would affect the outcome of the genotype call. For example, it is possible that a locus that is heterozygous is called as homozygous, because there are only a few observations of the second allele out of many observations of that locus. However, if information is available on differentiator tag sequences, this situation could be averted, because the over-represented allele would be seen to also have an over-represented differentiator tag sequence (i.e., the sequences with the overrepresented differentiator tag sequence all originated from the same single molecule). Therefore, the sequences and corresponding distribution of differentiator tag sequences can be used as an additional input to the genotype-calling algorithm to significantly improve the accuracy and confidence of the genotype calls.

In some aspects, the disclosure provides methods for analyzing a plurality of target sequences which are genetic loci or portions of genetic loci (e.g., a genetic locus of Table 1). The genetic loci may be analyzed by sequencing to obtain a genotype at one or more polymorphisms (e.g., SNPs). Exemplary polymorphisms are disclosed in Table 2. The skilled artisan will appreciate that other polymorphisms are known in the art and may

Petitioner Illumina, Inc., EX1004, p. 604

be identified, for example, by querying the Entrez Single Nucleotide Polymorphism database, for example, by searching with a GeneID from Table 1.

Table 1: Target Nucleic Acids

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
CYP21A2	1589	cytochrome P450, family 21, subfamily A, polypeptide 2	CAH1; CPS1; CA21H; CYP21; CYP21B; P450c21B; MGC150536; MGC150537; CYP21A2	201910	NG_008337.1	6p21.3
ABCC8	6833	ATP-binding cassette, sub-family C (CFTR/MRP), member 8	HI; SUR; HHF1; MRP8; PHHI; SUR1; ABC36; HRINS; TNDM2; ABCC8	600509	NG_008867.1	11p15.1
ATRX	546	alpha thalassemia/mental retardation syndrome X-linked (RAD54 homolog, S. cerevisiae)	SHS; XH2; XNP; ATR2; SFM1; RAD54; MRXHF1; RAD54L; ZNF-HX; MGC2094; ATRX	300032	NG_008838.1	Xq13.1-q21.1
ARSA	410	arylsulfatase A	MLD; ARSA	607574	NG_009260.1	22q13.31-qter; 22q13.33
PSAP	5660	Prosaposin	GLBA; SAP1; FLJ00245; MGC10993; PSAP	176801	NG_008835.1	10q21-q22
BTD	686	Biotinidase	BTD	609019	NG_008819.1	3p25

5

Petitioner Illumina, Inc., EX1004, p. 605

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
HLCS	3141	holocarboxylase synthetase (biotin-(proprionyl-Coenzyme A-carboxylase (ATP-hydrolysing)) ligase)	HCS; HLCS	609018	NC_000021.7	21q22.1; 21q22.13
BLM	641	Bloom syndrome, RecQ helicase-like	BS; RECQ2; RECQL2; RECQL3; MGC126616; MGC131618; MGC131620; BLM	604610	NG_007272.1	15q26.1
ASPA	443	aspartoacylase (Canavan disease)	ASP; ACY2; ASPA	608034	NG_008399.1	17pter-p13
CFTR	1080	cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7)	CF; MRP7; ABC35; ABCC7; CFTR/MRP; TNR-CFTR; dJ760C5.1; CFTR	602421	NC_000007.12	7q31.2
ASS1	445	argininosuccinate synthetase 1	ASS; CTLN1; ASS1	603470	NG_011542.1	9q34.1
MMACHC	25974	methylmalonic aciduria (cobalamin deficiency) cblC type, with homocystinuria	cblC; FLJ25671; DKFZp564I122; RP11-291L19.3; MMACHC	609831	NC_000001.9	1p34.1
IKBKAP	8518	inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase complex-associated protein	FD; DYS; ELP1; IKAP; IKI3; TOT1; FLJ12497; DKFZp781H1425; IKBKAP	603722	NG_008788.1	9q31
FANCC	2176	Fanconi anemia, complementation group C	FA3; FAC; FACC; FLJ14675; FANCC	227645	NG_011707.1	9q22.3
GALK1	2584	galactokinase 1	GK1; GALK; GALK1	604313	NG_008079.1	17q24
GALT	2592	galactose-1-phosphate uridylyltransferase	GALT	606999	NC_000009.10	9p13
GALE	2582	UDP-galactose-4-epimerase	SDR1E1; FLJ95174; FLJ97302; GALE	606953	NG_007068.1	1p36-p35
GBA	2629	glucosidase, beta; acid (includes glucosylceramidase)	GCB; GBA1; GLUC; GBA	606463	NG_009783.1	1q21
GJB2	2706	gap junction protein, beta 2, 26kDa	HID; KID; PPK; CX26; DFNA3; DFNB1; NSRD1; DFNA3A; DFNB1A; GJB2	121011	NG_008358.1	13q11-q12
GCDH	2639	glutaryl-Coenzyme A dehydrogenase	GCD; ACAD5; GCDH	608801	NG_009292.1	19p13.2
G6PC	2538	glucose-6-phosphatase, catalytic subunit	G6PT; GSD1; GSD1a; MGC163350; G6PC	232200	NG_011808.1	17q21
HBB	3043	hemoglobin, beta	CD113t-C; beta-globin; HBB	141900	NG_000007.3	11p15.5

Petitioner Illumina, Inc., EX1004, p. 606

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
BCKDHA	593	branched chain keto acid dehydrogenase E1, alpha polypeptide	MSU; MSUD1; OVD1A; BCKDE1A; FLJ45695; BCKDHA	608348	NC_000019.8	19q13.1-q13.2
BCKDHB	594	branched chain keto acid dehydrogenase E1, beta polypeptide	E1B; FLJ17880; dJ279A18.1; BCKDHB	248611	NG_009775.1	6q13-q15
DBT	1629	dihydrolipoamide branched chain transacylase E2	E2; E2B; BCATE2; MGC9061; DBT	248610	NG_011852.1	1p31
DLD	1738	dihydrolipoamide dehydrogenase	E3; LAD; DLDH; GCSL; PHE3; DLD	238331	NG_008045.1	7q31-q32
ACADM	34	acyl-Coenzyme A dehydrogenase, C-4 to C-12 straight chain	MCAD; ACAD1; MCADH; FLJ18227; FLJ93013; FLJ99884; ACADM	607008	NG_007045.1	1p31
MMAA	166785	methylmalonic aciduria (cobalamin deficiency) cblA type	cblA; MGC120010; MGC120011; MGC120012; MGC120013; MMAA	607481	NG_007536.1	4q31.22
MMAB	326625	methylmalonic aciduria (cobalamin deficiency) cblB type	ATR; cblB; MGC20496; MMAB	607568	NG_007096.1	12q24
MUT	4594	methylmalonyl Coenzyme A mutase	MCM; MUT	609058	NG_007100.1	6p12.3
MCOLN1	57192	mucolipin 1	ML4; MLIV; MST080; TRPML1; MSTP080; TRP-ML1; TRPM-L1; MCOLN1	605248	NC_000019.8	19p13.3-p13.2
ACTA1	58	actin, alpha 1, skeletal muscle	ACTA; ASMA; CFTD; MPFD; NEM1; NEM2; NEM3; CFTD1; CFTDM; ACTA1	102610	NG_006672.1	1q42.13
TPM3	7170	tropomyosin 3	TM3; TRK; NEM1; TM-5; TM30; TM30nm; TPMsk3; hscp30; MGC3261; FLJ41118; MGC14582; MGC72094; OK/SW-cl.5; TPM3	191030	NG_008621.1	1q21.2
TNNT1	7138	troponin T type 1 (skeletal, slow)	ANM; TNT; STNT; TNTS; FLJ98147; MGC104241; TNNT1	191041	NG_011829.1	19q13.4
NEB	4703	nebulin	NEM2; NEB177D; FLJ11305; FLJ36536; FLJ39568; FLJ39584; DKFZp686C1456; NEB	161650	NG_009382.1	2q22
SMPD1	6609	sphingomyelin phosphodiesterase 1, acid lysosomal	ASM; NPD; SMPD1	607608	NG_011780.1	11p15.4-p15.1
GLDC	2731	glycine dehydrogenase (decarboxylating)	GCE; NKH; GCSP; HYGN1; MGC138198; MGC138200; GLDC	238300	NC_000009.10	9p22

Petitioner Illumina, Inc., EX1004, p. 607

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
GCSH	2653	glycine cleavage system protein H (aminomethyl carrier)	GCE; NKH; GCSH	238330	NC_000016.8	16q23.2
AMT	275	aminomethyltransferase	GCE; NKH; GCST; AMT	238310	NC_000003.10	3p21.2-p21.1
OTC	5009	ornithine carbamoyltransferase	OCTD; MGC129967; MGC129968; MGC138856; OTC	300461	NG_008471.1	Xp21.1
PAH	5053	phenylalanine hydroxylase	PH; PKU; PKU1; PAH	612349	NG_008690.1	12q22-q24.2
DHPR	5860	quinoid dihydropteridine reductase	DHPR; PKU2; SDR33C1; FLJ42391; QDPR	612676	NG_008763.1	4p15.31
PTS	5805	6-pyruvoyltetrahydropterin synthase	PTPS; FLJ97081; PTS	261640	NG_008743.1	11q22.3-q23.3
PCCA	5095	propionyl Coenzyme A carboxylase, alpha polypeptide	PCCA	232000	NG_008768.1	13q32
PCCB	5096	propionyl Coenzyme A carboxylase, beta polypeptide	DKFZp451E113; PCCB	232050	NG_008939.1	3q21-q22
ACADS	35	acyl-Coenzyme A dehydrogenase, C-2 to C-3 short chain	SCAD; ACAD3; ACADS	606885	NG_007991.1	12q22-qter
DHCR7	1717	7-dehydrocholesterol reductase	SLOS; DHCR7	602858	NC_000011.8	11q13.2-q13.5
SMN1	6606	survival of motor neuron 1, telomeric	SMA; SMN; SMA1; SMA2; SMA3; SMA4; SMA@; SMNT; BCD541; T-BCD541; SMN1	600354	NG_008691.1	5q13
HEXA	3073	hexosaminidase A (alpha polypeptide)	TSD; MGC99608; HEXA	606869	NG_009017.1	15q23-q24
MYO7A	4647	myosin VIIA	DFNB2; MYU7A; NSRD2; USH1B; DFNA11; MYOVIIA; MYO7A	276903	NG_009086.1	11q13.5
USH1C	10083	Usher syndrome 1C (autosomal recessive, severe)	PDZ73; AIE-75; DFNB18; PDZ-45; PDZ-73; NY-CO-37; NY-CO-38; ush1cpst; PDZ-73/NY-CO-38; USH1C	605242	NC_000011.8	11p15.1-p14
CDH23	64072	cadherin-like 23	USH1D; DFNB12; FLJ00233; FLJ36499; KIAA1774; KIAA1812; MGC102761; DKFZp434P2350; CDH23	605516	NG_008835.1	10q21-q22
PCDH15	65217	protocadherin 15	USH1F; DFNB23; DKFZp667A1711; PCDH15	605514	NG_009191.1	10q21.1

Petitioner Illumina, Inc., EX1004, p. 608

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
SANS	124590	Usher syndrome 1G (autosomal recessive)	SANS; ANKS4A; FLJ33924; USH1G	607696	NG_007882.1	17q25.1
ARX	170302	aristaless related homeobox	ISSX; PRTS; MRX29; MRX32; MRX33; MRX36; MRX38; MRX43; MRX54; MRX76; MRX87; MRXS1; ARX	300382	NG_008281.1	Xp21
OPHN1	4983	oligophrenin 1	OPN1; MRX60; OPHN1	300127	NG_008960.1	Xq12
JARID1C	8242	lysine (K)-specific demethylase 5C	MRXJ; SMCX; MRXSJ; XE169; JARID1C; DXS1272E; KDM5C	314690	NG_008085.1	Xp11.22-p11.21
FTSJ1	24140	FtsJ homolog 1 (E. coli)	JM23; MRX9; SPB1; TRM7; CDLIV; MRX44; FTSJ1	300499	NG_008879.1	Xp11.23
SLC6A8	6535	solute carrier family 6 (neurotransmitter transporter, creatine), member 8	CRT; CT1; CRTR; MGC87396; SLC6A8	300036	NC_000023.9	Xq28
DLG3	1741	discs, large homolog 3 (Drosophila)	MRX; MRX90; NEDLG; NE-Dlg; SAP102; SAP-102; KIAA1232; DLG3	300189	NC_000023.9	Xq13.1
TM4SF2	7102	tetraspanin 7	AI5; MXS1; CD231; MRX58; CCG-B7; TM4SF2; TALLA-1; TM4SF2b; DXS1692E; TSPAN7	300096	NG_009160.1	Xp11.4
ZNF41	7592	zinc finger protein 41	MRX89; MGC8941; ZNF41	314995	NG_008238.1	Xp11.23
FACL4	2182	acyl-CoA synthetase long-chain family member 4	ACS4; FACL4; LACS4; MRX63; MRX68; ACSL4	300157	NG_008053.1	Xq22.3-q23
PQBP1	10084	polyglutamine binding protein 1	SHS; MRX55; MRXS3; MRXS8; NPW38; RENS1; PQBP1	300463	NC_000023.9	Xp11.23
PEX1	5189	peroxisomal biogenesis factor 1	ZWS1; PEX1	602136	NG_008341.1	7q21.2
PXMP3	5828	peroxisomal membrane protein 3, 35kDa	PAF1; PEX2; PMP3; PAF-1; PMP35; RNF72; PXMP3	170993	NG_008371.1	8q21.1
PEX6	5190	peroxisomal biogenesis factor 6	PAF2; PAF-2; PXAAA1; PEX6	601498	NG_008370.1	6p21.1
PEX10	5192	peroxisomal biogenesis factor 10	NALD; RNF69; MGC1998; PEX10	602859	NG_008342.1	1p36.32
PEX12	5193	peroxisomal biogenesis factor 12	PAF-3; PEX12	601758	NG_008447.1	17q12
PEX5	5830	peroxisomal biogenesis factor 5	PXR1; PTS1R; PTS1-BP; FLJ50634; FLJ50721; FLJ51948; PEX5	600414	NG_008448.1	12p13.31
PEX26	55670	peroxisomal biogenesis factor 26	FLJ20695; PEX26M1T; Pex26pM1T; PEX26	608666	NG_008339.1	22q11.21

Petitioner Illumina, Inc., EX1004, p. 609

The mutations listed in Table 2 are documented polymorphisms in several disease-associated genes (CFTR is mutated in cystic fibrosis, GBA is mutated in Gaucher disease, ASPA is mutated in Canavan disease, HEXA is mutated in Tay Sachs disease). The polymorphisms are of several types: insertion/deletion polymorphisms which will cause frameshifts (and thus generally interrupt protein function) unless the insertion/deletion length is a multiple of 3bp, and substitutions which can alter the amino acid sequence of the protein and in some cases cause complete inactivation by introduction of a stop codon.

Table 2: Non-limiting examples of polymorphisms

Gene name	GeneID	SNP ID	Mutation	SEQ ID NO:
CFTR	1080	rs63500661	TCACATCACCAAGTTAAAAAAAAAAA[A/G]G GGGCGGGGGGGCAGAATGAAAATT	1
CFTR	1080	rs63107760	AAACAAGGATGAATTAAGTTTTTTTT[-/T]AAAAAAGAAACATTTGGTAAGGGGA	2
CFTR	1080	rs62469443	ATCACCAAGTTAAAAAAAAAAAAGGG[A/G]C GGGGGGGCAGAATGAAAATTGCAT	3
CFTR	1080	rs62469442	CTATTGAACCAGAACCAAACAGGAAT[A/G]C CATAGCATTTTGTAAACTAAACTG	4
CFTR	1080	rs62469441	CAGGAGTTCAAGACCAGCCTACTAAA[A/C]C ACACACACACACACACACACACAC	5
CFTR	1080	rs62469439	GATTAAATAATAGTGTTTATGTACCC[C/G]GC TTATAGGAGAAGAGGGTGTGTGT	6
CFTR	1080	rs62469438	ATTGTTATCTTTTCATATAAGGTAAC[A/T]GA GGCCCAGAGAGATTAAATAACAT	7
CFTR	1080	rs62469437	TAATTTTAATTAAGTAAATTTAATTG[A/G]TA GATAAATAAGTAGATAAAAAATA	8
CFTR	1080	rs62469436	GTATAAAAAAAAAAAAAAAAAAAGTT[A/T]G AATGTTTTCTTGCATTCAGAGCCT	9
CFTR	1080	rs62469435	ATACTAAAAATTTAAAGTTCTCTTGC[A/G]AT ATATTTTCTTAATATCTTACATC	10
CFTR	1080	rs62469434	TGCTGGGATTACAGGCGTGAGCCACC[A/G]C GCCTGGCCTGATGGGACATATTTT	11
CFTR	1080	rs62469433	CTACAATATAAGTATAGTATTGCAAA[A/C]CC ATCAGGAAGGGTGTTAACTATTT	12
CFTR	1080	rs61763210	GTTGTCTCCAAACTTTTTTTCAGGTG[-/AGA]AGGTGGCCAACCGAGCTTCGGAAAG	13
CFTR	1080	rs61720488	TTTTTTCATAAAAGATTATATAAAGG[A/C]TA TTGCTTTTGAATCACAAACACTA	14
CFTR	1080	rs61481156	ATCTAGTGAGCAGTCAGGAAAGAGAA[C/T]T TCCAGATCCTGGAAATCAGGGTTA	15
CFTR	1080	rs61443875	TAGAGTATAAAAAAAAAAAAAAAAA[-/A]GTTTGAATGTTTTCTTGCATTCAGA	16

Petitioner Illumina, Inc., EX1004, p. 610

CFTR	1080	rs61312222	TGCAAATGCCAACTATCAAAGATATT[C/G]GA GTATACTGTCAATAAACTTCATA	17
CFTR	1080	rs61159372	TCCTCAACAGTTAGAAACAATATTTT[C/G]AG TGATTTCCCATGCCAACTTTACT	18
CFTR	1080	rs61094145	TTTTTGGTATTGTTGTTAAATAAGTG[A/G]GA ATTCAATACAGTATAATGTCTGT	19
CFTR	1080	rs61086387	CTTGAAATCGGATATATATATATATATA[- /TGTATATATATATATATATATATATATAT ACATATATATATATA]GTATTATCCCTGTTTTC ACAGTTTT	20
CFTR	1080	rs60996744	AGAGGGGCTGTGAAGGACACCAAGGA[A/G]G AGACTAAGAGCCAGGAGGGAAAAC	21
CFTR	1080	rs60960860	TAGAGTTTATTAGCTTTTACTACTCT[A/G]CTT AGTTACTTTGTGTTACAGAATA	22
CFTR	1080	rs60923902	ACTAGTGATGATGAGCTTCTTTTCAT[- /AT]GTTTGTTGGCTGCATAAATGTCTTC	23
CFTR	1080	rs60912824	GCAGAGAAAAGAGGGGCTGTGAAGGA[C/G]A CCAAGGAGGAGACTAAGAGCCAGG	24
CFTR	1080	rs60887846	TTCAGAGGTCTACCACTGGTGCATAC[G/T]CT AATCACAGTGTCGAAAATTTTAC	25
CFTR	1080	rs60793174	AAGAAAGAGCAAAAGAGGGCAAACTT[C/T]T CATACATTTTTGATGTCGAAACCA	26
CFTR	1080	rs60788575	CCTAAAGTTTAAAAAGAAAAAAAAAA[- /A]GGAAGAAGGAATTAAAAATCCAAAG	27
CFTR	1080	rs60760741	GTGTGTGTGTGTATATATATATATATAT[A/T]TA TATATTTTTTTTTTCCTGAGCCA	28
CFTR	1080	rs60456599	AAACTGTTGATGTTTTCATTTATTTA[C/G]ATC ATTGGAAAACTTTAGATTCTAG	29
CFTR	1080	rs60363249	TTTATCCATTCTTAACCAGAACAGAC[A/G]TT TTTTCAGAGCTGGTCCAGGAAAA	30
CFTR	1080	rs60355115	TTGAAATCGGATATATATATATATATAT[A/G]TA TATATATATATATATATATATAT	31
CFTR	1080	rs60308689	TAGTTTTTTATTTCCTCATATTATTT[- /T]CAGTGGCTTTTTCTTCCACATCTTT	32
CFTR	1080	rs60271242	ACATAGTTCTCAGTGGTACAACTACA[A/G]GT GATTTCTCTTTTCTTATTTCTGG	33
CFTR	1080	rs60010318	AGAGCAATGGCATCCCTTGTCTTGTG[C/T]TA TACAGGATGCAGCAATTTATAGG	34
CFTR	1080	rs59961323	TTCTGTCTACATAAGATGTCATACTA[A/G]AT TATCTTTTCCAGCATGCATTCAG	35
CFTR	1080	rs59961270	CAGGGTGGCATGTTAGGCAGTGCTTA[A/G]A ATAAATGAGTTGGTTATACAAGTA	36
CFTR	1080	rs59837506	AGGACACACACACACACACACACACA[- /CA]TGCACACACATTTAAATAGATGCAT	37
CFTR	1080	rs59572090	TAAAAAATTGGTATAATGAAATTGCA[C/T]TT GTAGTCTTTGGACATTTAAATCC	38
CFTR	1080	rs59548252	TTTCAATACTTAAGAGGTACGCAGAG[A/G]A AAGAGGGGCTGTGAAGGACACCAA	39
CFTR	1080	rs59519859	CAGCAATGAATATTTTGAGGCTGAGG[C/T]GC TGAGGGGTAAAATTGCAGCCTGG	40
CFTR	1080	rs59509837	TTATGGTTTATATTTTTGTGTCTTCT[- /CTTT]AACACATCTTTTCTAGCAGAATTCA	41

Petitioner Illumina, Inc., EX1004, p. 611

CFTR	1080	rs59417037	GTATTTTAGTTTTTTTTTTTTGTTTG[-/T]TTTGTTTTGTTTTGTTTTGTTTTTG	42
CFTR	1080	rs59159458	TGGGTGACTCCATTTTTACTTTTAGT[C/T]TGG TCTGTTGAGGCCTCGTGAGAGA	43
CFTR	1080	rs59048119	TATTTTCATGTATTTTAGTTTTTTTTT[-/TTTT]GTTTGTTTTGTTTTGTTTTGTTTTG	44
CFTR	1080	rs58970500	GTGTGTGTGTATATATATATATATAT[A/T]TA TATTTTTTTTTTCCTGAGCCAAA	45
CFTR	1080	rs58942292	AACCTATTAGCATGTCTGGCAGAAAA[-/A]TAGATACTTAATAAATTTCTTAAAT	46
CFTR	1080	rs58917054	GAGGCTTAGACAGTTTAAGTAACTCA[A/G]G CATGGTTACACAACTAGCTAGGGC	47
CFTR	1080	rs58837484	GTGTGAGTATTATGAGACCATATGTT[A/G]GG AGATTTTATTTGGTATTGAGGAT	48
CFTR	1080	rs58829491	GAAACCCCACCCCTTCTATAGTTTTC[C/T]CTT TAATATTTACAATGGAACCATT	49
CFTR	1080	rs58805195	CATATATATATAGTGTGTGTGTGTGT[A/G]TA TATATATATATATATATATTTTT	50
GBA	2629	rs60866785	CGAGCGAGAGAGAGAGAGAGAGAGAG[-/AG]GAGCCGGCGCGAGAACTACGCATGC	51
GBA	2629	rs60239603	GGCAGGTAATATCTAGTACCTTACTT[A/T]TA TTTCCTGAGCACATTCTACATTT	52
GBA	2629	rs56310840	GGCCAGGAATGGGAGTGCTTAGGTGC[A/G]G AGGTGGCACTGTTCCCGCAGCTGC	53
GBA	2629	rs41264927	TGAAAACTCCATCCCCTCAGGGTCAT[C/T]AG ATGAAGAGAAGACCACAGGGGTT	54
GBA	2629	rs41264925	TGTAGGTAAGGGTCACATGTGGGAGA[C/G]G CAGCTGTGGGTAGGTCAGCCCTGT	55
GBA	2629	rs36024691	CCAAGAAGGCGCCATTACACTCCAGC[-/C]TGGGCGACAGGGCGAGACTCCCTCA	56
GBA	2629	rs36024092	TGCCACACCCAGCTAATTTGTGTGTG[-/G]TATGTGTGTGTATGTATGTGTGTGT	57
GBA	2629	rs35682967	GTTCCTCCAGTAATTTTTTTTTTTTT[-/T]CGTTTTGAGACAGAGTCTTGCCCTG	58
GBA	2629	rs35033592	ATCATGCCCAGATAATTTTTTTTTTT[-/T]GTATTTTAGTAGACACAGGGTTTCA	59
GBA	2629	rs34732744	CGAGCGAGAGAGAGAGAGAGAGAGAG[-/AG]GAGCCGGCGCGAGAACTACGCATGC	60
GBA	2629	rs34620635	CCTGTGAGGGGCACATTCCTTAGTAG[-/C]TAAGGAGTTGGGGGTGTGAAGATCC	61
GBA	2629	rs34302637	ACAGGCTACTGGCTGGGCCCAGGCAA[-/A]GGGGGCCTTGGCAGGAAAAGTTCCT	62
GBA	2629	rs33949225	GCGAGAGAGAGAGAGAGAGAGAGAGG[-/AG]AGCCGGCGCGAGAACTACGCATGCG	63
GBA	2629	rs28678003	AAGAAGAAAAATAAAAAGAAAGTGGG[C/T]C AGACCGAGAGAACAGGAAGCCTGA	64
GBA	2629	rs28559737	AAGGACAAAGGCAAAGAGACAAAGGC[G/T]C AACACTGGGGGTCCCCAGAGAGTG	65
GBA	2629	rs28373017	TACCTAGTCACTTCCTGCCTCCATGG[C/T]GC AAAAGGGGATGGGTGTGCCTCTT	66
GBA	2629	rs12752133	CTCTTCCGAGGTTCCACCCTGAACAC[C/T]TT CCTGCTCCCTCGTGGTGTAGAGT	67

Petitioner Illumina, Inc., EX1004, p. 612

GBA	2629	rs12747811	TTCTGACTGGCAACCAGCCCCACTCT[C/T]TG GGAGCCCTCAGGAATGAACTTGC	68
GBA	2629	rs12743554	gctcagcctcccaggctggagtgcag[A/T]ggcgcgatctcggctcac cgcaacc	69
GBA	2629	rs12041778	CATGAACCACATCAAATGAGATTTAG[C/T]GG GAGTGGCACACACAGTCATGACC	70
GBA	2629	rs12034326	AAGCAGCCCTGGGGAGTCGGGGCGGG[A/G]C CTGGATTGGAAAAGAGACGGTCAC	71
GBA	2629	rs11558184	CTCCAAGTTCTGGGAGCAGAGTGTGC[A/G]G CTAGGCTCCTGGGATCGAGGGATG	72
GBA	2629	rs11430678	GTTCCTCCAGTAAttttttttttt[- /G/T]gtttgagacagagtcttgccctgt	73
GBA	2629	rs11264345	CTAGTACCTTACTTCCCTCAAGTTCA[A/T]TC ATCTCACAGATATTTCCTGAGCA	74
GBA	2629	rs10908459	aattagccgtgcgtggtggcgggtgc[C/T]gtaatcccacgtacttgg gaggct	75
GBA	2629	rs10796940	CCATGGCCAGCCGGGGAGGGGACGGG[A/C]A CACACAGACCCACACAGAGACTCA	76
GBA	2629	rs10668496	agcgagagagagagagagagagagag[- /AG]gagCCGGCGCGAGAACTACGCATGC	77
GBA	2629	rs7416991	CGTAGCAGTTAGCAGATGATAGGCGG[C/G/T] GAAATCTTATTTCACAGGGCATTAA	78
GBA	2629	rs4024049	CTGGCCCTGGTGACAGTGGGGCTGTG[C/T]GT GGGGCCAGAGCCTTCTCAGAGGT	79
GBA	2629	rs4024048	CAGATACTGGCCCTGGTGACAGTGGG[A/G]C TGTGCGTGGGGCCAGAGCCTTCTC	80
GBA	2629	rs4024047	GACAGATACTGGCCCTGGTGACAGTG[G/T]G GCTGTGCGTGGGGCCAGAGCCTTC	81
GBA	2629	rs3841430	GGCTCctctctctctctctctctc[- /TC]gctcgctctctcgctctctcgctct	82
GBA	2629	rs3754485	GTTTCAGACCAGCCTGGCCAACATAG[C/T]GA AACCCCATCTCTACTAAAAATAA	83
GBA	2629	rs3205619	AGTGGGCGATTGGATGGAGCTGAGTA[C/T]G GGGCCCATCCAGGCTAATCACACC	84
GBA	2629	rs2990227	CCGGGCTCCGTGAATGTTTGTCACAT[C/G]TC TGAAGAACGTATGAATTACATAA	85
GBA	2629	rs2990226	GAATCCCAACCCCGACGCTCGTCGCC[C/G]G GCTCCGTGAATGTTTGTCACATGT	86
GBA	2629	rs2990225	GCGAATCCCAACCCCGACGCTCGTCG[C/T]CG GGCTCCGTGAATGTTTGTCACAT	87
GBA	2629	rs2990224	TGGGCAGAAGTCAGGGTCCAAAGAAA[G/T]G GCAAAGAAAAGTGTcagtggctca	88
ASPA	443	rs63751297	TAAGAAAGACGTTTTTGATTTTTTTC[A/G]GA CTTCTCTGGCTCCACTACCCTGC	89
ASPA	443	rs62071301	CTGATTCCTGGCCAGGAGCGGTGGCT[C/T]AC GCCTGTAATCCCAGCGCTTTGGG	90
ASPA	443	rs62071300	TAAAAATGCTGATTCCTGGCCAGGAG[C/T]GG TGGCTCACGCCTGTAATCCCAGC	91
ASPA	443	rs62071299	TTTAAAAATGCTGATTCCTGGCCAGG[A/C]GC GGTGGCTCACGCCTGTAATCCCA	92
ASPA	443	rs62071297	CAAGACCTGTCAAAGATCTGAGAAAT[A/T]TT ACCCGACTTACAAGCTAACCATT	93

Petitioner Illumina, Inc., EX1004, p. 613

ASPA	443	rs61697033	ACTGTAATAAGTGCTGTAAAAGAAAT[A/G]C ACAAAATAATATAGCAGAGGGTAT	94
ASPA	443	rs60743592	CTTGAGGTCAGGAGTTCAAGACCAGT[C/T]TG GGCAACATGGGGAAAACCTTGTC	95
ASPA	443	rs60666840	AGGTTGCAGTGAGCCGAGATCATGCC[A/G]TT GCACTCCAGCCGGGGCAACAAAA	96
ASPA	443	rs60147514	ACAAGTGTCTTGAAATTATCTGTGAT[C/T]TG CTATAGAGCAATACTTTTGTAAA	97
ASPA	443	rs59930743	GTGGGTATATGCAGCTCTATGCACTA[C/T]CT GCTCATTTATTTGGTAAATCTAA	98
ASPA	443	rs59690349	TGTGTGTGTGTGCGTGTGTGTGTGTG[-/TGTGTGTG]ATCATAAGAGTGGCTGCAGCAA ACT	99
ASPA	443	rs59676360	AGTCTGGAGTGCAATGGTGCAATCTC[A/G]GC TCACTGCAGCCTCCACCTCCGGG	100
ASPA	443	rs59335404	CTCCTAATGGATATTTCCTAAATTTT[G/T]CTG AACAGAATTTAACTTGAGCTGG	101
ASPA	443	rs58879097	ATTTAAAAATGGATTTCTAGAAAAAC[A/G]AT CACATACTTGAATATTTTAGCAA	102
ASPA	443	rs58686774	CTATAAATGGGTAGCATGAGGGATTC[A/G]A GGAGGTGGCTGAAAGAAGCACGTA	103
ASPA	443	rs57511162	AAGAAACCAAGCATAGTAGAGTGTTA[A/G]A AAACCAAAGCAACTAAACAACTGT	104
ASPA	443	rs55859596	CGGGGCTCAGAACTTGTAACAGAAAA[A/T]T AAAATATACTCCACTCAAGGGAAT	105
ASPA	443	rs55742972	TACTACACTTCACGGATACTGTACTT[-/GTACTT]TTTTTCCAAATTGAAGGTTTTTGGC	106
ASPA	443	rs55640436	TTGTTTTTGTTTTTGTTTTTGTTTTT[-/GTTTTTGTTTTT]TGAGATGGAGTCTCGCTCT GTCGCC	107
ASPA	443	rs36225687	TTTGCCTTACTACACTTCACGGATAC[-/TGTACT]TGTACTTTTTTTCCAAATTGAAGGT	108
ASPA	443	rs36051310	GAGGTGGCTGAAAGAAGCACGTATCC[-/C]TGATGGCATGGTTGCGGGTTATATG	109
ASPA	443	rs36034906	GAGAAAAGCAGTTCCTGGAACACCCC[-/C]ACCCCTTAACCCCTTATCTCTGCTT	110
ASPA	443	rs36033666	TTACATATGTATACATGTGCCATGTT[-/T]GGTGTGCCGCACCCATTAACTCGTC	111
ASPA	443	rs35730123	CTTTTTCCAGATTTTTTTTTTTTTTT[-/T]GAGACAGAGTTTCACTCTTGTTGCC	112
ASPA	443	rs35629100	TTTGGAAATCTTAAGCTTTTATTTGG[-/G]TGTCACAGAGAAACAGGATCTGTAT	113
ASPA	443	rs35614631	TACTTTAAGTTTTAGGGTACATGTGC[-/A]CCATGTGCAGGTTTGTTACATATGT	114
ASPA	443	rs35225782	ATTCATGACCAGCCACATAAATGCAC[-/A]GTATTACTTCGCAAGCATGCCAATG	115
ASPA	443	rs35178659	GTGCACTAGAATTAGCTAAAGTGGGG[-/G]AAAAAAAGATGCATTTGATGGTCTA	116
ASPA	443	rs35095578	AACCTCCACCTCCCAGGTTCAAGAGA[-/A]TTCTCCTGCCTCAGCCTCCCAAGTA	117
ASPA	443	rs35002210	CCTCCCTGTGATCCGAAGTAGCAGAC[A/G]TA CTTAACTTCCATGGTGGATTGTT	118

Petitioner Illumina, Inc., EX1004, p. 614

ASPA	443	rs34744839	AAAACATTATTATATCTAGAAAAAAA[-/A]TGTATCTTAACCATTGTGGGAAGTG	119
ASPA	443	rs34680506	TTGAAGGTAAAATCATAGGGAGTTGG[-/G]AGCTGTCCTCTTGCGCTGAATCAGT	120
ASPA	443	rs34365618	ACTTGTGGCCTTTTTGGAGAGGTTAG[-/CA]ACTCTGAAAACTCTGTCCCTGGACC	121
ASPA	443	rs34275920	GAAGGAGAAAAAGAGAGGAAATAAGT[-/T]AAAATAATAAACACAATTAATAAAG	122
ASPA	443	rs34109510	TGTATACATGTGCCATGTTGGTGTGC[C/T]GC ACCCATTAACTCGTCATTTAGCA	123
ASPA	443	rs34054576	TCACCTGTCACCTCCTATAGAACTTT[-/C]CCCTGACCCTCCTCTATAGCATTAA	124
ASPA	443	rs34015272	ATAAATGATCATCATTCACAGTAGGG[-/G]TTTTGTTTTGTTTTTTTTCTGGAAA	125
ASPA	443	rs34002091	ACAGACATATCTACAAACACACTTTT[-/T]CACATATTTGTGTAAGTCATTTATG	126
ASPA	443	rs28940574	AAAGACAACTAAACTAACGCTCAATG[A/C]A AAAAGTATTCGCTGCTGTTTACAT	127
ASPA	443	rs28940279	TACCGTGTACCCCGTGTTTGTGAATG[A/C]GG CCGCATATTACGAAAAGAAAGAA	128
ASPA	443	rs17850703	CAGGGCTGGAGGTAAAACCATTTATT[A/G]CT AACCCCAGAGCAGTGAAGAAGTG	129
ASPA	443	rs17222495	TTCTTCATTGCCTATTGAAGAGAGAG[C/T]GG AATGCTTTGGTTGCCAGATATGG	130
ASPA	443	rs17175228	CACAAGATCTCATTACTCAGGAGCTG[C/T]CC AAGTGTCTAATGTACTTAGTTAA	131
ASPA	443	rs16953074	TTCTGTGTAACATTTCATTTAAGCAA[A/G]GG ATTCGGCAAATCAAAAATTGTCA	132
ASPA	443	rs16953070	TAAAACGTATTGAAGGTATTATTGAC[G/T]CT GTTGAAGCAAAGAGAACAAAACA	133
HEXA	3073	rs62022858	ATCTGCTCTTCCAGTTGGATGACAAG[C/T]CT TGCTGTCTAACACCTGCTGCAGA	134
HEXA	3073	rs62022857	CCATTTTTTGTTGTATTTTTTTTTTC[C/T]TGAA TACTTTTTATCGCAGTTGGTT	135
HEXA	3073	rs62017872	CCCTGTCTCTAAAAGAAAAAAAAAAA[A/G]A AAAAAAAAAGAAAACAAAACCCAA	136
HEXA	3073	rs62017871	AGTGGCTCCAAAAAGGTCATGGAACC[C/T]CT TGAGGATGATGCAAATTGACTCT	137
HEXA	3073	rs61662730	TAAAGTTACTTTTCTTTTATTGACTT[C/T]CCC TTATTTTTTAACCTTATGCTTT	138
HEXA	3073	rs61329913	CAGAGTTAAAAAAAAAAAAAAAAAAA[-/A]GGAAGTAGCAGCAACAGCTTGGAAA	139
HEXA	3073	rs60920713	GTTGCCCAGGGTTGAGTGCAGAGGCA[C/T]AT CTTGGCTCACAGCAACCTCTGCC	140
HEXA	3073	rs60783213	AAGGCTTTTTTTTTTTTTTTTTTTT[-/TTTT]GAGACAGAGTCTTGCTGTGTCACCC	141
HEXA	3073	rs60644867	GCCTACATTCTGCAAAGAGGAGGGAA[C/G]A TTCACAGCTCCATACTTGAACCCT	142
HEXA	3073	rs60288568	CCAAAGGAGAATAGCTCTAGGGGAGG[C/G]A GGTGGATGAGTATGCATGGGGGAG	143
HEXA	3073	rs59888548	GACTCCATCTCAAAAAAAAAAAAAAA[-/A]TGCAGTCTAATGGCAGAATTAGACT	144

Petitioner Illumina, Inc., EX1004, p. 615

HEXA	3073	rs59733856	TTATTTATTTATTTATTTATTTTTGA[A/G]ACA GGGTCTCTGTTGTCCAGGCTGG	145
HEXA	3073	rs59427837	TTTTGAGGCAGGGTCTCACTCTGTTG[C/T]CC AGGGTTGAGTGCAGAGGCACATC	146
HEXA	3073	rs59171976	CGCCTTGCGAAGGCCCCACAGCTTGC[C/T]TG TGACAAACGTTCATAGGCAAATG	147
HEXA	3073	rs58706602	GGAGGTCTGTACAAAGCACCACCTAC[C/T]TC ATGGGTCAGTTTCCACAGCAGAA	148
HEXA	3073	rs58696963	GAATCTTATAATTCACTGTGTACCTC[- /CCTC]TGTTTCATATTTTCGCAATTGAACT	149
HEXA	3073	rs58610850	AACATAGTATCTAATATAGCTTTACA[C/T]CC AAAGCCAAAATATGAATACACTG	150
HEXA	3073	rs58016062	TTGTTTTGTTTTGTTTGGGGGGGGGG[- /G]TTGTTTTTCTGAGAGGGAGTCTTGC	151
HEXA	3073	rs57733983	CATACCAAAGGGCAGCTGGAGGGATA[C/T]A GACGGAAGTCATGTGGAGAGTGAA	152
HEXA	3073	rs57476645	CAGGTGTGAGCCACCACGACCACCAA[A/T]T TAGCTCTTTTTACTCCTTCCCTTC	153
HEXA	3073	rs56870003	AGTGGTAGCTGATTTTGCTTCTGGAT[A/C]CT TTGCCACCTTCCCACTCTTTAAT	154
HEXA	3073	rs56338339	AAAGACCTGTTTCTTAAAAAAAAAAA[- /AGAAAAAAAAAAA]GAAAGAAAAGAAAAG AAAAAAACAG	155
HEXA	3073	rs55995352	TAAAAAATCTTTCAATGAGGAGATGT[C/T]CC CAGAGCAAGACAGCTGTAGGATG	156
HEXA	3073	rs55860138	AAAAGAAAAAAAAAAAAAAAAAAAA[- /A]GAAAACAAAACCCAAACCCATAAAG	157
HEXA	3073	rs55743646	CCTGTCTCTAAAAGAAAAAAAAAAAA[A/G]A AAAAAAAAGAAAACAAAACCCAAA	158
HEXA	3073	rs55665666	GTTATCATAGAAAAATATCACACTCT[- /GT]CTGTATCCCCACTTCCAGAAACTGT	159
HEXA	3073	rs36106892	CAGGAGCTCATAGAATTACATACAAT[- /C]TTTTTTTTTTTTTTTTGAGACAGCG	160
HEXA	3073	rs36091525	TTGAGAATCTTATAATTCACTGTGTA[- /CCTC]CCTCTGTTTCATATTTTCGCAATTG	161
HEXA	3073	rs35949555	CCACTACCACAGTGCCTAGAGAACAA[C/T]A TGTGTTTAATAATATTTAAATAAT	162
HEXA	3073	rs35827424	CCCTGTCTCTAAAAGAAAAAAAAAAA[- /A]AAAAAAAAAAGAAAACAAAACCCAA	163
HEXA	3073	rs35729578	CCATTATATCATTCATTTCCCACTCA[- /T]TTTCTTCATTCCAACCAAGATATAT	164
HEXA	3073	rs35649102	TCCGTCTCAAAAAAAAAAAAAAAAAG[- /A]GAAAGGAATTATTCTCATGTATACA	165
HEXA	3073	rs35118677	CTGGGGCAGTTAAAAAGAAAAACAAA[- /C]CCCTGGTCCCTGCCCTTGAGGAGAT	166
HEXA	3073	rs35005352	CTCCAGGGTCCCATTCCAGGACCACA[- /C]GCCTGCTACCTCTGCAGCTCACTCA	167
HEXA	3073	rs34736306	GGATTGACATATACCAGTTAGACGGA[- /T]TTTTTTTTTCCATAAACCAGGCTCA	168
HEXA	3073	rs34607939	ACAAATAATTACTACATATCTACAAC[A/G]TT CCAGATACAGAAGAAATGGCCAA	169

Petitioner Illumina, Inc., EX1004, p. 616

HEXA	3073	rs34496117	TAAACACACTTGAAACATCATATAAA[-/ATG]ATATTACTACAAGACTTAACCGTAA	170
HEXA	3073	rs34300017	ACACAGGTAATCCATGTTTATTATAG[-/A]AAAATGCCACATTACTCTTTATTGA	171
HEXA	3073	rs34206496	AGTTATCATAGAAAAATATCACACTC[-/TG]TCTGTATCCCCACTTCCAGAAACTG	172
HEXA	3073	rs34110830	AATGAACTTACAGGAAGGTAATATAT[-/G]GGAAATAAACATCTTATTGAATTTA	173
HEXA	3073	rs34093438	GGACCCCTGAAAGGCACAAGACACCC[-/T]TTCAGGTTCACACTTCCTGAAAGCT	174
HEXA	3073	rs34085965	CCACCAATCACCAGAGCCTTCTGCTC[A/G]GG GGTACCTGAGGGAAAACAAGCAA	175
HEXA	3073	rs34004907	AAAGACTGAAAAAACATTCATAACTA[-/T]TTTTCTTGTTATCCTCGGAAATGTC	176
HEXA	3073	rs28942072	TATCTTCATCTTGGAGGAGATGAGGT[C/T]GA TTTCACCTGCTGGAAGTCCAACC	177
HEXA	3073	rs28942071	TTGCCTATGAACGTTTGTCACACTTC[C/T]GCT GTGAGTTGCTGAGGCGAGGTGT	178
HEXA	3073	rs28941771	GCTTGCTGTTGGATACATCTCGCCAT[C/T]AC CTGCCACTCTCTAGCATCCTGGA	179
HEXA	3073	rs28941770	CCGGGGCTTGCTGTTGGATACATCTC[G/T]CC ATTACCTGCCACTCTCTAGCATC	180

3. Nucleic acid target length evaluation:

In some embodiments, aspects of the invention relate to methods for detecting nucleic acid deletions or insertions in regions containing nucleic acid sequence repeats.

Genomic regions that contain nucleic acid sequence repeats are often the site of genetic instability due to the amplification or contraction of the number of sequence repeats (e.g., the insertion or deletion of one or more units of the repeated sequence). Instability in the length of genomic regions that contain high numbers of repeat sequences has been associated with a number of hereditary and non hereditary diseases and conditions.

For example, "Fragile X syndrome, or Martin-Bell syndrome, is a genetic syndrome which results in a spectrum of characteristic physical, intellectual, emotional and behavioral features which range from severe to mild in manifestation. The syndrome is associated with the expansion of a single trinucleotide gene sequence (CGG) on the X chromosome, and results in a failure to express the FMR-1 protein which is required for normal neural development. There are four generally accepted forms of Fragile X syndrome which relate to the length of the repeated CGG sequence; Normal (29-31 CGG repeats), Premutation (55-200 CGG repeats), Full

Petitioner Illumina, Inc., EX1004, p. 617

Mutation (more than 200 CGG repeats), and Intermediate or Gray Zone Alleles (40 - 60 repeats)."

Other examples include cancer, which has been associated with microsatellite instability (MSI) involving an increase or decrease in the genomic copy number of nucleic acid repeats at one or more microsatellite loci (e.g., BAT-25 and/or BAT-26). The are currently many sequencing-based assays for determining the number of nucleic acid sequence repeats at a particular locus and identifying the presence of nucleic acid insertions or deletions. However, such techniques are not useful in a high throughput multiplex analysis where the entire length of a region may not be sequenced.

In contrast, in some embodiments, aspects of the invention relate to detecting the presence of an insertion or deletion at a genomic locus without requiring the locus to be sequenced (or without requiring the entire locus to be sequenced). Aspects of the invention are particularly useful for detecting an insertion or deletion in a nucleic acid region that contains high levels of sequence repeats. The presence of sequence repeats at a genetic locus is often associated with relatively high levels of polymorphism in a population due to insertions or deletions of one or more of the sequence repeats at the locus. The polymorphisms can be associated with diseases or predisposition to diseases (e.g., certain polymorphic alleles are recessive alleles associated with a disease or condition). However, the presence of sequence repeats often complicates the analysis of a genetic locus and increases the risk of errors when using sequencing techniques to determine the precise sequence and number of repeats at that locus.

In some embodiments, aspects of the invention relate to determining the size of a genetic locus by evaluating the capture frequency of a portion of that locus suspected of containing an insertion or deletion (e.g., due to the presence of sequence repeats) using a nucleic acid capture technique (e.g., a nucleic acid sequence capture technique based on molecular inversion probe technology). According to aspects of the invention, a statistically significant difference in capture efficiency for a genetic locus of interest in different biological samples (e.g., from different subjects) is indicative of different relative lengths in those samples. It should be appreciated that the length differences may be at one or both alleles of the genetic locus. Accordingly, aspects of the invention may be used to identify polymorphisms regardless of whether biological samples being interrogated at heterozygous or homozygous for the

Petitioner Illumina, Inc., EX1004, p. 618

polymorphisms. According to aspects of the invention, subjects that contain one or more loci with an insertion or deletion can be identified by analyzing capture efficiencies for nucleic acids obtained from one or more biological samples using appropriate controls (e.g., capture efficiencies for known nucleic acid sizes, capture efficiencies for other regions that are not suspected of containing an insertion or deletion in the biological sample(s), or predetermined reference capture efficiencies, or any combination thereof. However, it should be appreciated that aspects of the invention are not limited by the nature or presence of the control. In some embodiments, if a statistically significant variation in capture efficiency is detected, a subject may be identified as being at risk for a disease or condition associated with insertions or deletions at that genetic locus. In some embodiments, the subject may be analyzed in greater detail in order to determine the precise nature of the insertion or deletion and whether the subject is heterozygous or homozygous for one or more insertions or deletions. For example, gel electrophoresis of an amplification (e.g., PCR) product of the locus, or Southern blotting, or any combination thereof can be used as an orthogonal approach to verify the length of the locus. In some embodiments, a more exhaustive and detailed sequence analysis of the locus can be performed to identify the number and types of insertions and deletions. However, other techniques may be used to further analyze a locus identified as having an abnormal length according to aspects of the invention.

Accordingly, aspects of the invention relate to detecting abnormal nucleic acid lengths in genomic regions of interest. In some embodiments, the invention aims to estimate the size of genomic regions that are hard to be accessed, such as repetitive elements. However, it should be appreciated that methods of the invention do not require that the precise length be estimated. In some embodiments, it is sufficient to determine that one or more alleles with abnormal lengths are present at a locus of interest (e.g., based on the detection of abnormal capture efficiencies).

In a non-limiting example, fragile X can be used to illustrate aspects of the invention where the size of trinucleotide repeats (genotype) is linked to a symptom (phenotype). However, it should be appreciated that fragile X is a non-limiting example and similar analyses may be performed for other genetic loci (e.g., independently or simultaneously in multiplex analyses).

Use of molecular inversion probes (MIPs) has been demonstrated for detection of single nucleotide polymorphisms (Hardenbol et al. 2005 Genome Res 15:269-75) and for preparative amplification of large sets of exons (Porreca et al. 2007 Nat Methods 4:931-6, Krishnakumar et

Petitioner Illumina, Inc., EX1004, p. 619

al. 2008 Proc Natl Acad Sci USA 105:9296-301). In both cases, oligonucleotide probes are designed which have ends ('targeting arms') that hybridize up-stream and down-stream of the locus that is to be amplified.

In some embodiments, aspects of the invention are based on the recognition that the effect of length on probe capturing efficiency can be used in the context of an assay (e.g., a high throughput and/or multiplex assay) to allow the length of sequences to be determined without requiring sequencing of the entire region being evaluated. This is particularly useful for repeat regions that are prone to changes in size. As illustrated in FIG. 8, which is reproduced from Deng et al., Nature Biotech. 27:353-60, (see Supplemental Fig. 1G of Deng et al.,) illustrates that shorter sequences are captured with higher efficiency that longer sequences using MIPs. The statistical package R and its effects module were used for this analysis. A linear model was used, and each individual factor was assumed to be independent. The dashed lines represent a 95% confidence interval. Shorter target sequences were captured with higher efficiency than long target sequences ($p<2x10^{-16}$). However, the use of this differential capture efficiency for systematic sequence length analysis was not previously recognized.

In some embodiments, following probe hybridization, polymerase fill-in and ligation reactions are performed to convert the hybridized probe to a covalently-closed, circular molecule containing the desired target. PCR or rolling circle amplification plus exonuclease digestion of non-circularized material is performed to isolate and amplify the circular targets from the starting nucleic acid pool. Since one of the main benefits of the method is the potential for a high degree of multiplexing, generally thousands of targets are captured in a single reaction containing thousands of probes.

According to aspects of the invention, repetitive regions are surrounded by non repetitive unique sequences, which can be used to amplify the repeat-containing regions using, for example, PCR or padlock (MIP)-based method.

In addition to the repetitive regions, a probe (e.g., a MIP or padlock probe) can be designed to include at least a sequence that is sufficient to be uniquely identified in the genome (or target pool). After the probe is circularized and amplified, the amplicon can be end-sequenced so that the unique sequence can be identified and served as the "representative" of the repetitive region as illustrated in FIG. 9. FIG. 9 illustrates a non-limiting scheme of padlock (MIP) capture of a region that includes both repetitive regions (thick wavy line) and the adjacent

Petitioner Illumina, Inc., EX1004, p. 620

unique sequence (thick strait line). The regions of the probe are indicated with the targeting arms shown as regions "1" and "3." An intervening region that may be, or include, a sequencing primer binding site is shown as "2." After the padlock is circularized and amplified, it can be end-sequenced to obtain the sequence of the unique sequence, which represents the repetitive region of interest.

Although capturing efficiency is overall negatively correlated with target length, different probe sequences may have unique features. Therefore, multiple probes could be designed and tested so that an optimal one is chosen to be sensitive enough to differentiate repetitive sizes of roughly 0-150bp, 150-600bp, and beyond, which represent normal, premutation and full mutation of fragile X syndrome, respectively. However, it should be appreciated that other probe sizes and sequences can be designed, and optionally optimized, to distinguish a range of repeat region size differences (e.g., length differences of about 3-30 bases, about 30-60 bases, about 60-90 bases, about 90-120 bases, about 120-150 bases, about 150-300 bases, about 300-600 bases, about 600-900 bases, or any intermediate or longer length difference). It should be appreciated that a length difference may be an increase in size or a decrease in size.

In some embodiments, an initial determination of an unexpected capture frequency is indicative of the presence of size difference. In some embodiments, an increase in capture frequency is indicative of a deletion. In some embodiments, a decrease in capture frequency is indicative of an insertion. However, it should be appreciated that depending on specific sequence parameters and the relative sizes of the capture probes, the target region, and the deletions or insertions, a change in capture frequency can be associated with either an increase or decrease in target region length. In some embodiments, the precise nature of the change can be determined using one or more additional techniques as described herein.

Accordingly, in some aspects a MIP probe includes a linear nucleic acid strand that contains two hybridization sequences or targeting arms, one at each end of the linear probe, wherein each of the hybridization sequences is complementary to a separate sequence on a the same strand of a target nucleic acid, and wherein these sequences on the target nucleic acid flank the two ends of the target nucleic acid sequence of interest. It should be appreciated that upon hybridization, the two ends of the probe are inverted with respect to each other in the sense that both 5' and 3' ends of the probe hybridize to the same strand to separate regions flanking the target region (as illustrated in FIG. 9 for example).

Petitioner Illumina, Inc., EX1004, p. 621

In some embodiments, the hybridization sequences are between about 10-100 nucleotides long, for example between about 10-30, about 30-60, about 60-90, or about 20, about 30, about 40, or about 50 nucleotides long. However, other lengths may be used depending on the application. In some embodiments, the hybridization Tms of both targeting arms of a probe are designed or selected to be similar. In some embodiments, the hybridization Tms of the targeting arms of a plurality of probes designed to capture different target regions are selected or designed to be similar so that they can be used together in a multiplex reaction. Accordingly, a typical size of a MIP probe prior to fill in is about 60-80 nucleotides long. However, other sizes can be used depending on the sizes of the targeting arms and any other sequences (e.g., primer binding or tag sequences) that are present in the MIP probe. In some embodiments, MIP probes are designed to avoid sequence-dependent secondary structures. In some embodiments, MIP probes are designed such that the targeting arms do not overlap with known polymorphic regions. In some embodiments, targeting arms that can be used for capturing the repeat region of the Fragile X locus can have the following sequences or complementary to these sequences depending on the strand that is captured.

left: CTCCUITYCGGTITCACTTC (SEQ ID NO: 181)

right: ATCTTCTCTTCAGCCCTGCT (SEQ ID NO: 182)

The typical captured size using these targeting arms is about 100 nucleotides in length (e.g., about 30 repeats of a tri-nucleotide repeat).

In some embodiments, the number of reads obtained for the "representative" of the repetitive region is not informative to estimate the target length because it is dependent on the total number of reads obtained. To overcome this, it is useful to include one or more probes that target other "control" regions where no or minimal polymorphism exists among populations. Because of the systematic consistency of capturing efficiency (see, e.g., FIG. 9), the ratio of reads obtained for the repetitive "representative" to reads obtained for the control region(s) will be tuned using DNA with defined numbers of repeats. Ultimately, the ratio can serve as a measure of the repeat length as illustrated in FIG. 10. FIG. 10 illustrates a non-limiting hypothetical relationship between target gap size and the relative number of reads of the repetitive region, which is measured by the ratio of the repeat "representative" reads vs. the "control" region reads. The unit of y-axis is arbitrary.

Petitioner Illumina, Inc., EX1004, p. 622

In some embodiments, to better tell targets with similar size range apart, the whole repetitive region can be sequenced by making a shotgun library (e.g., by making a shotgun library from a captured sequence, for example a sequence captured using a MIP probe). The longer the repeat is, the more short reads of repeats will be obtained. Therefore, the target length will contribute twice to the relative number of "repetitive" reads, which will gain better resolution of differentiating targets. In some embodiments, the expectation is that the number of reads from any given repeat will be a direct function of the number of repeats present. However, in some embodiments, a Poisson sampling induced spread may need to be considered and in some embodiments may be sufficiently large to limit the resolution.

When a precise measurement of the length of both alleles from a diploid sample is desired, further manipulations may be required. This is because the capture efficiency measured will actually be the average efficiency of the two alleles. To effectively achieve separate measurements for each allele, barcodes (e.g., sequence tags) can be used that allow the efficiency of individual capture events (from individual genomic loci) to be followed. FIG. 11A-C shows the approach. For a given locus, MIPs are synthesized to contain one of a large number differentiator tags in their backbone such that the probability of any two MIPs in a reaction having the same differentiator tag sequence is low. MIP capture is performed on the sample; the reaction will be biased for shorter target lengths, and therefore the reaction product will be comprised of more 'short' circles than 'long' circles. Each circle should bear a unique differentiator tag sequence. Then, linear RCA (1RCA) is performed on the circles. In the 1RCA reaction, circles are converted into long, linear concatemers of themselves. The 1RCA reaction for a given circle stops when the concatemer has reached a 'fixed' length (based on the processivity/error rate of the polymerase). Concatemers derived from smaller circles will therefore contain more copies of the differentiator tag, and concatemers derived from larger circles will contain fewer copies of the differentiator tag. The number of each differentiator tag sequence is counted, for example, by next-generation sequencing.

When number of occurrences is plotted against differentiator tag ID, the data will naturally cluster into two groups reflecting the lengths of the two alleles in the diploid sample. The allele lengths can therefore be read directly off this graph, after absolute length calibration using known standards. In some embodiments, a sequencing technique (e.g., a next-generation sequencing technique) is used to sequence part of one or more captured targets (e.g., or

Petitioner Illumina, Inc., EX1004, p. 623

amplicons thereof) and the sequences are used to count the number of different barcodes that are present. Accordingly, in some embodiments, aspects of the invention relate to a highly-multiplexed qPCR reaction.

Other non-limiting examples of loci at which insertions or deletions or repeat sequences may be associated with a disease or condition are provided in Tables 3 and 4. It should be appreciated that the presence of an abnormal length at any one or more of these loci may be evaluated according to aspects of the invention. In some embodiments, two or more of these loci or other loci may be evaluated in a single multiplex reaction using different probes designed to hybridize under the same reaction conditions to different target nucleic acid in a biological sample.

Table 3.

Polyglutamine (PolyQ) Diseases

Type	Gene	Normal/wildtype	Pathogenic
DRPLA (Dentatorubropallidoluysian atrophy)	ATN1 or DRPLA	6 - 35	49 - 88
HD (Huntington's disease)	HTT (Huntingtin)	10 - 35	35+
SBMA (Spinobulbar muscular atrophy or Kennedy disease)	Androgen receptor on the X chromosome.	9 - 36	38 - 62
SCA1 (Spinocerebellar ataxia Type 1)	ATXN1	6 - 35	49 - 88
SCA2 (Spinocerebellar ataxia Type 2)	ATXN2	14 - 32	33 - 77
SCA3 (Spinocerebellar ataxia Type 3 or Machado-Joseph disease)	ATXN3	12 - 40	55 - 86
SCA6 (Spinocerebellar ataxia Type 6)	CACNA1A	4 - 18	21 - 30
SCA7 (Spinocerebellar ataxia Type 7)	ATXN7	7 - 17	38 - 120
SCA17 (Spinocerebellar ataxia Type 17)	TBP	25 - 42	47 - 63

Table 4.

Petitioner Illumina, Inc., EX1004, p. 624

Non-Polyglutamine Diseases

Type	Gene	Codon	Normal/wildtype	Pathogenic
FRAXA (Fragile X syndrome)	*FMR1*, on the X-chromosome	CGG	6 - 53	230+
FXTAS (Fragile X-associated tremor/ ataxia syndrome)	*FMR1*, on the X-chromosome	CGG	6 - 53	55-200
FRAXE (Fragile XE mental retardation)	AFF2 or FMR2, on the X-chromosome	GCC	6 - 35	200+
FRDA (Friedreich's ataxia)	FXN or X25, (frataxin)	GAA	7 - 34	100+
DM (Myotonic dystrophy)	DMPK	CTG	5 - 37	50+
SCA8 (Spinocerebellar ataxia Type 8)	OSCA or SCA8	CTG	16 - 37	110 – 250
SCA12 (Spinocerebellar ataxia Type 12)	PPP2R2B or SCA12	CAG On 5' end	7 - 28	66 - 78

The following examples illustrate aspects and embodiments of the invention and are not intended to be limiting or restrictive. Many variations of the invention will become apparent to those skilled in the art upon review of this specification. The full scope of the invention should be determined by reference to the claims, along with their full scope of equivalents, and the specification, along with such variations.

4. Increasing detection sensitivity:

In some embodiments, aspects of the invention relate to methods for increasing the sensitivity of nucleic acid detection assays.

There are currently many genomic assays that utilize next-generation (e.g., polony-based) sequencing to generate data, including genome resequencing, RNA-seq for gene expression, bisulphite sequencing for methylation, and Immune-seq, among others. In order to make

Petitioner Illumina, Inc., EX1004, p. 625

quantitative measurements (including genotype calling), these methods utilize the counts of sequencing reads of a given genomic locus as a proxy for the representation of that sequence in the original sample of nucleic acids. The majority of these techniques require a preparative step to construct a high-complexity library of DNA molecules that is representative of a sample of interest. Current assays use one of several alternative nucleic acid preparative techniques (e.g., amplification, for example PCR-based amplification; sequence-specific capture, for example, using immobilized capture probes; or target capture into a circularized probe followed by a sequence analysis step. In order to reduce errors associated with the unpredictability (stochastic nature) of nucleic acid isolation and sequence analysis techniques, current methods involve oversampling a target nucleic acid preparation in order to increase the likelihood that all sequences that are present in the original nucleic acid sample will be represented in the final sequence data. For example, a genomic sequencing library may contain an over- or under-representation of particular sequences from a source nucleic acid sample (e.g., genome preparation) as a result of stochastic variations in the library construction process. Such variations can be particularly problematic when they result in target sequences from a genome being absent or undetectable in a sequencing library. For example, an under-representation of particular allelic sequences (e.g., heterozygotic alleles) from a genome in a sequencing library can result in an apparent homozygous representation in a sequencing library.

In contrast, aspects of the invention relate to basing a nucleic acid sequence analysis on results from two or more different nucleic acid preparatory techniques that have different systematic biases in the types of nucleic acids that they sample rather than simply oversampling the target nucleic acid. According to some embodiments, different techniques have different sequence biases that are systematic and not simply due to stochastic effects during nucleic acid capture or amplification. Accordingly, in some embodiments, the degree of oversampling required to overcome variations in nucleic acid preparation needs to be sufficient to overcome the biases. In some embodiments, the invention provides methods that reduce the need for oversampling by combining nucleic acid and/or sequence results obtained from two or more different nucleic acid preparative techniques that have different biases.

According to the invention, different techniques have different characteristic or systematic biases. For example, one technique may bias a sample analysis towards one particular allele at a genetic locus of interest, whereas a different technique would bias the

Petitioner Illumina, Inc., EX1004, p. 626

sample analysis towards a different allele at the same locus. Accordingly, the same sample may be identified as being different depending on the type of technique that is used to prepare nucleic acid for sequence analysis. This effectively represents a sensitivity issue, because each technique has a different relative sensitivities for polymorphic sequences of interest.

According to aspects of the invention, the sensitivity of a nucleic acid analysis can be increased by combining the sequences from different nucleic acid preparative steps and using the combined sequence information for a diagnostic assay (e.g., for a making a call as to whether a subject is homozygous or heterozygous at a genetic locus of interest).

Currently, the ability of DNA sequencing to detect mutations is limited by the ability of the upstream sample isolation (e.g., by amplification, immobilization enrichment, circularization capture, etc.) methods to reliably isolate the locus of interest.

If one wishes to make heterozygote base-calls for a diploid genome (e.g. a human sample presented for molecular diagnostic sequencing), it is important in some embodiments that the isolation method produces near- or perfectly- uniform amounts of the two alleles to be sequenced (at least sufficiently uniform to be "called" unambiguously as a heterozygote or a homozygote for a locus of interest).

Sample preparative methods may fall into three classes: 1) single- or several target amplification (e.g., uniplex PCR, 'multiplex' PCR), 2) multi-target hybridization enrichment (e.g., Agilent SureSelect 'hybrid capture' [Gnirke et al 2009, *Nature methods* 27:182-9], Roche/Nimblegen 'sequence capture' [Hodges et al 2007, *Nature genetics* 39:1522-7], and 3) multi-target circularization selection (e.g. molecular inversion probes or padlock probes, [Porreca et al 2007, *Nature methods* 4:931-6, Turner et al 2009, *Nature methods* 6:315-6], 'selectors' [Dahl et al 2005, *Nucleic acids research* 33:e71]). Each of these methods can result in a pool of isolated product that does not adequately represent the input abundance distribution. For example, the two alleles at a heterozygous position can become skewed far from their input 50:50 ratio to something that results in a missed basecall during downstream sequencing. For example, if the ratio was skewed from 50:50 to 10:90, and the sample was sequenced to 10X average coverage, there is a high probability that one of the two alleles would not be observed once in the ten sequencing reads. This would reduce the sensitivity of the sequencing method by converting a heterozygous position to homozygous (where potentially the 'mutant' allele was the one not observed). In some embodiments, a skewed ratio is a particular issue that decreases the

Petitioner Illumina, Inc., EX1004, p. 627

sensitivity of detecting mutations present in a heterogeneous tumor tissue. For example, if only 10% of the cells analyzed in a heterogeneous sample harbored a heterozygous mutation, the mutation would be expected to be present in 5% of sequence reads, not 50%. In this scenario, the need for robust, sensitive detection may be even more acute.

The methods disclosed herein are based, in part, on the discovery that certain classes of isolation methods have different modes of bias. The disclosure provide methods for increasing the sensitivity of the downstream sequencing by using a combination of multiple isolation methods (e.g., one or more from at least two of the classes disclosed herein) for a sample. This is particularly important in molecular diagnostics where high sensitivity is required to minimize the chances of 'missing' a disease-associated mutation. For example, given a nominal false-negative error rate of $1x10^{-3}$ for sequencing following circularization selection, and a false-negative error rate of $1x10^{-3}$ for sequencing following hybridization enrichment, one can achieve a final false-negative rate of $1x10^{-6}$ by performing both techniques on the sample (assuming failures in each method are fully independent). For a recessive disease with carrier frequency of 0.1, caused by a single fully-penetrant mutant allele, the number of missed carrier diagnoses would decrease from 1000 per million patients tested to 1 per million patients tested. Furthermore, if the testing was used in the context of prenatal carrier screening, the number of affected children born as a result of missing the carrier call in one parent would decrease from 25 per million to 25 per billion born.

Additionally, the disclosure provides combinations of preparative methods to effectively increase sequencing coverage in regions containing disease-associated alleles. Since heterozygote error rate is largely tied to both deviations from 50:50 allele representation, and in the case of next-generation DNA sequencing deviations from average abundance (such that less abundant isolated targets are more likely to be undersampled at one or both alleles), selectively increasing coverage in these regions will also selectively increase sensitivity. Furthermore, MIPs that detect presence or absence of specific known disease-associated mutations can be used to increase sensitivity selectively. In some embodiments, these MIPs would have a targeting arm whose 3' -most region is complementary to the expected mutation, and has a fill-in length of 0 or more bp. Thus, the MIP will form only if the mutation is present, and its presence will be detected by sequencing.

Petitioner Illumina, Inc., EX1004, p. 628

Additionally, algorithms disclosed herein may be used to determine base identity with varying levels of stringency depending on whether the given position has any known disease-associated alleles. Stringency can be reduced in such positions by decreasing the minimum number of observed mutant reads necessary to make a consensus base-call. This will effectively increase sensitivity for mutant allele detection at the cost of decreased specificity.

An embodiment of the invention combines MIPs plus hybridization enrichment, plus optionally extra MIPs targeted to specific known, common disease-associated loci, e.g., to detect the presence of a polymorphism in a target nucleic acid. A non-limiting example is illustrated in FIG. 12 that illustrates a schematic using MIPs plus hybridization enrichment, plus optionally extra MIPs targeted to specific known, common disease-associated loci, e.g., to detect the presence of a polymorphism in a target nucleic acid.

FIGs. 13 and 14 illustrate different capture efficiencies for MIP-based captures. FIG. 13 shows a graph of per-target abundance with MIP capture. In this graph, bias largely drives the heterozygote error rate, since targets which are less abundant here are less likely to be covered in sufficient depth during sequencing to adequately sample both alleles. This is from Turner et al 2009, Nature methods 6:315-6. Hybridization enrichment results in a qualitatively similar abundance distribution, but the abundance of a given target is likely not correlated between the two methods. FIG. 14 shows a graph of correlation between two MIP capture reactions from Ball et al 2009, Nature biotechnology 27:361-8. Each point represents the target abundance in replicate 1 and replicate 2. Pearson correlation r=0.956. This indicates that MIP capture reproducibly biases targets to specific abundances. Hybridization enrichment is similarly correlated from one capture to the next.

According to aspects of the invention, such biases can be detected or overcome by systematically combining different capture and/or analytical techniques in an assay that interrogates a plurality of loci in a plurality of subject samples.

Accordingly, it should be appreciated that in any of the embodiments described herein (e.g., tiling/staggering, tagging, size-detection, sensitivity enhancing algorithms, or any combination thereof), aspects of the invention involve preparing genomic nucleic acid and/or contacting them with one or more different probes (e.g., capture probes, hybridization probes, MIPs, others etc.). In some embodiments, the amount of genomic nucleic acid used per subject ranges from 1 ng to 10 micrograms (e.g., 500 ng to 5 micrograms). However, higher or lower

Petitioner Illumina, Inc., EX1004, p. 629

amounts (e.g., less than 1 ng, more than 10 micrograms, 10-50 micrograms, 50-100 micrograms or more) may be used. In some embodiments, for each locus of interest, the amount of probe used per assay may be optimized for a particular application. In some embodiments, the ratio (molar ratio, for example measured as a concentration ratio) of probe to genome equivalent (e.g., haploid or diploid genome equivalent, for example for each allele or for both alleles of a nucleic acid target or locus of interest) ranges from 1/100, 1/10, 1/1, 10/1, 100/1, 1000/1. However, lower, higher, or intermediate ratios may be used.

In some embodiments, the amount of target nucleic acid and probe used for each reaction is normalized to avoid any observed differences being caused by differences in concentrations or ratios. In some embodiments, in order to normalize genomic DNA and probe, the genomic DNA concentration is read using a standard spectrophotometer or by fluorescence (e.g., using a fluorescent intercalating dye). The probe concentration may be determined experimentally or using information specified by the probe manufacturer.

Similarly, once a locus has been captured (e.g., on a MIP or other probe or in another form), it may be amplified and/or sequenced in a reaction involving one or more primers. The amount of primer added for each reaction can range from 0.1 pmol to 1 nmol, 0.15 pmol to 1.5 nmol (for example around 1.5 pmol). However, other amounts (e.g., lower, higher, or intermediate amounts) may be used.

In some embodiments, it should be appreciated that one or more intervening sequences (e.g., sequence between the first and second targeting arms on a MIP capture probe), identifier or tag sequences, or other probe sequences that are not designed to hybridize to a target sequence (e.g., a genomic target sequence) should be designed to avoid excessive complementarity (to avoid cross-hybridization) to target sequences or other sequences (e.g., other genomic sequences) that may be in a biological sample. For example, these sequences may be designed have a sufficient number of mismatches with any genomic sequence (e.g., at least 5, 10, 15, or more mismatches out of 30 bases) or as having a Tm (e.g., a mismatch Tm) that is lower (e.g., at least 5, 10, 15, 20, or more degrees C lower) than the hybridization reaction temperature.

It should be appreciated that a targeting arm as used herein may be designed to hybridize (e.g., be complementary) to either strand of a genetic locus of interest if the nucleic acid being analyzed is DNA (e.g., genomic DNA). However, in the context of MIP probes, whichever strand is selected for one targeting arm will be used for the other one. However, in the context of

Petitioner Illumina, Inc., EX1004, p. 630

RNA analysis, it should be appreciated that a targeting arm should be designed to hybridize to the transcribed RNA. It also should be appreciated that MIP probes referred to herein as "capturing" a target sequence are actually capturing it by template-based synthesis rather than by capturing the actual target molecule (other than for example in the initial stage when the arms hybridize to it or in the sense that the target molecule can remain bound to the extended MIP product until it is denatured or otherwise removed).

It should be appreciated that in some embodiments a targeting arm may include a sequence that is complementary to one allele or mutation (e.g., a SNP or other polymorphism, a mutation, etc.) so that the probe will preferentially hybridize (and capture) target nucleic acids having that allele or mutation. However, in many embodiments, each targeting arm is designed to hybridize (e.g., be complementary) to a sequence that is not polymorphic in the subjects of a population that is being evaluated. This allows target sequences to be captured and/or sequenced for all alleles and then the differences between subjects (e.g., calls of heterozygous or homozygous for one or more loci) can be based on the sequence information and/or the frequency as described herein.

It should be appreciated that sequence tags (also referred to as barcodes) may be designed to be unique in that they do not appear at other positions within a probe or a family of probes and they also do not appear within the sequences being targeted. Thus they can be used to uniquely identify (e.g., by sequencing or hybridization properties) particular probes having other characteristics (e.g., for particular subjects and/or for particular loci).

It also should be appreciated that in some embodiments probes or regions of probes or other nucleic acids are described herein as comprising or including certain sequences or sequence characteristics (e.g., length, other properties, etc.). However, it should be appreciated that in some embodiments, any of the probes or regions of probes or other nucleic acids consist of those regions (e.g., arms, central regions, tags, primer sites, etc., or any combination thereof) of consist of those sequences or have sequences with characteristics that consist of one or more characteristics (e.g., length, or other properties, etc.) as described herein in the context of any of the embodiments (e.g., for tiled or staggered probes, tagged probes, length detection, sensitivity enhancing algorithms or any combination thereof).

It should be appreciated that probes, primers, and other nucleic acids designed or used herein may be synthetic, natural, or a combination thereof. Accordingly, as used herein, the term

Petitioner Illumina, Inc., EX1004, p. 631

"nucleic acid" refers to multiple linked nucleotides (i.e., molecules comprising a sugar (e.g., ribose or deoxyribose) linked to an exchangeable organic base, which is either a pyrimidine (e.g., cytosine (C), thymidine (T) or uracil (U)) or a purine (e.g., adenine (A) or guanine (G)). "Nucleic acid" and "nucleic acid molecule" may be used interchangeably and refer to oligoribonucleotides as well as oligodeoxyribonucleotides. The terms shall also include polynucleosides (i.e., a polynucleotide minus a phosphate) and any other organic base containing nucleic acid.

The organic bases include adenine, uracil, guanine, thymine, cytosine and inosine. Unless otherwise stated, nucleic acids may be single or double stranded. The nucleic acid may be naturally or non-naturally occurring. Nucleic acids can be obtained from natural sources, or can be synthesized using a nucleic acid synthesizer (i.e., synthetic).

Harvest and isolation of nucleic acids are routinely performed in the art and suitable methods can be found in standard molecular biology textbooks. (See, for example, Maniatis' Handbook of Molecular Biology.) The nucleic acid may be DNA or RNA, such as genomic DNA, mitochondrial DNA, mRNA, cDNA, rRNA, miRNA, or a combination thereof. Non-naturally occurring nucleic acids such as bacterial artificial chromosomes (BACs) and yeast artificial chromosomes (YACs) can also be used.

The invention also contemplates the use of nucleic acid derivatives. As will be described herein, the use of certain nucleic acid derivatives may increase the stability of the nucleic acids of the invention by preventing their digestion, particularly when they are exposed to biological samples that may contain nucleases. As used herein, a nucleic acid derivative is a non-naturally occurring nucleic acid or a unit thereof. Nucleic acid derivatives may contain non-naturally occurring elements such as non-naturally occurring nucleotides and non-naturally occurring backbone linkages.

Nucleic acid derivatives may contain backbone modifications such as but not limited to phosphorothioate linkages, phosphodiester modified nucleic acids, phosphorothiolate modifications, combinations of phosphodiester and phosphorothioate nucleic acid, methylphosphonate, alkylphosphonates, phosphate esters, alkylphosphonothioates, phosphoramidates, carbamates, carbonates, phosphate triesters, acetamidates, carboxymethyl esters, methylphosphorothioate, phosphorodithioate, p-ethoxy, and combinations thereof. The backbone composition of the nucleic acids may be homogeneous or heterogeneous.

Petitioner Illumina, Inc., EX1004, p. 632

Nucleic acid derivatives may contain substitutions or modifications in the sugars and/or bases. For example, they include nucleic acids having backbone sugars which are covalently attached to low molecular weight organic groups other than a hydroxyl group at the 3' position and other than a phosphate group at the 5' position (e.g., an 2'-0-alkylated ribose group). Nucleic acid derivatives may include non-ribose sugars such as arabinose. Nucleic acid derivatives may contain substituted purines and pyrimidines such as C-5 propyne modified bases, 5-methylcytosine, 2-aminopurine, 2-amino-6-chloropurine, 2,6-diaminopurine, hypoxanthine, 2-thiouracil and pseudoisocytosine. In some embodiments, substitution(s) may include one or more substitutions/modifications in the sugars/bases, groups attached to the base, including biotin, fluorescent groups (fluorescein, cyanine, rhodamine, etc), chemically-reactive groups including carboxyl, NHS, thiol, etc., or any combination thereof.

A nucleic acid may be a peptide nucleic acid (PNA), locked nucleic acid (LNA), DNA, RNA, or co-nucleic acids of the same such as DNA-LNA co-nucleic acids. PNA are DNA analogs having their phosphate backbone replaced with 2-aminoethyl glycine residues linked to nucleotide bases through glycine amino nitrogen and methylenecarbonyl linkers. PNA can bind to both DNA and RNA targets by Watson-Crick base pairing, and in so doing form stronger hybrids than would be possible with DNA or RNA based oligonucleotides in some cases.

PNA are synthesized from monomers connected by a peptide bond (Nielsen, P.E. et al. Peptide Nucleic Acids, Protocols and Applications, Norfolk: Horizon Scientific Press, p. 1-19 (1999)). They can be built with standard solid phase peptide synthesis technology. PNA chemistry and synthesis allows for inclusion of amino acids and polypeptide sequences in the PNA design. For example, lysine residues can be used to introduce positive charges in the PNA backbone. All chemical approaches available for the modifications of amino acid side chains are directly applicable to PNA. Several types of PNA designs exist, and these include single strand PNA (ssPNA), bisPNA and pseudocomplementary PNA (pcPNA).

The structure of PNA/DNA complex depends on the particular PNA and its sequence. ssPNA binds to single stranded DNA (ssDNA) preferably in antiparallel orientation (i.e., with the N-terminus of the ssPNA aligned with the 3' terminus of the ssDNA) and with a Watson-Crick pairing. PNA also can bind to DNA with a Hoogsteen base pairing, and thereby forms triplexes with double stranded DNA (dsDNA) (Wittung, P. et al., *Biochemistry* 36:7973 (1997)).

Petitioner Illumina, Inc., EX1004, p. 633

A locked nucleic acid (LNA) is a modified RNA nucleotide. An LNA form hybrids with DNA, which are at least as stable as PNA/DNA hybrids (Braasch, D.A. et al., *Chem & Biol.* 8(1):1-7(2001)). Therefore, LNA can be used just as PNA molecules would be. LNA binding efficiency can be increased in some embodiments by adding positive charges to it. LNAs have been reported to have increased binding affinity inherently.

Commercial nucleic acid synthesizers and standard phosphoramidite chemistry are used to make LNAs. Therefore, production of mixed LNA/DNA sequences is as simple as that of mixed PNA/peptide sequences. The stabilization effect of LNA monomers is not an additive effect. The monomer influences conformation of sugar rings of neighboring deoxynucleotides shifting them to more stable configurations (Nielsen, P.E. et al. Peptide Nucleic Acids, Protocols and Applications, Norfolk: Horizon Scientific Press, p. 1-19 (1999)). Also, lesser number of LNA residues in the sequence dramatically improves accuracy of the synthesis. Most of biochemical approaches for nucleic acid conjugations are applicable to LNA/DNA constructs.

These and other aspects of the invention are illustrated by the following non-limiting examples.

EXAMPLES

The following examples illustrate non-limiting embodiments of the invention.

Example 1: Design a set of capture probes for a human target exon

All targets are captured as a set of partially-overlapping subtargets. For example, in the tiling approach, a 200 bp target exon might be captured as a set of 12 subtargets, each 60bp in length (FIG. 1). Each subtarget is chosen such that it partially overlaps two or three other targets.

In some embodiments, all probes are composed of three regions: 1) a 20 bp 'targeting arm' comprised of sequence which hybridizes immediately upstream from the sub-target, 2) a 30 bp 'constant region' comprised of sequence used as a pair of amplification priming sites, and 3) a second 20 bp 'targeting arm' comprised of sequence which hybridizes immediately downstream from the sub-target. Targeting arm sequences will be different for each capture probe in a set, while constant region sequence will be the same for all probes in the set, allowing all captured targets to be amplified with a single set of primers. Targeting arm sequences should be designed

Petitioner Illumina, Inc., EX1004, p. 634

such that any given pair of 20 bp sequences is unique in the target genome (to prevent spurious capture of undesired sites). Additionally, melting temperatures should be matched for all probes in the set such that hybridization efficiency is uniform for all probes at a constant temperature (e.g., 60 C). Targeting arm sequences should be computationally screened to ensure they do not form strong secondary structure that would impair their ability to basepair with the genomic target.

Hybridize capture probes to human genomic sample

Assemble hybridization reaction:

- 1.0 ul capture probe mix (~2.5 pmol)
- 2.0 ul 10X Ampligase buffer (Epicentre)
- 6.0 ul 500 ng/ul human genomic DNA (~16.7 fmol)
- 11 ul dH20

In a thermal cycler, heat reaction to 95C for 5 min to denature genomic DNA, then cool to 60C. Allow to incubate at 60C for 40 hours.

Convert hybridized probes into covalently-closed circular products containing subtargets

Prepare fill-in / ligation reaction mixture:

- 0.25 ul 2 mM dNTP mix (Invitrogen)
- 2.5 ul 10X Ampligase buffer (Epicentre)
- 5.0 ul 5 U/ul Taq Stoffel fragment (Applied Biosystems)
- 12.5 ul 5 U/ul Ampligase (Epicentre)
- 4.75 ul dH20

Add 1.0 ul of this mix to the hybridized probe reaction, and incubate at 60C for 10 hours.

Purify circularized probe / subtarget products from un-reacted probes and genomic DNA

Prepare exonuclease reaction mixture:

- 21 ul fill-in / ligation reaction product
- 2.0 ul 10X exonuclease I buffer (New England Biolabs)

Petitioner Illumina, Inc., EX1004, p. 635

- 2.0 ul 20 U/ul exonuclease I (New England Biolabs)
- 2.0 ul 100 U/ul exonuclease III (New England Biolabs)

Incubate at 37C for 60 min, then heat-inactivate by incubating at 80C for 15 min. Immediately cool to 4C for storage.

Amplify circular material by PCR using primers specific to the 'constant region' of the probes

Prepare PCR mixture:

- 5.0 ul 10X Accuprime reaction buffer (Invitrogen)
- 1.5 ul 10 uM CP-2-FA (5'-GCACGATCCGACGGTAGTGT-3') (SEQ ID NO:183)
- 1.5 ul 10 uM CP-2-RA (5'-CCGTAATCGGGAAGCTGAAG-3') (SEQ ID NO:184)
-
- 0.4 ul 25 mM dNTP mix (Invitrogen)
- 2.0 ul heat-inactivated exonuclease reaction mix
- 1.5 ul 10X SybrGreen (Invitrogen)
- 0.4 ul 2.5 U/ul Accuprime Pfx polymerase (Invitrogen)
- 37.7 ul dH20

Thermal cycle in real-time thermal cycler according to the following protocol, but stop cycling before amplification yield plateaus (generally 8-12 cycles):

1. 95C for 5 min
2. 95C for 30 sec
3. 58C for 60 sec
4. 72C for 60 sec
5. goto 2, N more times

Prepare a shotgun next-generation sequencing library for analysis

- Purify desired amplicon population from non-specific amplification products by gel extraction.

Petitioner Illumina, Inc., EX1004, p. 636

- Concatemerize amplicons into high-molecular weight products suitable for shearing
- Mechanically shear, using either a nebulizer, BioRuptor, Hydroshear, Covaris, or similar instrument. DNA should be sheared into fragments several hundred basepairs in length.
- Ligate adapters required for amplification by the sequencing platform used. If necessary, purify ligated product from unligated product and adapters.

Example 2: Use of differentiator tag sequences to detect and correct bias in a MIP-capture reaction of a set of exon targets

The first step in performing the detection/correction is to determine how many differentiator tag sequences are necessary for the given sample. In this example, 1000 genomic targets corresponding to 1000 exons were captured. Since the differentiator tag sequence is part of the probe, it will measure/report biases that occur from the earliest protocol steps. Also, being located in the backbone, the differentiator tag sequence can easily be sequenced from a separate priming site, and therefore not impact the total achievable read-length for the target sequence. MIP probes are synthesized using standard column-based oligonucleotide synthesis by any number of vendors (e.g. IDT), and differentiator tag sequences are introduced as 'degenerate' positions in the backbone. Each degenerate position increases the total number of differentiator tag sequences synthesized by a factor of 4, so a 10 nt degenerate region implies a differentiator tag sequence complexity of ~1e6 species.

Hybridize capture probes to human genomic sample

Assemble hybridization reaction:

- 1.0 ul capture probe mix (~2.5 pmol)
- 2.0 ul 10X Ampligase buffer (Epicentre)
- 6.0 ul 500 ng/ul human genomic DNA (~16.7 fmol)
- 11 ul dH20

In a thermal cycler, heat reaction to 95C for 5 min to denature genomic DNA, then cool to 60C. Allow to incubate at 60C for 40 hours.

Convert hybridized probes into covalently-closed circular products containing subtargets

Petitioner Illumina, Inc., EX1004, p. 637

Prepare fill-in / ligation reaction mixture:

- 0.25 ul 2 mM dNTP mix (Invitrogen)
- 2.5 ul 10X Ampligase buffer (Epicentre)
- 5.0 ul 5 U/ul Taq Stoffel fragment (Applied Biosystems)
- 12.5 ul 5 U/ul Ampligase (Epicentre)
- 4.75 ul dH20

Add 1.0 ul of this mix to the hybridized probe reaction, and incubate at 60C for 10 hours.

Purify circularized probe / subtarget products from un-reacted probes and genomic DNA

Prepare exonuclease reaction mixture:

- 21 ul fill-in / ligation reaction product
- 2.0 ul 10X exonuclease I buffer (New England Biolabs)
- 2.0 ul 20 U/ul exonuclease I (New England Biolabs)
- 2.0 ul 100 U/ul exonuclease III (New England Biolabs)

Incubate at 37C for 60 min, then heat-inactivate by incubating at 80C for 15 min. Immediately cool to 4C for storage.

Amplify circular material by PCR using primers specific to the 'constant region' of the probes

Prepare PCR mixture:

- 5.0 ul 10X Accuprime reaction buffer (Invitrogen)
- 1.5 ul 10 uM CP-2-FA (5'-GCACGATCCGACGGTAGTGT-3') (SEQ ID NO: 183)
-
- 1.5 ul 10 uM CP-2-RA (5'-CCGTAATCGGGAAGCTGAAG-3') (SEQ ID NO: 184)
-
- 0.4 ul 25 mM dNTP mix (Invitrogen)
- 2.0 ul heat-inactivated exonuclease reaction mix
- 1.5 ul 10X SybrGreen (Invitrogen)
- 0.4 ul 2.5 U/ul Accuprime Pfx polymerase (Invitrogen)

Petitioner Illumina, Inc., EX1004, p. 638

- 37.7 ul dH20

Thermal cycle in real-time thermal cycler according to the following protocol, but stop cycling before amplification yield plateaus (generally 8-12 cycles):

6. 95C for 5 min
7. 95C for 30 sec
8. 58C for 60 sec
9. 72C for 60 sec
10. goto 2, N more times

Prepare a shotgun next-generation sequencing library for analysis

- Purify desired amplicon population from non-specific amplification products by gel extraction.
- Concatemerize amplicons into high-molecular weight products suitable for shearing
- Mechanically shear, using either a nebulizer, BioRuptor, Hydroshear,
- Covaris, or similar instrument. DNA should be sheared into fragments several hundred basepairs in length.
- Ligate adapters required for amplification by the sequencing platform used. If necessary, purify ligated product from unligated product and adapters.

Perform sequencing of library according to manufacturer's directions (e.g. Illumina,ABI, etc), reading both the target sequence and the differentiator tag sequence.

Analyze data by correcting for any biases detected by quantitation of differentiator tag sequence abundance

Construct a table of target: differentiator tag abundances from the read data, e.g.:

Target ID	Differentiator Tag sequence ID	Count
1	3547	1
2	4762	1
1	9637	1
1	1078	5
3	4762	1
1	2984	1

Petitioner Illumina, Inc., EX1004, p. 639

All 'count' entries should be '1', since any particular target:differentiator tag mapping will not occur more than once by chance, and therefore will only be observed if bias was present somewhere in the sample preparation process. For any target:differentiator tag combination observed more than once, all such reads are 'collapsed' into a single read before consensus basecalls are determined. This will cancel the effect of bias on consensus basecall accuracy. FIG. 5 depicts a method for making diploid genotype calls in which repeat target:differentiator tag combination are collapsed.

Example 3: Differentiator tag sequence design for MIP capture reactions

For a set of targets, the number of differentiator tag sequences necessary to be confident (within some statistical bounds) that a certain differentiator tag sequence will not be observed more than once by chance in combination with a certain target sequence was determined. The total number of unique differentiator tag sequences for a certain differentiator tag sequence length is determined as $4^{(\text{Length in nucleotides of the differentiator tag sequence})}$. For a molecular inversion probe capture reaction that uses MIP probes having differentiator tag sequences, the probability of performing the capture reaction and capturing one or more copies of a target sequence having the same differentiator tag sequence is calculated as: $p = 1 - [N! / (N-M)!] / [NAM]$, wherein N is the total number of possible unique differentiator tag sequences and M is the number of target sequence copies in the capture reaction. Thus, by varying the differentiator tag sequence length it is possible to perform a MIP capture reaction in which the probability of capturing one or more copies of a target sequence having the same differentiator tag sequence is set at a predetermined probability value.

For example, for a differentiator tag sequence of 15 nucleotides in length, there are 1,073,741,824 possible differentiator tag sequences. A MIP capture reaction in which MIP probes, each having a differentiator tag sequence of 15 nucleotides, are combined with 10000 target sequence copies (e.g., genome equivalents), the probability of capturing one or more copies of a target sequence having the same differentiator tag sequence is 0.05. In this example, the MIP reaction will produce very few (usually 0, but occasionally 1 or more) targets where multiple copies are tagged with the same differentiator tag sequence. FIG. 6 depicts results of a simulation for 100000 capture reactions having 15 nucleotide differentiator tag sequences and 10000 target sequences.

Petitioner Illumina, Inc., EX1004, p. 640

Example 4: Assessment of the probability for obtaining enough sequencing reads to make accurate base-calls at multiple independent loci, as a function of sequencing coverage.

Monte Carlo simulations were performed to determine sequencing coverage requirements. The simulations assume 10000 genomic copies of a given locus (target) half mom alleles and half dad alleles. The simulations further assume 1% efficiency of capture for the MIP reaction. The simulation samples from a capture mix 100 times without replacement to create a set of 100 capture products. The simulation then samples from the set of 100 capture products with replacement (assuming unbiased amplification) to generate 'reads' from either mom or dad. The number of reads sampled depends on the coverage. The number of independent reads from both mom and dad necessary to make a high-quality base-call (assumed to be 10 or 20 reads) were then determined. The process was repeated 1000 times for each coverage level, and the fraction of times that enough reads from both parents were successfully obtained was determined. This fraction was raised to the power 1000, assuming we have 1000 independent loci that must obtain successful base-calls, plotted (See FIG. 7). Result show that roughly 50x coverage is required to capture each allele >= 10x with > 0.95 probability.

Example 5: MIP capture of 'target' locus and 'control' loci

In some embodiments, to accurately quantify the efficiency of target locus capture, at least three sets of control loci are captured in parallel that have a priori been shown to serve as proxies for various lengths of target locus. For example, if the target locus is expected to have a length between 50 and 1000 bp, then sets of control loci having lengths of 50, 250, and 1000 bp could be captured (e.g. 20 loci per set should provide adequate protection from outliers), and their abundance digitally measured by sequencing. These loci should be chosen such that minimal variation in efficiency between samples and on multiple runs of the same sample is observed (and are therefore 'efficiency invariant'). These will serve as 'reference' points that define the shape of the curve of abundance-vs-length. Determining the length of the target is then simply a matter of 'reading' the length from the appropriate point on the calibration curve.

In some embodiments, the statistical confidence one has in the estimate of target length from this method is driven largely by three factors: 1) reproducibility / variation of the abundance data used to generate the calibration curve; 2) goodness of fit of the regression to the 'control' datapoints; 3) reproducibility of abundance data for the target locus being measured. Statistical bounds on 1) and 2) will be known in advance, having been measured during

Petitioner Illumina, Inc., EX1004, p. 641

development of the assay. Additionally, statistical bounds on 3) will be known in general in advance, since assay development should include adequate population sampling and measure of technical reproducibility. Standard statistical methods should be used to combine these three measures into a single P value for any given experimental measure of target abundance.

In some embodiments, given the set of calibration observations, and a linear regression fit to that data, the regression can be used to predict the length value for n observations of the target locus whose length is unknown. First, choose an acceptable range for the confidence interval of the length estimate. For example, in the case of distinguishing "normal" (87-93bp) from "premutation" (165-600bp) potential cases of Fragile X, the goal is to measure length to sufficient precision to distinguish 93 bp from 165 bp. The predicted response value, computed when n observations is substituted into the equation for the regressed line, will have arbitrary precision. However, if for example a 95% confidence level is desired, that 95% confidence interval must be sufficiently short that it does not overlap both the "normal" and "premutation" length ranges. Continuing the example, if one calculates a length of 190 from n=400 MIP observations, and based on the regression from calibration data, the 95% confidence interval is 190 +/- 20 bp, one can conclude the sample represents a "premutation" length with 95% certainty. Conversely, if the calibration data were less robust, error estimates of the regression would be higher, leading to larger confidence intervals on the predicted response value. In some embodiments, if the 95% CI were calculated as 190 +/- 100 bp from n=400, one could not determine whether the predicted response value corresponds to a "normal" or "premutation" length.

In some embodiments, the confidence interval for a predicted response is calculated as:

The estimate for the response $\hat{y}$ *is identical to the estimate for the mean of the response:* $\hat{y} = b_0 + b_1 x^*$. *The confidence interval for the predicted value is given by* $\hat{y} \pm t^* s_{\hat{y}}$, *where* $\hat{y}$ *is the fitted value corresponding to* x^*. *The value* t^* *is the upper (1 - C)/2 critical value for the t(n - 2) distribution.*

In some embodiments, a technique for analyzing a locus of interest can involve the following steps.

Petitioner Illumina, Inc., EX1004, p. 642

Convert hybridized probes into covalently-closed circular products containing subtargets

Prepare fill-in / ligation reaction mixture:

- 0.25 ul 2 mM dNTP mix (Invitrogen)
- 2.5 ul 10X Ampligase buffer (Epicentre)
- 5.0 ul 5 U/ul Taq Stoffel fragment (Applied Biosystems)
- 12.5 ul 5 U/ul Ampligase (Epicentre)
- 4.75 ul dH20

Add 1.0 ul of this mix to the hybridized probe reaction, and incubate at 60C for 10 hours.

Purify circularized probe / subtarget products from un-reacted probes and 30 genomic DNA

Prepare exonuclease reaction mixture:

- 21 ul fill-in / ligation reaction product
- 2.0 ul 10X exonuclease I buffer (New England Biolabs)
- 2.0 ul 20 U/ul exonuclease I (New England Biolabs)
- 2.0 ul 100 U/ul exonuclease III (New England Biolabs)

Incubate at 37C for 60 min, then heat-inactivate by incubating at 80C for 15 min. Immediately cool to 4C for storage.

Amplify circular material by PCR using primers specific to the 'constant region' of the probes

Prepare PCR mixture:

- 5.0 ul 10X Accuprime reaction buffer (Invitrogen)
- 1.5 ul 10 uM CP-2-FA-Ilmn (platform-specific amplification sequence plus 'circle constant region'-specific sequence)
- 1.5 ul 10 uM CP-2-RA-Ilmn (platform-specific amplification sequence plus 'circle constant region'-specific sequence)
- 0.4 ul 25 mM dNTP mix (Invitrogen)
- 2.0 ul heat-inactivated exonuclease reaction mix
- 1.5 ul 10X SybrGreen (Invitrogen)
- 0.4 ul 2.5 U/ul Accuprime Pfx polymerase (Invitrogen)

Petitioner Illumina, Inc., EX1004, p. 643

- 37.7 ul dH20

Thermal cycle in real-time thermal cycler according to the following protocol, but stop cycling before amplification yield plateaus (generally 8-12 cycles):

11. 95C for 5 min
12. 95C for 30 sec
13. 58C for 60 sec
14. 72C for 60 sec
15. goto 2, N more times

Perform sequencing (e.g., next-generation sequencing) on sample for digital quantitation according to manufacturer's instructions (e.g., Illumina, ABI)

Example 6: MIP-capture reaction of a set of exon target nucleic acids

MIP probes are synthesized using standard column-based oligonucleotide synthesis by any number of vendors (e.g. IDT).

Hybridize capture probes to human genomic sample

Assemble hybridization reaction:

- 1.0 ul capture probe mix (-2.5 pmol)
- 2.0 ul 10X Ampligase buffer (Epicentre)
- 6.0 ul 500 ng/ul human genomic DNA (-16.7 fmol)
- 11 ul dH20

In a thermal cycler, heat reaction to 95C for 5 min to denature genomic DNA, then cool to 60C. Allow to incubate at 60C for 40 hours.

Convert hybridized probes into covalently-closed circular products containing target nucleic acids

Prepare fill-in / ligation reaction mixture:

- 0.25 ul 2 mM dNTP mix (Invitrogen)
- 2.5 ul 10X Ampligase buffer (Epicentre)
- 5.0 ul 5 U/ul Taq Stoffel fragment (Applied Biosystems)
- 12.5 ul 5 U/ul Ampligase (Epicentre)
- 4.75 ul dH20

Add 1.0 ul of this mix to the hybridized probe reaction, and incubate at 60C for 10 hours.

Petitioner Illumina, Inc., EX1004, p. 644

Purify, circularized probe/target nucleic acid products from un-reacted probes and genomic DNA

Prepare exonuclease reaction mixture:

- 21 ul fill-in / ligation reaction product
- 2.0 ul 10X exonuclease I buffer (New England Biolabs)
- 2.0 ul 20 U/ul exonuclease I (New England Biolabs)
- 2.0 ul 100 U/ul exonuclease III (New England Biolabs)

Incubate at 37C for 60 min, then heat-inactivate by incubating at 80C for 15 min. Immediately cool to 4C for storage.

Amplify circular material by PCR using primers specific to the 'constant region' of the probes

Prepare PCR mixture:

- 5.0 ul 10X Accuprime reaction buffer (Invitrogen)
- 1.5 ul 10 uM CP-2-FA (5'-GCACGATCCGACGGTAGTGT-3') (SEQ ID NO: 183)
-
- 1.5 ul 10 uM CP-2-RA (5'-CCGTAATCGGGAAGCTGAAG-3') (SEQ ID NO: 184)
- 0.4 ul 25 mM dNTP mix (Invitrogen)
- 2.0 ul heat-inactivated exonuclease reaction mix
- 1.5 ul 10X SybrGreen (Invitrogen)
- 0.4 ul 2.5 U/ul Accuprime Pfx polymerase (Invitrogen)
- 37.7 ul dH20

Thermal cycle in real-time thermal cycler according to the following protocol, but stop cycling before amplification yield plateaus (generally 8-12 cycles):

16. 95C for 5 min
17. 95C for 30 sec
18. 58C for 60 sec
19. 72C for 60 sec
20. goto 2, N more times

Petitioner Illumina, Inc., EX1004, p. 645

Prepare a shotgun next-generation sequencing library for analysis

- Purify desired amplicon population from non-specific amplification products by gel extraction.
- Concatemerize amplicons into high-molecular weight products suitable for shearing
- Mechanically shear, using either a nebulizer, BioRuptor, Hydroshear, Covaris, or similar instrument. DNA should be sheared into fragments several hundred basepairs in length.
- Ligate adapters required for amplification by the sequencing platform used. If necessary, purify ligated product from unligated product and adapters.

Perform sequencing of library according to manufacturer's directions (e.g. Illumina, ABI, etc), reading the target sequence to determine abundance of the target nucleic acid.

Example 7: Use of MIPs, hybridization, and mutation-detection MIPs to genotype a set of 1000 targets

MIPs, hybridization, and mutation-detection MIPs are used to genotype a set of 1000 targets. The protocol permits detection of any of 50 specific known point mutations

First, separate MIP, hybridization, and mutation-detection MIP reactions are performed on a biological sample. A MIP capture reaction is performed essentially as described in Turner et al 2009, Nature methods 6:315-6. A set of MIPs is designed such that each probe in the set flanks one of the 1000 targets. Separately, a hybridization enrichment reaction is performed using the Agilent SureSelect procedure. Prior to selection, the genomic DNA to be enriched is converted into a shotgun sequencing library using Illumina's 'Fragment Library' kit and protocol. Agilent's web interface is used to design a set of probes which will hybridize to the target nucleic acids. Separately, a set of probes are designed (mutation-detection MIPs) which will form MIPs only if mutations (e.g., specific polymorphisms) are present. Each mutation-detection MIP has a 3'-most base identity that is specific for a single known mutation. A reaction with this set of mutation-detection MIPs is performed to selectively detect the presence of any mutant alleles.

Once all three reactions have been performed, the two MIP reactions are combined (e.g., at potentially non-equimolar ratios to further increase sensitivity of mutation detection) into a

Petitioner Illumina, Inc., EX1004, p. 646

single tube, and run as one sample on the next-generation DNA sequencing instrument. The hybridization-enriched reaction is run as a separate sample on the next-generation DNA sequencing instrument. Reads from each 'sample' are combined by a software algorithm which forms a consensus diploid genotype at each position in the target set by evaluating the total coverage at each position, the origin of each read in that total coverage, the quality score of each individual read, and the presence (or absence) of any reads derived from mutation-specific MIPs overlapping the region.

It should be appreciated that the preceding examples are non-limiting and aspects of the invention may be implemented as described herein using alternative techniques and/or protocols that are available to one or ordinary skill in the art.

It will be clear that the methods may be practiced other than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present disclosure are possible in light of the above teachings and, therefore, are within the scope of the claims. Preferred features of each aspect of the Disclosure are as for each of the other aspects mutatis mutandis. The documents including patents, patent applications, journal articles, or other disclosures mentioned herein are hereby incorporated by reference in their entirety. In the event of conflict, the disclosure of present application controls, other than in the event of clear error.

METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS

RELATED APPLICATIONS

This application is a continuation of U.S. Application No. 15/231,687, filed August 8, 2016, which is a continuation of U.S. Application No. 13/266,862, filed March 13, 2012, which is a national phase application that claims the benefit of, and priority to, international PCT Application No. PCT/US2010/001293, filed April 30, 2010, which claims the benefit under 35 U.S.C. §119(e) of U.S. Provisional Application No. 61/174,470, filed April 30, 2009, U.S. Provisional Application No. 61/178,923, filed May 15, 2009, U.S. Provisional Application No. 61/179,358, filed May 18, 2009, and U.S. Provisional Application No. 61/182,089, filed May 28, 2009, the entire contents of each of which are incorporated herein by reference.

FIELD OF INVENTION

The invention relates to methods and compositions for determining genotypes in patient samples.

BACKGROUND OF THE INVENTION

Information about the genotype of a subject is becoming more important and relevant for a range of healthcare decisions as the genetic basis for many diseases, disorders, and physiological characteristics is further elucidated. Medical advice is increasingly personalized, with individual decisions and recommendations being based on specific genetic information. Information about the type and number of alleles at one or more genetic loci impacts disease risk, prognosis, therapeutic options, and genetic counseling amongst other healthcare considerations.

For cost-effective and reliable medical and reproductive counseling on a large scale, it is important to be able correctly and unambiguously identify the allelic status for many different genetic loci in many subjects.

Numerous technologies have been developed for detecting and analyzing nucleic acid sequences from biological samples. These technologies can be used to genotype subjects and determine the allelic status of any locus of interest. However, they are not sufficiently robust and

Petitioner Illumina, Inc., EX1004, p. 648

cost-effective to be scaled up for reliable high throughput analysis of many genetic loci in large numbers of patients. The frequency of incorrect or ambiguous calls is too high for current technology to manage large numbers of patient samples without involving expensive and time-consuming steps to resolve uncertainties and provide confidence in the information output.

SUMMARY OF THE INVENTION

Aspects of the invention relate to preparative and analytical methods and compositions for evaluating genotypes, and in particular, for determining the allelic identity (or identities in a diploid organism) of one or more genetic loci in a subject.

Aspects of the invention are based, in part, on the identification of different sources of ambiguity and error in genetic analyses, and, in part, on the identification of one or more approaches to avoid, reduce, recognize, and/or resolve these errors and ambiguities at different stages in a genetic analysis.

According to aspects of the invention, certain types of genetic information can be under-represented or over-represented in a genetic analysis due to a combination of stochastic variation and systematic bias in any of the preparative stages (e.g., capture, amplification, etc.), determining stages (e.g., allele-specific detection, sequencing, etc.), data interpretation stages (e.g., determining whether the assay information is sufficient to identify a subject as homozygous or heterozygous), and/or other stages.

According to aspects of the invention, error or ambiguity may be apparent in a genetic analysis, but not readily resolved without running additional samples or more expensive assays (*e.g.*, array-based assays may report no-calls due to noisy/low signal). According to further aspects of the invention, error or ambiguity may not be accounted for in a genetic analysis and incorrect base calls may be made even when the evidence for them is limited and/or not statistically significant (*e.g.*, next-generation sequencing technologies may report base calls even if the evidence for them is not statistically significant). According to further aspects of the invention error or ambiguity may be problematic for a multi-step genetic analysis because it is apparent but not readily resolved in one or more steps of the analysis and not apparent or accounted for in other steps of the analysis.

In some embodiments, sources of error and ambiguity in one or more steps can be addressed by capturing and/or interrogating each target locus of interest with one or more sets of

Petitioner Illumina, Inc., EX1004, p. 649

overlapping probes that are designed to overcome any systematic bias or stochastic effects that may impact the complexity and/or fidelity of the genetic information that is generated.

In some embodiments, sources of error and ambiguity in one or more steps can be addressed by capturing and/or interrogating each target locus of interest with at least one set of probes, wherein different probes are labeled with different identifiers that can be used to track the assay reactions and determine whether certain types of genetic information are under-represented or over-represented in the information that is generated.

In some embodiments, errors and ambiguities associated with the analysis of regions containing large numbers of sequence repeats are addressed by systematically analyzing frequencies of certain nucleic acids at particular stages in an assay (e.g., at a capture, sequencing, or detection stage). It should be appreciated that such techniques may be particularly useful in the context of a standardized protocol that is designed to allow many different loci to be evaluated in parallel without requiring different assay procedures for each locus. In some embodiments, the use of a single detection modality (*e.g.*, sequencing) to assay multiple types of genetic lesions (*e.g.*, point mutations, insertions/deletions, length polymorphisms) is advantageous in the clinical setting. In some embodiments of the invention, methods are provided that facilitate the use of multiple sample preparation steps in parallel, coupled with multiple analytical processes following sequence detection. Thus, in some embodiments of the invention, an improved workflow is provided that reduces error and uncertainty when simultaneously assaying different types of genetic lesions across multiple loci in multiple patients.

In some embodiments, aspects of the invention provide methods for overcoming preparative and/or analytical bias by combining two or more techniques, each having a different bias (e.g., a known bias towards under-representation or over-representation of one or more types of sequences), and using the resulting data to determine a genetic call for a subject with greater confidence.

It should be appreciated that in some embodiments, aspects of the invention relate to multiplex diagnostic methods. In some embodiments, multiplex diagnostic methods comprise capturing a plurality of genetic loci in parallel (e.g., one or more genetic loci from Table 1). In some embodiments, the genetic loci possess one or more polymorphisms (e.g., one or more polymorphisms from Table 2) the genotypes of which correspond to disease causing alleles.

Petitioner Illumina, Inc., EX1004, p. 650

Accordingly, in some embodiments, the disclosure provides methods for assessing multiple heritable disorders in parallel. In some embodiments, methods are provided for diagnosing multiple heritable disorders in parallel at a pre-implantation, prenatal, perinatal, or postnatal stage. In some embodiments, the disclosure provides methods for analyzing multiple genetic loci (e.g., a plurality of target nucleic acids selected from Table 1) from a patient sample, such as a blood, pre-implantation embryo, chorionic villus or amniotic fluid sample, or other sample (e.g., other biological fluid or tissue sample such as a biopsy sample) as aspects of the invention are not limited in this respect.

Other samples may include tumor tissue or circulating tumor cells. In some embodiments, a patient sample (*e.g.*, a tumor tissue or cell sample) is mosaic for one or more mutations of interest, and thus, may require higher sensitivity than is needed for a germline mutation analysis. In some embodiments, a sample comprises cells from a non-host organism (*e.g.*, bacterial or viral infections in a human subject) or a sample for environmental monitoring (*e.g.*, bacterial, viral, fungal composition of a soil, water, or air sample).

Accordingly, in some embodiments, aspects of the methods disclosed herein relate to genotyping a polymorphism of a target nucleic acid. In some embodiments, the genotyping may comprise determining that one or more alleles of the target nucleic acid are heterozygous or homozygous. In further embodiments, the genotyping may comprise determining the sequence of a polymorphism and comparing that sequence to a control sequence that is indicative of a disease risk. In some embodiments, the polymorphism is selected from a locus in Table 1 or Table 2. However, it should be appreciated that any locus associated with a disease or condition of interest may be used.

In some embodiments, a diagnosis, prognosis, or disease risk assessment is provided to a subject based on a genotype determined for that subject at one or more genetic loci (e.g., based on the analysis of a biological sample obtained from that subject). In some embodiments, an assessment is provided to a couple, based on their respective genotypes at one or more genetic loci, of the risk of their having one or more children having a genotype associated with a disease or condition (e.g., a homozygous or heterozygous genotype associated with a disease or condition). In some embodiments, a subject or a couple may seek genetic or reproductive counseling in connection with a genotype determined according to embodiments of the

Petitioner Illumina, Inc., EX1004, p. 651

invention. In some embodiments, genetic information from a tumor or circulating tumor cells is used to determine prognosis and guide selection of appropriate drugs/treatments.

It should be appreciated that any of the methods or compositions described herein may be used in combination with any of the medical evaluations associated with one or more genetic loci as described herein.

In some embodiments, aspects of the invention provide effective methods for overcoming challenges associated with systematic errors (bias) and/or stochastic effects in multiplex genomic capture and/or analysis (including sequencing analysis). In some embodiments, aspects of the invention are useful to avoid, reduce and/or account for variability in one or more sampling and/or analytical steps. For example, in some embodiments, variability in target nucleic acid representation and unequal sampling of heterozygous alleles in pools of captured target nucleic acids can be overcome.

Accordingly, in some embodiments, the disclosure provides methods that reduce variability in the detection of target nucleic acids in multiplex capture methods. In other embodiments, methods improve allelic representation in a capture pool and, thus, improve variant detection outcomes. In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve the use of different sets of multiple probes (e.g., molecular inversion probes MIPs) that capture overlapping regions of a target nucleic acid to achieve a more uniform representation of the target nucleic acids in a capture pool compared with methods of the prior art. In other embodiments, methods reduce bias, or the risk of bias, associated with large scale parallel capture of genetic loci, e.g., for diagnostic purposes. In other embodiments, methods are provided for increasing reproducibility (e.g., by reducing the effect of polymorphisms on target nucleic acid capture) in the detection of a plurality of genetic loci in parallel. In further embodiments, methods are provided for reducing the effect of probe synthesis and/or probe amplification variability on the analysis of a plurality of genetic loci in parallel.

According to some aspects, methods of analyzing a plurality of genetic loci are provided. In some embodiments, the methods comprise contacting each of a plurality of target nucleic acids with a probe set, wherein each probe set comprises a plurality of different probes, each probe having a central region flanked by a 5' region and a 3' region that are complementary to nucleic acids flanking the same strand of one of a plurality of subregions of the target nucleic

Petitioner Illumina, Inc., EX1004, p. 652

acid, wherein the subregions of the target nucleic acid are different, and wherein each subregion overlaps with at least one other subregion, isolating a plurality of nucleic acids each having a nucleic acid sequence of a different subregion for each of the plurality of target nucleic acids, and analyzing the isolated nucleic acids.

In other embodiments, methods comprise contacting each of a plurality of target nucleic acids with a probe set, wherein each probe set comprises a plurality of different probes, each probe having a central region flanked by a 5' region and a 3' region that are complementary to nucleic acids flanking the same strand of one of a plurality of subregions of the target nucleic acid, wherein the subregions of the target nucleic acid are different, and wherein a portion of the 5' region and a portion of the 3' region of a probe have, respectively, the sequence of the 5' region and the sequence of the 3' region of a different probe, isolating a plurality of nucleic acids each having a nucleic acid sequence of a different subregion for each of the plurality of target nucleic acids, and analyzing the isolated nucleic acids.

Aspects of the disclosure are based, in part, on the discovery of methods for overcoming problems associated with systematic and random errors (bias) in genome capture, amplification and sequencing methods, namely high variability in the capture and amplification of nucleic acids and disproportionate representation of heterozygous alleles in sequencing libraries. Accordingly, in some embodiments, the disclosure provides methods that reduce errors associated with the variability in the capture and amplification of nucleic acids. In other embodiments, the methods improve allelic representation in sequencing libraries and, thus, improve variant detection outcomes. In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve the use of differentiator tag sequences to uniquely tag individual nucleic acid molecules. In some embodiments, the differentiator tag sequence permit the detection of bias based on the occurrence of combinations of differentiator tag and target sequences observed in a sequencing reaction. In other embodiments, the methods reduce errors caused by bias, or the risk of bias, associated with the capture, amplification and sequencing of genetic loci, e.g., for diagnostic purposes.

Aspects of the invention relate to providing sequence tags (referred to as differentiator tags) that are useful to determine whether target nucleic acid sequences identified in an assay are from independently isolated target nucleic acids or from multiple copies of the same target

Petitioner Illumina, Inc., EX1004, p. 653

nucleic acid molecule (e.g., due to bias in a preparative step, for example, amplification). This information can be used to help analyze a threshold number of independently isolated target nucleic acids from a biological sample in order to obtain sequence information that is reliable and can be used to make a genotype conclusion (e.g., call) with a desired degree of confidence. This information also can be used to detect bias in one or more nucleic acid preparative steps.

In some embodiments, the methods disclosed herein are useful for any application where reduction of bias, e.g., associated with genomic isolation, amplification, sequencing, is important. For example, detection of cancer mutations in a heterogeneous tissue sample, detection of mutations in maternally-circulating fetal DNA, and detection of mutations in cells isolated during a preimplantation genetic diagnostic procedure.

Accordingly, in some aspects, methods of genotyping a subject are provided. In some embodiments, the methods comprise determining the sequence of at least a threshold number of independently isolated nucleic acids, wherein the sequence of each isolated nucleic acid comprises a target nucleic acid sequence and a differentiator tag sequence, wherein the threshold number is a number of unique combinations of target nucleic acid and differentiator tag sequences, wherein the isolated nucleic acids are identified as independently isolated if they comprise unique combinations of target nucleic acid and differentiator tag sequences, and wherein the target nucleic acid sequence is the sequence of a genomic locus of a subject.

In some embodiments, the isolated nucleic acids are products of a circularization selection-based preparative method, e.g., molecular inversion probe capture products. In other embodiments, the isolated nucleic acids are products of an amplification-based preparative methods. In other embodiments, the isolated nucleic acids are products of hybridization-based preparative methods.

Circularization selection-based preparative methods selectively convert regions of interest (target nucleic acids) into a covalently-closed circular molecule which is then isolated typically by removal (usually enzymatic, e.g. with exonuclease) of any non-circularized linear nucleic acid. Oligonucleotide probes (e.g., molecular inversion probes) are designed which have ends that flank the region of interest (target nucleic acid) and, optionally, primer sites, e.g., sequencing primer sites. The probes are allowed to hybridize to the genomic target, and enzymes are used to first (optionally) fill in any gap between probe ends and second ligate the probe closed. Following circularization, any remaining (non-target) linear nucleic acid is

Petitioner Illumina, Inc., EX1004, p. 654

typically removed, resulting in isolation (capture) of target nucleic acid. Circularization selection-based preparative methods include molecular inversion probe capture reactions and 'selector' capture reactions. In some embodiments, molecular inversion probe capture of a target nucleic acid is indicative of the presence of a polymorphism in the target nucleic acid.

In amplification-based (e.g., PCR-based or LCR-based, etc.) preparative methods, genomic loci (target nucleic acids) are isolated directly by means of a polymerase chain reaction or ligase chain reaction (or other amplification method) that selectively amplifies each locus using one or more oligonucleotide primers. It is to be understood that primers will be sufficiently complementary to the target sequence to hybridize with and prime amplification of the target nucleic acid. Any one of a variety of art known methods may be utilized for primer design and synthesis. One or more of the primers may be perfectly complementary to the target sequence. Degenerate primers may also be used. Primers may also include additional nucleic acids that are not complementary to target sequences but that facilitate downstream applications, including for example restriction sites and differentiator tag sequences. Amplification-based methods include amplification of a single target nucleic acid and multiplex amplification (amplification of multiple target nucleic acids in parallel).

Hybridization-based preparative methods involve selectively immobilizing target nucleic acids for further manipulation. It is to be understood that one or more oligonucleotides (immobilization oligonucleotides), which comprise differentiator tag sequences, and which may be from 15 to 170 nucleotides in length, are used which hybridize along the length of a target region of a genetic locus to immobilize it. In some embodiments, immobilization oligonucleotides, are either immobilized before hybridization is performed (e.g., Roche/Nimblegen 'sequence capture'), or are prepared such that they include a moiety (e.g. biotin) which can be used to selectively immobilize the target nucleic acid after hybridization by binding to e.g., streptavidin-coated microbeads (e.g. Agilent 'SureSelect').

It should be appreciated that any of the circularization, amplification, and/or hybridization based methods described herein may be used in connection with one or more of the tiling/staggering, tagging, size-detection, and/or sensitivity enhancing algorithms described herein.

In some embodiments, the methods disclosed herein comprise determining the sequence of molecular inversion probe capture products, each comprising a molecular inversion probe and

Petitioner Illumina, Inc., EX1004, p. 655

a target nucleic acid, wherein the sequence of the molecular inversion probe comprises a differentiator tag sequence and, optionally, a primer sequence, and wherein the target nucleic acid is a captured genomic locus of a subject, and genotyping the subject at the captured genomic locus based on the sequence of at least a threshold number of unique combinations of target nucleic acid and differentiator tag sequences of molecular inversion probe capture products.

In some embodiments, the methods disclosed herein comprise obtaining molecular inversion probe capture products, each comprising a molecular inversion probe and a target nucleic acid, wherein the sequence of the molecular inversion probe comprises a differentiator tag sequence and, optionally, a primer sequence, wherein the target nucleic acid is a captured genomic locus of the subject, amplifying the molecular inversion probe capture products, and genotyping the subject by determining, for each target nucleic acid, the sequence of at least a threshold number of unique combinations of target nucleic acid and differentiator tag sequence of molecular inversion probe capture products. In certain embodiments, obtaining comprises capturing target nucleic acids from a genomic sample of the subject with molecular inversion probes, each comprising a unique differentiator tag sequence. In specific embodiments, capturing is performed under conditions wherein the likelihood of obtaining two or more molecular inversion probe capture products with identical combinations of target and differentiator tag sequences is equal to or less than a predetermined value, optionally wherein the predetermined value is about 0.05.

In one embodiment, the threshold number for a specific target nucleic acid sequence is selected based on a desired statistical confidence for the genotype. In some embodiments, the methods further comprising determining a statistical confidence for the genotype based on the number of unique combinations of target nucleic acid and differentiator tag sequences.

According to some aspects, methods of analyzing a plurality of genetic loci are provided. In some embodiments, the methods comprise obtaining a plurality of molecular inversion probe capture products each comprising a molecular inversion probe and a target nucleic acid, wherein the sequence of the molecular inversion probe comprises a differentiator tag sequence and, optionally, a primer sequence (e.g., a sequence that is complementary to the sequence of a nucleic acid that is used as a primer for sequencing or other extension reaction), amplifying the plurality of molecular inversion probe capture products, determining numbers of occurrence of combinations of target nucleic acid and differentiator tag sequence of molecular inversion probe

Petitioner Illumina, Inc., EX1004, p. 656

capture products in the amplified plurality, and if the number of occurrence of a specific combination of target nucleic acid sequence and differentiator tag sequence exceeds a predetermined value, detecting bias in the amplification of the molecular inversion probe comprising the specific combination. In some embodiments, the methods further comprise genotyping target sequences in the plurality, wherein the genotyping comprises correcting for bias, if detected.

In some embodiments, the target nucleic acid is a gene (or portion thereof) selected from Table 1. In some embodiments, the genotyping comprises determining the sequence of a target nucleic acid (e.g., a polymorphic sequence) at one or more (both) alleles of a genome (a diploid genome) of a subject. In certain embodiments, the genotyping comprises determining the sequence of a target nucleic acid at both alleles of a diploid genome of a subject, wherein in the target nucleic acid comprises, or consists of, a sequence of Table 1, Table 2, or other locus of interest.

In some embodiments, aspects of the invention provide methods and compositions for identifying nucleic acid insertions or deletions in genomic regions of interest without determining the nucleotide sequences of these regions. Aspects of the invention are particularly useful for detecting nucleic acid insertions or deletions in genomic regions containing nucleic acid sequence repeats (e.g., di- or tri-nucleotide repeats). However, the invention is not limited to analyzing nucleic acid repeats and may be used to detect insertions or deletions in any target nucleic acid of interest. Aspects of the invention are particularly useful for analyzing multiple loci in a multiplex assay.

In some embodiments, aspects of the invention relate to determining whether an amount of target nucleic acid that is captured in a genomic capture assay is higher or lower than expected. In some embodiments, a statistically significant deviation from an expected amount (e.g., higher or lower) is indicative of the presence of a nucleic acid insertion or deletion in the genomic region of interest. In some embodiments, the amount is a number of nucleic acid molecules that are captured. In some embodiments, the amount is a number of independently captured nucleic acid molecules in a sample. It should be appreciated that the captured nucleic acids may be literally captured from a sample, or their sequences may be captured without actually capturing the original nucleic acids in the sample. For example, nucleic acid sequences

Petitioner Illumina, Inc., EX1004, p. 657

may be captured in an assay that involves a template-based extension of nucleic acids having the region of interest, in the sample.

Aspects of the invention are based on the recognition that the efficiency of certain capture techniques is affected by the length of the nucleic acid being captured.

Accordingly, an increase or decrease in the length of a target nucleic acid (e.g., due to an insertion or deletion of a repeated sequence) can alter the capture efficiency of that nucleic acid. In some embodiments, a difference in the capture efficiency (e.g., a statistically significant difference in the capture efficiency) of a target nucleic acid is indicative of an insertion or deletion in the target nucleic acid. It should be appreciated that the capture efficiency for a target nucleic acid may be evaluated based on an amount of captured nucleic acid (e.g., number of captured nucleic acid molecules) relative to a control amount (e.g., based on an amount of control nucleic acid that is captured). However, the invention is not limited in this respect and other techniques for evaluating capture efficiency also may be used.

According to aspects of the invention, evaluating the capture efficiency as opposed to determining the sequence of the entire repeat region reduces errors associated with sequencing through repeat regions. Repeat sequences often give rise to stutters or skips in sequencing reactions that make it very difficult to accurately determine the number of repeats in a target region without running multiple sequencing reactions under different conditions and carefully analyzing the results. Such procedures are cumbersome and not readily scalable in a manner that is consistent with high throughput analyses of target nucleic acids. In some embodiments, repeat regions may be longer than the length of the individual sequence read, making length determination on the basis of a single read impossible. For example, when using next-generation sequencing the repeat regions may be longer than the length of the individual sequence read, making length determination on the basis of a single read impossible. Accordingly, aspects of the invention are useful to increase the sensitivity of detecting insertions or deletions in target regions, particularly target regions containing repeated sequences.

In some embodiments, aspects of the invention relate to capturing genomic nucleic acid sequences using a molecular inversion probe (e.g., MIP or Padlock probe) technique, and determining whether the amount (e.g., number) of captured sequences is higher or lower than expected. In some embodiments, the amount (e.g., number) of captured sequences is compared to an amount (e.g., number) of sequences captured in a control assay. The control assay may

Petitioner Illumina, Inc., EX1004, p. 658

involve analyzing a control sample that contains a nucleic acid from the same genetic locus having a known sequence length (e.g., a known number of nucleic acid repeats). However, a control may involve analyzing a second (e.g., different) genetic locus that is not expected to contain any insertions or deletions. The second genetic locus may be analyzed in the same sample as the locus being interrogated or in a different sample where its length has been previously determined. The second genetic locus may be a locus that is not characterized by the presence of nucleic acid repeats (and thus not expected to contain insertions or deletions of the repeat sequence).

In some embodiments, a target nucleic acid region that is being evaluated may be determined by the identity of the targeting arms of a probe that is designed to capture the target region (or sequence thereof). For example, the targeting arms of a MIP probe may be designed to be complementary (e.g., sufficiently complementary for selective hybridization and//or polymerase extension and/or ligation) to genomic regions flanking a target region suspected of containing an insertion or deletion. It should be appreciated that two targeting arms may be designed to be complementary (e.g., sufficiently complementary for selective hybridization and/or polymerase extension and/or ligation) to the two flanking regions that are immediately adjacent (e.g., immediately 5' and 3', respectively) to a region of a sequence repeat on one strand of a genomic nucleic acid. However, one or both targeting arms may be designed to hybridize several bases (e.g., 1-5, 5-10, 10-25, 25-50, or more) upstream or downstream from the repeat region in such a way that the captured sequence includes a region of unique genomic sequence that on one or both sides of the repeat region. This unique region can then be used to identify the captured target (e.g., based on sequence or hybridization information).

In some embodiments, two or more (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10 or more) different loci may be interrogated in parallel in a single assay (e.g., in a multiplex assay). In some embodiments, the ratio of captured nucleic acids for each locus may be used to determine whether a nucleic acid insertion or deletion is present in one locus relative to the other. For example, the ratio may be compared to a control ratio that is representative of the two loci when neither one has an insertion or deletion relative to control sequences (e.g., sequences that are normal or known to be associated with healthy phenotypes for those loci). However, the amount of captured nucleic acids may be compared to any suitable control as discussed herein.

Petitioner Illumina, Inc., EX1004, p. 659

The locus of a captured sequence may be identified by determining a portion of unique sequence 5' and/or 3' to the repeat region in the target nucleic acid suspected of containing a deletion or insertion. This does not require sequencing the captured repeat region itself. However, some or all of the repeat region also could be sequenced as aspects of the invention are not limited in this respect.

Aspects of the invention may be combined with one or more sequence-based assays (e.g., SNP detection assays), for example in a multiplex format, to determine the genotype of one or more regions of a subject.

In some embodiments, methods of detecting a polymorphism in a nucleic acid in a biological sample are provided. In some embodiments, the methods comprise evaluating the efficiency of capture at one or more loci and determining whether one or both alleles at that locus contain an insertion or deletion relative to a control locus (e.g., a locus indicative of a length of repeat sequence that is associated with a healthy phenotype).

Accordingly, aspects of the invention relate to methods for determining whether a target nucleic acid has an abnormal length by evaluating the capture efficiency of a target nucleic acid in a biological sample from a subject, wherein a capture efficiency that is different from a reference capture efficiency is indicative of the presence, in the biological sample, of a target nucleic acid having an abnormal length. It should be appreciated that the term "abnormal" is a relative term based on a comparison to a "normal" length. In some embodiments, a normal length is a length that is associated with a normal (e.g., healthy or non-carrier phenotype). Accordingly, an abnormal length is a length that is either shorter or longer than the normal length. In some embodiments, the presence of an abnormal length is indicative of an increased risk that the locus is associated with a disease or a disease carrier phenotype. In some embodiments, the abnormal length is indicative that the subject is either has a disease or condition or is a carrier of a disease or condition (e.g., associated with the locus). However, it should be appreciated that the description of embodiments relating to detecting the presence of an abnormal length also support detecting the presence of a length that is different from an expected or control length.

In some embodiments, aspects of the invention relate to estimating the length of a target nucleic acid (e.g., of a sub-target region within a target nucleic acid). In some embodiments, aspects of the invention relate to methods for estimating the length of a target nucleic acid by

Petitioner Illumina, Inc., EX1004, p. 660

contacting the target nucleic acid with a plurality of detection probes under conditions that permit hybridization of the detection probes to the target nucleic acid, wherein each detection probe is a polynucleotide that comprises a first arm that hybridizes to a first region of the target nucleic acid and a second arm that hybridizes to a second region of the target nucleic acid, wherein the first and second regions are on a common strand of the target nucleic acid, and wherein the nucleotide sequence of the target between the 5' end of the first region and the 3' end of the second region is the nucleotide sequence of a sub-target nucleic acid; and capturing a plurality of sub-target nucleic acids that are hybridized with the plurality of detection probes; and measuring the frequency of occurrence of a sub-target nucleic acid in the plurality of sub-target nucleic acids, wherein the frequency of occurrence of the sub-target nucleic acid in the plurality of sub-target nucleic acids is indicative of the length of the sub-target nucleic acid. It should be appreciated that methods for estimating a nucleic acid length may involve comparing a capture efficiency for a target nucleic acid region to two or more reference efficiencies for known nucleic acid lengths in order to determine whether the target nucleic acid region is smaller, intermediate, or larger in size than the known control lengths. In some embodiments, a series of nucleic acids of known different lengths may be used to provide a calibration curve for evaluating the length of a target nucleic acid region of interest.

In some embodiments, the capture efficiency of a target region suspected of having a deletion or insertion is determined by comparing the capture efficiency to a reference indicative of a normal capture efficiency. In some embodiments, the capture efficiency is lower than the reference capture efficiency. In some embodiments, the subject is identified as having an insertion in the target region. In some embodiments, the capture efficiency is higher than the reference capture efficiency. In some embodiments, the subject is identified as having a deletion in the target region. In some embodiments, the subject is identified as being heterozygous for the insertion. In some embodiments, the subject is identified as being heterozygous for the deletion.

In some embodiments of any of the methods described herein (e.g., tiling/staggering, tagging, size-detection, and/or sensitivity enhancement) aspects of the invention relate to capturing a sub-target nucleic acid (or a sequence of a sub-target nucleic acid). In some embodiments, a molecular inversion probe technique is used. In some embodiments, a molecular inversion probe is a single linear strand of nucleic acid that comprises a first targeting arm at its 5' end and a second targeting arm at its 3' end, wherein the first targeting arm is capable of

Petitioner Illumina, Inc., EX1004, p. 661

specifically hybridizing to a first region flanking one end of the sub-target nucleic acid, and wherein the second targeting arm is capable of specifically hybridizing to a second region flanking the other end of the sub target nucleic acid on the same strand of the target nucleic acid. In some embodiments, the first and second targeting arms are between about 10 and about 100 nucleotides long. In some embodiments, the first and second targeting arms are about 10-20, 20-30, 30-40, or 40-50 nucleotides long. In some embodiments, the first and second targeting arms are about 20 nucleotides long. In some embodiments, the first and second targeting arms have the same length. In some embodiments, the first and second targeting arms have different lengths. In some embodiments, each pair of first and second targeting arms in a set of probes has the same length. Accordingly, if one of the targeting arms is longer, the other one is correspondingly shorter. This allows for a quality control step in some embodiments to confirm that all captured probe/target sequence products have the same length after a multiplexed plurality of capture reactions. In some embodiments, a set of probes may be designed to have the same length if the intervening region is varied to accommodate any differences in the length of either one or both of the first and second targeting arms.

In some embodiments, the hybridization Tms of the first and second targeting arms are similar. In some embodiments, the hybridization Tms of the first and second targeting arms are within 2-5 °C of each other. In some embodiments, the hybridization Tms of the first and second targeting arms are identical. In some embodiments, the hybridization Tms of the first and second targeting arms are close to empirically-determined optima but not necessarily identical.

In some embodiments, the first and second targeting arms of a molecular inversion probe have different Tms. For example, the Tm of the first targeting arm (at the 5' end of the molecular inversion probe) may be higher than the Tm of the second targeting arm (at the 3' end of the molecular inversion probe). According to aspects of the invention, and without wishing to be bound by theory, a relatively high Tm for the first targeting arm may help avoid or prevent the first targeting arm from being displaced after hybridization by the extension product of the 3' end of the second targeting arm. It should be appreciated that a reference to the Tm of a targeting arm as used herein relates to the Tm of hybridization of the targeting arm to a nucleic acid having the complementary sequence (e.g., the region of the target nucleic acid that has a sequence that is complementary to the sequence of the targeting arm). It also should be appreciated that the Tms of the targeting arms described herein may be calculated using any

Petitioner Illumina, Inc., EX1004, p. 662

appropriate method. For example, in some embodiments an experimental method (e.g., a gel shift assay, a hybridization assay, a melting curve analysis, for example in a PCR machine with a SYBR dye by stepping through a temperature ramp while monitoring signal level from an intercalating dye, for example, bound to a double-stranded DNA, etc.) may be used to determine one or more Tms empirically. In some embodiments, an optimal Tm may be determined by evaluating the number of products formed (e.g., for each of a plurality of MIP probes), and determining the optimal Tm as the center point in a histogram of Tm for all targeting arms. In some embodiments, a predictive algorithm may be used to determine a Tm theoretically. In some embodiments, a relatively simple predictive algorithm may be used based on the number of G/C and A/T base pairs when the sequence is hybridized to its target and/or the length of the hybridized product (e.g., for example, 64.9 + 41 * ([G+C] - 16.4) / (A + T + G + C), see for example, Wallace,R.B., Shaffer,J., Murphy,R.F., Bonner,J., Hirose,T., and Itakura,K. (1979) Nucleic Acids Res 6:3543-3557). In some embodiments, a more complex algorithm may be used to account for the effects of base stacking entropy and enthalpy, ion concentration, and primer concentration (see, for example, SantaLucia J (1998), Proc Natl Acad Sci USA, 95:1460-5). In some embodiments an algorithm may use modified parameters (e.g., nearest-neighbor parameters for basepair entropy/enthalpy values). It should be appreciated that any suitable algorithm may be used as aspects of the invention are not limited in this respect. However, it also should be appreciated that different methodologies may results in different calculated or predicted Tms for the same sequences. Accordingly, in some embodiments, the same empirical and/or theoretical method is used to determine the Tms of different sequences for a set of probes to avoid a negative impact of any systematic difference in the Tm determination or prediction when designing a set of probes with predetermined similarities or differences for different Tms.

In some embodiments, the Tm of the first targeting arm may be about 1 °C, about 2 °C, about 3 °C, about 4 °C, about 5 °C, or more than about 5 °C higher than the Tm of the second targeting arm. In some embodiments, each probe in a plurality of probes (e.g., each probe in a set of 5-10, each probe in a set of at least 10, each probe in a set of 10-50, each probe in a set of 50-100, each probe in a set of 100-500, each probe in a set of 500-1,000, each probe in a set of 1,000-1,500, each probe in a set of 1,500-2,000, each probe in a set of 2,000-3,000, 3,000-5,000, 5,000-10,000 or each probe in a set of at least 5,000 different probes) has a unique first targeting arm (e.g., they all have different sequences) and a unique second targeting arm (e.g., they all

Petitioner Illumina, Inc., EX1004, p. 663

have different sequences). In some embodiments, for at least 10% of the probes (e.g., at least 25%, 25%-50%, 50%- 75%, 75%-90%, 90%-95% or over 95%, or all of the probes) the first targeting arm has a Tm for its complementary sequence that is higher (e.g., about 1 °C, about 2 °C, about 3 °C, about 4 °C, about 5 °C, or more than about 5 °C higher) than the Tm of the second targeting arm for its complementary sequence. In some embodiments, each of the first targeting arms have similar or identical Tms for their respective complementary sequences and each of the second targeting arms have similar or identical Tms for their respective complementary sequences (and the first targeting arms have higher Tms than the second targeting arms). For example, in some embodiments, the Tm of the first arm(s) may be about 58 °C and the Tm of the second arm(s) may be about 56 °C. In some embodiments, the Tm of the first arm(s) may be about 68 °C, and the Tm of the second arm(s) may be about 65 °C. It should be appreciated that in some embodiments the similarity (e.g., within a range of 1 °C, 2 °C, 3 °C, 4 °C, 5 °C) or identity of the Tms for the different targeting arms should be based either on empirical data for each arm or based on the same predictive algorithm for each arm (e.g., Wallace,R.B., Shaffer,J., Murphy,R.F., Bonner,J., Hirose,T., and Itakura,K. (1979) Nucleic Acids Res 6:3543-3557, SantaLucia J (1998), Proc Natl Acad Sci USA, 95:1460-5, or other algorithm).

In some embodiments, the Tm of the first targeting arm of a molecular inversion probe (at the 5' end of the molecular inversion probe) is selected to be sufficiently stable to prevent displacement of the first targeting arm from its complementary sequence on a target nucleic acid. In some embodiments, the Tm of the first targeting arm is 50-55 °C, at least 55 °C, 55-60 °C, at least 60 °C, 60-65 °C, at least 65 °C, at least 70 °C, at least 75 °C, or at least 80 °C. As discussed above, it should be appreciated that the for a particular targeting arm may be determined empirically or theoretically. Different theoretical models may be used to determine a Tm and it should be appreciated that the predicted Tm for a particular sequence may be different depending on the algorithm used for the prediction. In some embodiments, each probe in a plurality of probes (e.g., each probe in a set of 5-10, each probe in a set of at least 10, each probe in a set of 10-50, each probe in a set of 50-100, each probe in a set of 100-500, or each probe in a set of at least 500 different probes) has a different first targeting arm (e.g., different sequences) but each different first targeting arm has a similar or identical Tm for its complementary sequence on a target nucleic acid. It should be appreciated that in some embodiments the

Petitioner Illumina, Inc., EX1004, p. 664

similarity (e.g., within a range of I C, 2 C, 3 C, 4 C, 5 C) or identity of the Tms for the different targeting arms should be based either on empirical data for each arm or based on the same predictive algorithm for each arm (e.g., Wallace,R.B., Shaffer,J., Murphy,R.F., Bonner,J., Hirose,T., and Itakura,K. (1979) Nucleic Acids Res 6:3543- 3557, SantaLucia J (1998), Proc Natl Acad Sci USA, 95:1460-5, or other algorithm).

In some embodiments, the sub-target nucleic acid contains a nucleic acid repeat. In some embodiments, the nucleic acid repeat is a dinucleotide or trinucleotide repeat. In some embodiments, the sub-target nucleic acid contains 10-100 copies of the nucleic acid repeat in the absence of an abnormal increase or decrease in nucleic acid repeats. In some embodiments, the sub-target nucleic acid is a region of the Fragile-X locus that contains a nucleic acid repeat. In some embodiments, one or both targeting arms hybridize to a region on the target nucleic acid that is immediately adjacent to a region of nucleic acid repeats. In some embodiments, one or both targeting arms hybridize to a region on the target nucleic acid that is separated from a region of nucleic acid repeats by a region that does not contain any nucleic acid repeats. In some embodiments, the molecular inversion probe further comprises a primer-binding region that can be used to sequence the captured sub-target nucleic acid and optionally the first and/or second targeting arm.

In some embodiments, aspects of the invention relate to evaluating the length of a plurality of different target nucleic acids in a biological sample. In some embodiments, the plurality of target nucleic acids are analyzed using a plurality of different molecular inversion probes. In some embodiments, each different molecular inversion probe comprises a different pair of first and second targeting arms at each of the 3' and 5' ends. In some embodiments, each different molecular inversion probe comprises the same primer-binding sequence.

In some embodiments, aspects of the invention relate to analyzing nucleic acid from a biological sample obtained from a subject. In some embodiments, the biological sample is a blood sample. In some embodiments, the biological sample is a tissue sample, specific cell population, tumor sample, circulating tumor cells, or environmental sample. In some embodiments, the biological sample is a single cell. In some embodiments, nucleic acids are analyzed in biological samples obtained from a plurality of different subjects. In some embodiments, nucleic acids from a biological sample are analyzed in multiplex reactions. It should be appreciated that a biological sample contains a plurality of copies of a genome derived

Petitioner Illumina, Inc., EX1004, p. 665

from a plurality of cells in the sample. Accordingly, a sample may contain a plurality of independent copies of a target nucleic acid region of interest, the capture efficiency of which can be used to evaluate its size as described herein.

In some embodiments, aspects of the invention relate to evaluating a nucleic acid capture efficiency by determining an amount of target nucleic acid that is captured (e.g., an amount of sub-target nucleic acid sequences that are captured). In some embodiments, the amount of target nucleic acid that is captured is determined by determining a number of independently captured target nucleic acid molecules (e.g., the amount of independently captured molecules that have the sequence of the sub-target region). In some embodiments, the amount of target nucleic acid that is captured is compared to a reference amount of captured nucleic acid. In some embodiments, the reference amount is determined by determining a number of independently captured molecules of a reference nucleic acid. In some embodiments, the reference nucleic acid is a nucleic acid of a different locus in the biological sample that is not suspected of containing a deletion or insertion. In some embodiments, the reference nucleic acid is a nucleic acid of known size and amount that is added to the capture reaction. As described herein, a number of independently captured nucleic acid sequences can be determined by contacting a nucleic acid sample with a preparation of a probe (e.g., a MIP probe as described herein). It should be appreciated that the preparation may comprise a plurality of copies of the same probe and accordingly a plurality of independent copies of the target region may be captured by different probe molecules. The number of probe molecules that actually capture a sequence can be evaluated by determining an amount or number of captured molecules using any suitable technique. This number is a reflection of both the number of target molecules in the sample and the efficiency of capture of those target molecules, which in turn is related to the size of the target molecules as described herein. Accordingly, the capture efficiency can be evaluated by controlling for the abundance of the target nucleic acid, for example by comparing the number or amount of captured target molecules to an appropriate control (e.g., a known size and amount of control nucleic acid, or a different locus that should be present in the same amount in the biological sample and is not expected to contain any insertions or deletions). It should be appreciated that other factors may affect the capture efficiency of a particular target nucleic acid region (e.g., the sequence of the region, the GC content, the presence of secondary structures, etc.). However, these factors also can be accounted for by using appropriate controls (e.g.,

Petitioner Illumina, Inc., EX1004, p. 666

known sequences having similar properties, the same sequences, other genomic sequences expected to be present in the biological sample at the same frequency, etc., or any combination thereof).

In some embodiments, aspects of the invention relate to identifying a subject as having an insertion or deletion in one or more alleles of a genetic locus if the capture efficiency for that genetic locus is statistically significantly different than a reference capture efficiency.

It should be appreciated that hybridization conditions used for any of the capture techniques described herein (e.g., MIP capture techniques) can be based on known hybridization buffers and conditions.

In some embodiments, the methods disclosed herein are useful for any application where the detection of deletions or insertions is important.

In some embodiments, aspects of the invention relate to basing a nucleic acid sequence analysis on results from two or more different nucleic acid preparatory techniques that have different systematic biases in the types of nucleic acids that they sample. According to the invention, different techniques have different sequence biases that are systematic and not simply due to stochastic effects during nucleic acid capture or amplification. Accordingly, the degree of oversampling required to overcome variations in nucleic acid preparation needs to be sufficient to overcome the biases (e.g., an oversampling of 2-5 fold, 5-10 fold, 5-15 fold, 15-20 fold, 20-30 fold, 30-50 fold, or intermediate to higher fold).

According to some embodiments, different techniques have different characteristic or systematic biases. For example, one technique may bias a sample analysis towards one particular allele at a genetic locus of interest, whereas a different technique would bias the sample analysis towards a different allele at the same locus. Accordingly, the same sample may be identified as being different depending on the type of technique that is used to prepare nucleic acid for sequence analysis. This effectively represents a sensitivity limitation, because each technique has different relative sensitivities for polymorphic sequences of interest.

According to aspects of the invention, the sensitivity of a nucleic acid analysis can be increased by combining the sequences from different nucleic acid preparative steps and using the combined sequence information for a diagnostic assay (e.g., for a making a call as to whether a subject is homozygous or heterozygous at a genetic locus of interest).

Petitioner Illumina, Inc., EX1004, p. 667

In some embodiments, the invention provides a method of increasing the sensitivity of a nucleic acid detection assay by obtaining a first preparation of a target nucleic acid using a first preparative method on a biological sample, obtaining a second preparation of a target nucleic acid using a second preparative method on the biological sample, assaying the sequences obtained in both first and second nucleic acid preparations, and using the sequence information from both first and second nucleic acid preparations to determine the genotype of the target nucleic acid in the biological sample, wherein the first and second preparative methods have different systematic sequence biases. In some embodiments, the first and second nucleic acid preparations are combined prior to performing a sequence assay. In some embodiments, separate sequence assays are performed on the first and second nucleic acid preparations and the sequence information from both assays are combined to determine the genotype of the target nucleic acid in the biological sample. In some embodiments, the first preparative method is an amplification-based, a hybridization-based, or a circular probe-based preparative method. In some embodiments, the second method is an amplification-based, a hybridization-based, or a circular probe-based preparative method. In some embodiments, the first and second methods are of different types (e.g., only one of them is an amplification-based, a hybridization-based, or a circular probe-based preparative method, and the other one is one or the other two types of method). Accordingly, in some embodiments the second preparative method is an amplification-based, a hybridization-based, or a circular probe-based preparative method, provided that the second method is different from the first method. However, in some embodiments, both methods may be of the same type, provided they are different methods (e.g., both are amplification based or hybridization-based, but are different types of amplification or hybridization methods, e.g., with different relative biases).

In amplification-based (e.g., PCR-based or LCR-based, etc.) preparative methods, genomic loci (target nucleic acids) are isolated directly by means of a polymerase chain reaction or ligase chain reaction (or other amplification method) that selectively amplifies each locus using a pair of oligonucleotide primers. It is to be understood that primers will be sufficiently complementary to the target sequence to hybridize with and prime amplification of the target nucleic acid. Any one of a variety of art known methods may be utilized for primer design and synthesis. One or both of the primers may be perfectly complementary to the target sequence. Degenerate primers may also be used. Primers may also include additional nucleic acids that are

Petitioner Illumina, Inc., EX1004, p. 668

not complementary to target sequences but that facilitate downstream applications, including for example restriction sites and identifier sequences (e.g., source sequences). PCR based methods may include amplification of a single target nucleic acid and multiplex amplification (amplification of multiple target nucleic acids in parallel).

Hybridization-based preparative may methods involve selectively immobilizing target nucleic acids for further manipulation. It is to be understood that one or more oligonucleotides (immobilization oligonucleotides), which in some embodiments may be from 10 to 200 nucleotides in length, are used which hybridize along the length of a target region of a genetic locus to immobilize it. In some embodiments, immobilization oligonucleotides are either immobilized before hybridization is performed (e.g., Roche/Nimblegen 'sequence capture'), or are prepared such that they include a moiety (e.g., biotin) which can be used to selectively immobilize the target nucleic acid after hybridization by binding to e.g., streptavidin-coated microbeads (e.g., Agilent 'SureSelect').

Circularization selection-based preparative methods selectively convert each region of interest into a covalently-closed circular molecule which is then isolated by removal (usually enzymatic, e.g., with exonuclease) of any non-circularized linear nucleic acid. Oligonucleotide probes are designed which have ends that flank the region of interest. The probes are allowed to hybridize to the genomic target, and enzymes are used to first (optionally) fill in any gap between probe ends and second ligate the probe closed. In some embodiments, following circularization, any remaining (non-target) linear nucleic acid can be removed, resulting in isolation (capture) of target nucleic acid. Circularization selection-based preparative methods include molecular inversion probe capture reactions and 'selector' capture reactions. However, other techniques may be used as aspects of the invention are not limited in this respect. In some embodiments, molecular inversion probe capture of a target nucleic acid is indicative of the presence of a polymorphism in the target nucleic acid.

A variety of methods may be used to evaluate and compare bias profiles of each preparative technique. Next-generation sequencing may be used to quantitatively measure the abundance of each isolated target nucleic acid obtained from a certain preparative method. This abundance may be compared to a control abundance value (e.g., a known starting abundance of the target nucleic acid) and/or with an abundance determined through the use of an alternative preparative method. For example, a set of target nucleic acids may be isolated by one or more of

Petitioner Illumina, Inc., EX1004, p. 669

the three preparative methods; the target nucleic acid may be observed x times using the amplification technique, y times using the hybridization enrichment technique, and z times using the circularization selection technique. A pairwise correlation coefficient may be computed between each abundance value (e.g., x and y, x and z, and y and z) to assess bias in nucleic acid isolation between pairs of preparative methods. Since the mechanisms of isolation are different in each approach, the abundances will usually be different and largely uncorrelated with each other.

In some embodiments, the invention provides a method of obtaining a nucleic acid preparation that is representative of a target nucleic acid in a biological sample by obtaining a first preparation of a target nucleic acid using a first preparative method on a biological sample, obtaining a second preparation of a target nucleic acid using a second preparative method on the biological sample, and combining the first and second nucleic acid preparations to obtain a combined preparation that is representative of the target nucleic acid in the biological sample.

In some embodiments of any of the methods described herein, a third preparation of the target nucleic acid is obtained using a third preparative method that is different from the first and second preparative methods, wherein the first, second, and third preparative methods all have different systematic sequence biases. In some embodiments of any of the methods described herein, the different preparative methods are used for a plurality of different loci in the biological sample to increase the sensitivity of a multiplex nucleic acid analysis. In some embodiments, the target nucleic acid has a sequence of a gene selected from Table 1.

However, it should be appreciated that a genotyping method of the invention may include several steps, each of which independently may involve one or more different preparative techniques described herein. In some embodiments, a nucleic acid preparation may be obtained using one or more (e.g., 2, 3, 4, 5, or more) different techniques described herein (e.g., amplification, hybridization capture, circular probe capture, etc., or any combination thereof) and the nucleic acid preparation may be analyzed using one or more different techniques (e.g., amplification, hybridization capture, circular probe capture, etc., or any combination thereof) that are selected independently of the techniques used for the initial preparation.

In some embodiments, aspects of the invention also provide compositions, kits, devices, and analytical methods for increasing the sensitivity of nucleic acid assays. Aspects of the invention are particularly useful for increasing the confidence level of genotyping analyses.

Petitioner Illumina, Inc., EX1004, p. 670

However, aspects of the invention may be used in the context of any suitable nucleic acid analysis, for example, but not limited to, a nucleic acid analysis that is designed to determine whether more than one sequence variant is present in a sample.

In some embodiments, aspects of the invention relate to a plurality of nucleic acid probes (e.g., 10-50, 50-100, 100-250, 250-500, 500-1,000, 1,000-2,000, 2,000-5,000, 5,000-7,500, 7,500-10,000, or lower, higher, or intermediate number of different probes).

In some embodiments, each probe or each of a subset of probes (e.g., 10-25%, 25-50%, 50-75%, 75-90%, or 90-99%) has a different first targeting arm. In some embodiments, each probe or each probe of a subset of probes (e.g., 10-25%, 25-50%, 50-75%, 75-90%, or 90-99%) has a different second targeting arm. In some embodiments, the first and second targeting arms are separated by the same intervening sequence. In some embodiments, the first and second targeting arms are complementary to target nucleic acid sequences that are separated by the same or a similar length (e.g., number of nucleic acids, for example, 0-25, 25-50, 50-100, 100-250, 250-500, 500-1,000, 1,000-2,500 or longer or intermediate number of nucleotides) on their respective target nucleic acids (e.g., genomic loci). In some embodiments, each probe or a subset of probes (e.g., 10-30 25%, 25-50%, 50-75%, 75-90%, or 90-99%) includes a first primer binding sequence. In some embodiments, the primer binding sequence is the same (e.g., it can be used to prime sequencing or other extension reaction). In some embodiments, each probe or a subset of probes (e.g., 10-25%, 25-50%, 50-75%, 75-90%, or 90-99%) includes a unique identifier sequence tag (e.g., that is predetermined and can be used to distinguish each probe).

In some embodiments, the methods disclosed herein are useful for any application where sensitivity is important. For example, detection of cancer mutations in a heterogenous tissue sample, detection of mutations in maternally-circulating fetal DNA, and detection of mutations in cells isolated during a preimplantation genetic diagnostic procedure.

According to some aspects of the invention, methods of detecting a polymorphism in a nucleic acid in a biological sample are provided. In some embodiments, the methods comprise obtaining a nucleic acid preparation using a preparative method (e.g., any of the preparative methods disclosed herein) on a biological sample, and performing a molecular inversion probe capture reaction on the nucleic acid preparation, wherein a molecular inversion probe capture (e.g., using a mutation-detection MIP) of a target nucleic acid of the nucleic acid preparation is

Petitioner Illumina, Inc., EX1004, p. 671

indicative of the presence of a mutation (polymorphism) in the target nucleic acid, optionally wherein the polymorphism is selected from Table 2.

According to some aspects of the invention, methods of genotyping a nucleic acid in a biological sample are provided. In some embodiments, the methods comprise obtaining a nucleic acid preparation using a preparative method on a biological sample, sequencing a target nucleic acid of the nucleic acid preparation, and performing a molecular inversion probe capture reaction on the biological sample, wherein a molecular inversion probe capture of the target nucleic acid in the biological sample is indicative of the presence of a polymorphism in the target nucleic acid, genotyping the target nucleic acid based on the results of the sequencing and the capture reaction.

In some embodiments of the methods disclosed herein, the target nucleic acid has a sequence of a gene selected from Table 1.

It should be appreciated that any one or more embodiments described herein may be used for evaluating multiple genetic markers in parallel. Accordingly, in some embodiments, aspects of the invention relate to determining the presence of one or more markers (e.g., one or more alleles) at multiple different genetic loci in parallel.

Accordingly, the risk or presence of multiple heritable disorders may be evaluated in parallel. In some embodiments, the risk of having offspring with one or more heritable disorders may be evaluated. In some embodiments, an evaluation may be performed on a biological sample of a parent or a child (e.g., at a pre-implantation, prenatal, perinatal, or postnatal stage). In some embodiments, the disclosure provides methods for analyzing multiple genetic loci (e.g., a plurality of target nucleic acids selected from Table 1 or 2) from a patient sample, such as a blood, pre-implantation embryo, chorionic villus or amniotic fluid sample. A patient or subject may be a human. However, aspects of the invention are not limited to humans and may be applied to other species (e.g., mammals, birds, reptiles, other vertebrates or invertebrates) as aspects of the invention are not limited in this respect. A subject or patient may be male or female. In some embodiments, in connection with reproductive genetic counseling, samples from a male and female member of a couple may be analyzed. In some embodiments, for example, in connection with an animal breeding program, samples from a plurality of male and female subjects may be analyzed to determine compatible or optimal breeding partners or strategies for particular traits or to avoid one or more diseases or conditions. Accordingly,

Petitioner Illumina, Inc., EX1004, p. 672

reproductive risks may be determined and/or reproductive recommendations may be provided based on information derived from one or more embodiments of the invention.

However, it should be appreciated that aspects of the invention may be used in connection with any medical evaluation where the presence of one or more alleles at a genetic locus of interest is relevant to a medical determination (e.g., risk or detection of disease, disease prognosis, therapy selection, therapy monitoring, etc.). Further aspects of the invention may be used in connection with detection, in tumor tissue or circulating tumor cells, of mutations in cellular pathways that cause cancer or predict efficacy of treatment regimens, or with detection and identification of pathogenic organisms in the environment or a sample obtained from a subject, e.g., a human subject.

These and other aspects of the invention are described in more detail in the following description and non-limiting examples and drawings.

BRIEF DESCRIPTION OF DRAWINGS

FIG. 1 illustrates a non-limiting embodiment of a tiled probe layout;

FIG. 2 illustrates a non-limiting embodiment of a staggered probe layout;

FIG. 3 illustrates a non-limiting embodiment of an alternating staggered probe layout;

FIGS. 4A-C depict various non-limiting methods for combining differentiator tag sequence and target sequences (NNNN depicts a differentiator tag sequence);

FIG. 5 depicts a non-limiting method for genotyping based on target and 5 differentiator tag sequences;

FIG. 6 depicts non-limiting results of a simulation of a MIP capture reaction;

FIG. 7 depicts a non-limiting graph of sequencing coverage;

FIG. 8 illustrates that shorter sequences are captured with higher efficiency that longer sequences using MIPs;

FIG. 9 illustrates a non-limiting scheme of padlock (MIP) capture of a region that includes both repetitive regions (thick wavy line) and the adjacent unique sequence (thick strait line);

FIG. 10 illustrates a non-limiting hypothetical relationship between target gap size and the relative number of reads of the repetitive region;

FIG. 11A depicts MIP capture of FMR1 repeat regions from a diploid genome;

Petitioner Illumina, Inc., EX1004, p. 673

FIG. 11B depicts preparative methods for biallelic resolution of FMR I repeat region lengths in a diploid genome using MIP capture probes and unique differentiator tags;

FIG. 11C depicts an analysis of FMR1 repeat region lengths in a diploid genome;

FIG. 12 is a schematic of an embodiment of an algorithm of the invention;

FIG. 13 illustrates a non-limiting example of a graph of per-target abundance with MIP capture; and,

FIG. 14 shows a non-limiting a graph of correlation between two MIP capture reactions.

DETAILED DESCRIPTION

Aspects of the invention relate to preparative and analytical methods and compositions for evaluating genotypes, and in particular, for determining the allelic identity (or identities in a diploid organism) of one or more genetic loci in a subject.

Aspects of the invention are based, in part, on the identification of different sources of ambiguity and error in genetic analyses, and, in part, on the identification of one or more approaches to avoid, reduce, recognize, and/or resolve these errors and ambiguities at different stages in a genetic analysis. Aspects of the invention relate to methods and compositions for addressing bias and/or stochastic variation associated with one or more preparative and/or analytical steps of a nucleic acid evaluation technology. In some embodiments, preparative methods can be adapted to avoid or reduce the risk of bias skewing the results of a genetic analysis. In some embodiments, analytical methods can be adapted to recognize and correct for data variations that may give rise to misinterpretation (e.g., incorrect calls such as homozygous when the subject is actually heterozygous or heterozygous when the subject is actually homozygous). Methods of the invention may be used for any type of mutation, for example a single base change (e.g., insertion, deletion, transversion or transition, etc.), a multiple base insertion, deletion, duplication, inversion, and/or any other change or combination thereof.

In some embodiments, additional or alternative techniques may be used to address loci characterized by multiple repeats of a core sequence where the length of the repeat is longer than a typical sequencing read thereby making it difficult to determine whether a deletion or duplication of one or more core sequence units has occurred based solely on a sequence read.

In some embodiments, increased confidence in an assay result may be obtained by i) selecting two or more different preparative and/or analytical techniques that have different biases

Petitioner Illumina, Inc., EX1004, p. 674

(e.g., known to have different biases), ii) evaluating a patient sample using the two or more different techniques, iii) comparing the results from the two or more different techniques, and/or iv) determining whether the results are consistent for the two or more different techniques. In some embodiments, if determining in step (iv) indicates that the results are consistent (e.g., the same) then increased confidence in the assay result is obtained. In other embodiments, if determining in step (iv) indicates that the results are inconsistent (e.g., that the results are ambiguous) then one or more additional preparative and/or analytical techniques, which have a different bias (e.g., known to have a different bias) compared with the two or more different preparative and/or analytical techniques selected in step (i), are used to evaluate the patient sample, and the results of the one or more additional preparative and/or analytical techniques are compared with the results from step (ii) to resolve the inconsistency.

In some embodiments, two or more independent samples may be obtained from a subject and independently analyzed. In some embodiments, two or more independent samples are obtained at approximately the same time point. In some embodiments, two or more independent samples are obtained at multiple different time points. In some embodiments, the use of two or more independent sample facilitates the elimination, normalization, and/or quantification of stochastic measurement noise. It is to be appreciated that two or more independent samples may be obtained in connection with any of the methods disclosed herein, including, for example, methods for pathogen profiling in a human or other animal subjects, monitoring tumor progression/regression, analyzing circulating tumor cells, analyzing fetal cells in maternal circulation, and analyzing/monitoring/profiling of environmental pathogens.

In some embodiments, one or more of the techniques described herein may be combined in a single assay protocol for evaluating multiple patient samples in parallel.

It should be appreciated that aspects of the invention may be useful for high throughput, cost-effective, yet reliable, genotyping of multiple patient samples (e.g., in parallel, for example in multiplex reactions). In some embodiments, aspects of the invention are useful to reduce the error frequency in a multiplex analysis. Certain embodiments may be particularly useful where multiple reactions (e.g., multiple loci and/or multiple patient samples) are being processed. For example, 10-25, 25-50, 50-75, 75-100 or more loci may be evaluated for each subject out of any number of subject samples that may be processed in parallel (e.g., 1-25, 25-50, 50-100, 100-500, 500-1,000, 1,000-2,500, 2,500-5,000 or more or intermediate numbers of patient samples). It

Petitioner Illumina, Inc., EX1004, p. 675

should be appreciated that different embodiments of the invention may involve conducting two or more target capture reactions and/or two or more patient sample analyses in parallel in a single multiplex reaction. For example, in some embodiments a plurality of capture reactions (e.g., using different capture probes for different target loci) may be performed in a single multiplex reaction on a single patient sample. In some embodiments, a plurality of captured nucleic acids from each one of a plurality of patient samples may be combined in a single multiplex analysis reaction. In some embodiments, samples from different subjects are tagged with subject-specific (e.g., patient-specific) tags (e.g., unique sequence tags) so that the information from each product can be assigned to an identified subject. In some embodiments, each of the different capture probes used for each patient sample have a common patient-specific tag. In some embodiments, the capture probes do not have patient-specific tags, but the captured products from each subject may be amplified using one or a pair of amplification primers that are labeled with a patient-specific tag. Other techniques for associating a patient-specific tag with the captured product from a single patient sample may be used as aspects of the invention are not limited in this respect. It should be appreciated that patient-specific tags as used herein may refer to unique tags that are assigned to identified patients in a particular assay. The same tags may be used in a separate multiplex analysis with a different set of patient samples (e.g., from different patients) each of which is assigned one of the tags. In some embodiments, different sets of unique tags may be used in sequential (e.g., alternating) multiplex reactions in order to reduce the risk of contamination from one assay to the next and allow contamination to be detected on the basis of the presence of tags that are not expected to be present in a particular assay.

Embodiments of the invention may be used for any of a number of different settings: reproductive settings, disease screening, identifying subjects having cancer, identifying subjects having increased risk for a disease, stratifying a population of subjects according to one or more of a number of factors, for example responsiveness to a particular drug, lack or not of an adverse reaction (or risk therefore) to a particular drug, and/or providing information for medical records (e.g., homozygosity, heterozygosity at one or more loci). It should be appreciated that the invention is not limited to genomic analysis of patient samples. For example, aspects of the invention may be useful for high throughput genetic analysis of environment samples to detect pathogens.

Petitioner Illumina, Inc., EX1004, p. 676

In some embodiments, the methods disclosed herein are useful for diagnosis of one or more heritable disorders. In some embodiments, a heritable disorder that may be diagnosed with the methods disclosed herein is a genetic disorder that is prevalent in the Ashkenazi Jewish population. In some embodiments, the heritable disorders are selected from: 21-Hydroxylase-Defiocient Congenital Adrenal Hyperplasia; ABCC8-Related Hyperinsulinism; Alpha-Thalassemia, includes Constant Spring, & MR associated; Arylsulfatase A Deficiency-Metyachromatic Leukodystrophy; Biotinidase Deficiency Holocarboxylase Synthetase Deficiency; Bloom's Syndrome; Canavan Disease; CFTR-Related Disorders-cystic fibrosis; Citrullinemia Type 1; Combined MMA & Homocystinuria-dblC; Dystrophinopathies (DMD & BMD); Familial Dysautonomia; Fanconi Anemia-FANCC; Galactosemia-Classical: Galactokinase Defiency & Galactose Epimerase Deficiency; Gaucher Disease; GJB2-Related DFNB 1 Nonsyndromic Hearing Loss and Deafness; Glutaric acidemia Type 1; Hemoglobinopathies beta-chain disorders; Glycogen Storage Disease Type 1A; Maple Syrup Urine Disease; Types 1A, 1B, 2, 3; Medium Chain Acyl-Coenzyme A; Dehydrogenase Deficiency-MCADD; Methylmalonic Acidemia; Mucolipidosis IV; Nemaline Myopathy; Nieman-Pick Type A-Acid Sphingomyelinase Deficiency; Non-Ketotic Hyperglycinemia-Glycine Encephalopathy; Ornithine Transcarbamylase Deficiency; PKU Phenylalanine Hydroxylase Deficiency; Propionic Acidemia; Short Chain Acyl-CoA Dehydrogenase Deficiency-SCADD; Smith-Lemli-Opitz Syndrome; Spinal Muscular Atrophy (SMNI)-SMA; Tay Sachs-HexA Deficiency; Usher Synbdrome-Type I (Type IB, Type IC, Type ID, Type IF, Type IG) ; X-Linked Mental Retardation ARX-Related Disorders; X-Linked Mental Retardation with Cerebellar Cypoplasia and sistinctive Facial Appearance; X-Linked Mental Retardation; includes 9, 21, 30, 46, 58, 63, 88, 89; X linked mental retardation: FM1-Related Disorders-FRXA, Fragile X MR; X-linked SMR: Renpenning Syndrome 1; Zellweger Spectrum disorders - Peroxisomal Bifunctional Enzyme Deficiencies including Zellweger, NALD, and/or infantile Refsums. However, all of these, subsets of these, other genes, or combinations thereof may be used.

According to some aspects, the disclosure relates to multiplex diagnostic methods. In some embodiments, multiplex diagnostic methods comprise capturing a plurality of genetic loci in parallel (e.g., a genetic locus of Table 1). In some embodiments, genetic loci possess one or more polymorphisms (e.g., a polymorphism of Table 2) the genotypes of which correspond to

Petitioner Illumina, Inc., EX1004, p. 677

disease causing alleles. Accordingly, in some embodiments, the disclosure provides methods for assessing multiple heritable disorders in parallel.

In some embodiments, methods are provided for diagnosing multiple heritable disorders in parallel at a pre-implantation, prenatal, perinatal, or postnatal stage. In some embodiments, the disclosure provides methods for analyzing multiple genetic loci (e.g., a plurality of target nucleic acids selected from Table 1) from a patient sample, such as a blood, pre-implantation embryo, chorionic villus or amniotic fluid sample. A patient or subject may be a human. However, aspects of the invention are not limited to humans and may be applied to other species (e.g., mammals, birds, reptiles, other vertebrates or invertebrates) as aspects of the invention are not limited in this respect. A subject or patient may be male or female. In some embodiments, in connection with reproductive genetic counseling, samples from a male and female member of a couple may be analyzed. In some embodiments, for example, in connection with an animal breeding program, samples from a plurality of male and female subjects may be analyzed to determine compatible or optimal breeding partners or strategies for particular traits or to avoid one or more diseases or conditions.

However, it should be appreciated that any other diseases may be studied and/or risk factors for diseases or disorders including, but not limited to allergies, responsiveness to treatment, cancer tumor profiling for treatment and prognosis, monitoring and identification of patient infections, and monitoring of environmental pathogens.

1. Reducing representational bias in multiplex amplification reactions:

In some embodiments, aspects of the invention relate to methods that reduce bias and increase reproducibility in multiplex detection of genetic loci, e.g., for diagnostic purposes.

Molecular inversion probe technology is used to detect or amplify particular nucleic acid sequences in potentially complex mixtures. Use of molecular inversion probes has been demonstrated for detection of single nucleotide polymorphisms (Hardenbol et al. 2005 Genome Res 15:269-75) and for preparative amplification of large sets of exons (Porreca et al. 2007 Nat Methods 4:931-6, Krishnakumar et al. 2008 Proc Natl Acad Sci USA 105:9296-301). One of the main benefits of the method is in its capacity for a high degree of multiplexing, because generally thousands of targets may be captured in a single reaction containing thousands of probes. However, challenges associated with, for example, amplification efficiency (See, e.g.,

Petitioner Illumina, Inc., EX1004, p. 678

Turner EH, et al., Nat Methods. 2009 Apr 6:1-2.) have limited the practical utility of the method in research and diagnostic settings.

Aspects of the disclosure are based, in part, on the discovery of effective methods for overcoming challenges associated with systematic errors (bias) in multiplex genomic capture and sequencing methods, namely high variability in target nucleic acid representation and unequal sampling of heterozygous alleles in pools of captured target nucleic acids (e.g., isolated from a biological sample). Accordingly, in some embodiments, the disclosure provides methods that reduce variability in the detection of target nucleic acids in multiplex capture methods. In other embodiments, methods improve allelic representation in a capture pool and, thus, improve variant detection outcomes. In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve the use of different sets of multiple probes (e.g., molecular inversion probes MIPs) that capture overlapping regions of a target nucleic acid to achieve a more uniform representation of the target nucleic acids in a capture pool compared with methods of the prior art. In other embodiments, methods reduce bias, or the risk of bias, associated with large scale parallel capture of genetic loci, e.g., for diagnostic purposes. In other embodiments, methods are provided for increasing reproducibility (e.g., by reducing the effect of polymorphisms on target nucleic acid capture) in the detection of a plurality of genetic loci in parallel. In further embodiments, methods are provided for reducing the effect of probe synthesis and/or probe amplification variability on the analysis of a plurality of genetic loci in parallel.

In some aspects, the disclosure provides probe sets that comprise a plurality of different probes. As used herein, a 'probe' is a nucleic acid having a central region flanked by a 5' region and a 3' region that are complementary to nucleic acids flanking the same strand of a target nucleic acid or subregion thereof. An exemplary probe is a molecular inversion probe (MIP). A 'target nucleic acid' may be a genetic locus.

Exemplary genetic loci are disclosed herein in Table 1 (RefSeqGene Column).

While probes have been typically designed to meet certain constraints (e.g. melting temperature, G/C content, etc.) known to partially affect capture/amplification efficiency (Ball et al (2009) Nat Biotech 27:361-8 AND Deng et al (2009) Nat Biotech 27:353-60), a set of constraints which is sufficient to ensure either largely uniform or highly reproducible capture/amplification efficiency has not previously been achieved.

Petitioner Illumina, Inc., EX1004, p. 679

As disclosed herein, uniformity and reproducibility can be increased by designing multiple probes per target, such that each base in the target is captured by more than one probe. In some embodiments, the disclosure provides multiple MIPs per target to be captured, where each MIP in a set designed for a given target nucleic acid has a central region and a 5' region and 3' region ('targeting arms') which hybridize to (at least partially) different nucleic acids in the target nucleic acid (immediately flanking a subregion of the target nucleic acid). Thus, differences in efficiency between different targeting arms and fill-in sequences may be averaged across multiple MIPs for a single target, which results in more uniform and reproducible capture efficiency.

In some embodiments, the methods involve designing a single probe for each target (a target can be as small as a single base or as large as a kilobase or more of contiguous sequence).

It may be preferable, in some cases, to design probes to capture molecules (e.g., target nucleic acids or subregions thereof) having lengths in the range of 1 - 200 bp (as used herein, a bp refers to a base pair on a double-stranded nucleic acid - however, where lengths are indicated in bps, it should be appreciated that single-stranded nucleic acids having the same number of bases, as opposed to base pairs, in length also are contemplated by the invention). However, probe design is not so limited. For example, probes can be designed to capture targets having lengths in the range of up to 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 1000, or more bps, in some cases.

It is to be appreciated that the length of a capture molecule (e.g., a target nucleic acid or subregion thereof) is selected based upon multiple considerations. For example, where analysis of a target involves sequencing, e.g., with a next-generation sequencer, the target length should typically match the sequencing read-length so that shotgun library construction is not necessary. However, it should be appreciated that captured nucleic acids may be sequenced using any suitable sequencing technique as aspects of the invention are not limited in this respect.

It is also to be appreciated that some target nucleic acids are too large to be captured with one probe. Consequently, it may be necessary to capture multiple subregions of a target nucleic acid in order to analyze the full target.

In some embodiments, a subregion of a target nucleic acid is at least 1 bp. In other embodiments, a subregion of a target nucleic acid is at least 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 bp or more. In other embodiments, a

Petitioner Illumina, Inc., EX1004, p. 680

subregion of a target nucleic acid has a length that is up to 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or more percent of a target nucleic acid length.

The skilled artisan will also appreciate that consideration is made, in the design of MIPs, for the relationship between probe length and target length. In some embodiments, MIPs are designed such that they are several hundred basepairs (e.g., up to 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 bp or more) longer than corresponding target (e.g., subregion of a target nucleic acid, target nucleic acid).

In some embodiments, lengths of subregions of a target nucleic acid may differ.

For example, if a target nucleic acid contains regions for which probe hybridization is not possible or inefficient, it may be necessary to use probes that capture subregions of one or more different lengths in order to avoid hybridization with problematic nucleic acids and capture nucleic acids that encompass a complete target nucleic acid.

Aspects of the invention involve using multiple probes, e.g., MIPs, to amplify each target nucleic acid. In some embodiments, the set of probes for a given target can be designed to 'tile' across the target, capturing the target as a series of shorter sub targets. In some embodiments, where a set of probes for a given target is designed to 'tile' across the target, some probes in the set capture flanking non-target sequence). Alternately, the set can be designed to 'stagger' the exact positions of the hybridization regions flanking the target, capturing the full target (and in some cases capturing flanking non-target sequence) with multiple probes having different targeting arms, obviating the need for tiling. The particular approach chosen will depend on the nature of the target set. For example, if small regions are to be captured, a staggered-end approach might be appropriate, whereas if longer regions are desired, tiling might be chosen. In all cases, the amount of bias-tolerance for probes targeting pathological loci can be adjusted ('dialed in') by changing the number of different MIPs used to capture a given molecule.

In some embodiments, the 'coverage factor', or number of probes used to capture a basepair in a molecule, is an important parameter to specify. Different numbers of probes per target are indicated depending on whether one is using the tiling approach (see, e.g., FIG. 1) or one of the staggered approaches (see, e.g., FIGs. 2 or 3).

FIG. 1 illustrates a non-limiting embodiment of a tiled probe layout showing ten captured sub-targets tiled across a single target. Each position in the target is covered by three sub-targets such that MIP performance per base pair is averaged across three probes.

Petitioner Illumina, Inc., EX1004, p. 681

FIG. 2 illustrates a non-limiting embodiment of a staggered probe layout showing the targets captured by a set of three MIPs. Each MIP captures the full target, shown in black, plus (in some cases) additional extra-target sequence, shown in gray, such that the targeting arms of each MIP fall on different sequence. Each position in the target is covered by three sub-targets such that MIP performance per basepair is averaged across three probes. Targeting arms land immediately adjacent to the black or gray regions shown. It should be appreciated that in some embodiments, the targeting arms (not shown) can be designed so that they do not overlap with each other.

FIG. 3 illustrates a non-limiting embodiment of an alternating staggered probe layout showing the targets captured by a set of three MIPs. Each MIP captures the full target, shown in black, plus (in some cases) additional extra-target sequence, shown in gray, such that the targeting arms of each MIP fall on different sequence. Each position in the target is covered by three sub-targets such that MIP performance per basepair is averaged across three probes. Targeting arms land immediately adjacent to the black or gray regions shown.

It should be appreciated that for any of the layouts, the targeting arms on adjacent tiled or staggered probes may be designed to either overlap, not overlap, or overlap for only a subset of the probes.

In certain embodiments for any of the layouts, a coverage factor of about 3 to about 10 is used. However, the methods are not so limited and coverage factors of up to 2, 3, 4, 5, 6, 7, 8, 9, 10, 20 or more may be used. It is to be appreciated that the coverage factor selected may depend the probe layout being employed. For example, in the tiling approach, for a desired coverage factor, the number of probes per target is typically a function of target length, sub-target length, and spacing between adjacent sub-target start locations (step size). For example, for a desired coverage factor of 3, a 200 bp target with a start-site separation of 20 bp and sub-target length of 60 bp may be encompassed with 12 MIPs (FIG. 1). Thus, a specific coverage factor may be achieved by varying the number of probes per target nucleic acid and the length of the molecules captured. In the staggered approach, a fixed-length target nucleic acid is captured as several subregions or as 'super-targets', which are molecules comprising the target nucleic acid and additional flanking nucleic acids, which may be of varying lengths. For example, a target of 50 bp can be captured at a coverage factor of 3 with 3 probes in either a 'staggered' (FIG. 2) or 'alternating staggered' configuration (FIG. 3).

Petitioner Illumina, Inc., EX1004, p. 682

The coverage factor will be driven by the extent to which detection bias is tolerable. In some cases, where the bias tolerance is small, it may be desirable to target more subregions of target nucleic acid with, perhaps, higher coverage factors. In some embodiments, the coverage factor is up to 2, 3, 4, 5, 6, 7, 8, 9, 10 or more.

In some embodiments, when a tiled probe layout is used, when the target length is greater than 1 bp and when a step size (distance between the 5'-end of a target and the 5' end of its adjacent target) is less than the length of a target or subregion thereof, it is possible to compute probe number for a particular target based on target length (T), sub target length (S), and coverage factor (C), such that probe number = T/(S/C) + (C-1).

In some aspects, the disclosure provides methods to increase the uniformity of amplification efficiency when multiple molecules are amplified in parallel; methods to increase the reproducibility of amplification efficiency; methods to reduce the contribution of targeting probe variability to amplification efficiency; methods to reduce the effect on a given target nucleic acid of polymorphisms in probe hybridization regions; and/or methods to simplify downstream workflows when multiplex amplification by MIPs is used as a preparative step for analysis by nucleic acid sequencing.

Polymorphisms in the target nucleic acid under the regions flanking a target can interfere with hybridization, polymerase fill-in, and/or ligation. Furthermore, this may occur for only one allele, resulting in allelic drop-out, which ultimately decreases downstream sequencing accuracy. In some embodiments, using a set of MIPs having multiple hybridization sites for the capture of any given target, the probability of loss from polymorphism is substantially decreased because not all targeting arms in the set of MIPs will cover the location of the mutation.

Probes for MIP capture reactions may be synthesized on programmable microarrays because of the large number of sequences required. Because of the low synthesis yields of these methods, a subsequent amplification step is required to produce sufficient probe for the MIP amplification reaction. The combination of multiplex oligonucleotide synthesis and pooled amplification results in uneven synthesis error rates and representational biases. By synthesizing multiple probes for each target, variation from these sources may be averaged out because not all probes for a given target will have the same error rates and biases.

Multiplex amplification strategies disclosed herein may be used analytically, as in detection of SNPs, or preparatively, often for next-generation sequencing or other sequencing

Petitioner Illumina, Inc., EX1004, p. 683

techniques. In the preparative setting, the output of an amplification reaction is generally the input to a shotgun library protocol, which then becomes the input to the sequencing platform. The shotgun library is necessary in part because next-generation sequencing yields reads significantly shorter than amplicons such as exons. In addition to the bias-reduction afforded by the multi-tiled approach described here, tiling also obviates the need for shotgun library preparation. Since the length of the capture molecule can be specified when the probes, e.g., MIPs, are designed, it can be chosen to match the readlength of the sequencer. In this way, reads can 'walk' across an exon by virtue of the start position of each capture molecule in the probe set for that exon. Reducing analytical errors associated with bias in nucleic acid preparations:

In some embodiments, aspects of the invention relate to preparative steps in DNA sequencing-related technologies that reduce bias and increase the reliability and accuracy of downstream quantitative applications.

There are currently many genomics assays that utilize next-generation (polony-based) sequencing to generate data, including genome resequencing, RNA-seq for gene expression, bisulphite sequencing for methylation, and Immune-seq, among others. In order to make quantitative measurements (including genotype calling), these methods utilize the counts of sequencing reads of a given genomic locus as a proxy for the representation of that sequence in the original sample of nucleic acids. The majority of these techniques require a preparative step to construct a high-complexity library of DNA molecules that is representative of a sample of interest. This may include chemical or biochemical treatment of the DNA (e.g., bisulphite treatment), capture of a specific subset of the genome (e.g., padlock probe capture, solution hybridization), and a variety of amplification techniques (e.g., polymerase chain reaction, whole genome amplification, rolling circle amplification).

Systematic and random errors are common problems associated with genome amplification and sequencing library construction techniques. For example, genomic sequencing library may contain an over- or under-representation of particular sequences from a source genome as a result of errors (bias) in the library construction process. Such bias can be particularly problematic when it results in target sequences from a genome being absent or undetectable in the sequencing libraries. For example, an under representation of particular allelic sequences (e.g., heterozygotic alleles) from a genome in a sequencing library can result in an apparent homozygous representation in a sequencing library. As most downstream

Petitioner Illumina, Inc., EX1004, p. 684

sequencing library quantification techniques depend on stochastic counting processes, these problems have typically been addressed by sampling enough (over-sampling) to obtain a minimum number of observations necessary to make statistically significant decisions. However, the strategy of oversampling is generally limited to elimination of low-count Poisson noise, and the approach wastes resources and increases the expense required to perform such experiments. Moreover, oversampling can result in a reduced statistical confidence in certain conclusions (e.g., diagnostic calls) based on the data. Accordingly, new approaches are needed for overcoming bias in sequencing library preparatory methods.

Aspects of the disclosure are based, in part, on the discovery of methods for overcoming problems associated with systematic and random errors (bias) in genome capture, amplification and sequencing methods, namely high variability in the capture and amplification of nucleic acids and disproportionate representation of heterozygous alleles in sequencing libraries. Accordingly, in some embodiments, the disclosure provides methods that reduce variability in the capture and amplification of nucleic acids. In other embodiments, the methods improve allelic representation in sequencing libraries and, thus, improve variant detection outcomes. In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve the use of differentiator tag sequences to uniquely tag individual nucleic acid molecules. In some embodiments, the differentiator tag sequence permits the detection of bias based on the frequency with which pairs of differentiator tag and target sequences are observed in a sequencing reaction. In other embodiments, the methods reduce errors caused by bias, or the risk of bias, associated with the capture, amplification and sequencing of genetic loci, e.g., for diagnostic purposes.

Aspects of the invention relate to associating unique sequence tags (referred to as differentiator tag sequences) with individual target molecules that are independently captured and/or analyzed (e.g., prior to amplification or other process that may introduce bias). These tags are useful to distinguish independent target molecules from each other thereby allowing an analysis to be based on a known number of individual target molecules. For example, if each of a plurality of target molecule sequences obtained in an assay is associated with a different differentiator tag, then the target sequences can be considered to be independent of each other and a genotype likelihood can be determined based on this information. In contrast, if each of the plurality of target molecule sequences obtained in the assay is associated with the same

Petitioner Illumina, Inc., EX1004, p. 685

differentiator tag, then they probably all originated from the same target molecule due to over-representation (e.g., due to biased amplification) of this target molecule in the assay. This provides less information than the situation where each nucleic acid was associated with a different differentiator tag. In some embodiments, a threshold number of independently isolated molecules (e.g., unique combinations of differentiator tag and target sequences) is analyzed to determine the genotype of a subject.

In some embodiments, the invention relates to compositions comprising pools (libraries) of preparative nucleic acids that each comprise "differentiator tag sequences" for detecting and reducing the effects of bias, and for genotyping target nucleic acid sequences. As used herein, a "differentiator tag sequence" is a sequence of a nucleic acid (a preparative nucleic acid), which in the context of a plurality of different isolated nucleic acids, identifies a unique, independently isolated nucleic acid. Typically, differentiator tag sequences are used to identify the origin of a target nucleic acid at one or more stages of a nucleic acid preparative method. For example, in the context of a multiplex nucleic acid capture reaction, differentiator tag sequences provide a basis for differentiating between multiple independent, target nucleic acid capture events. Also, in the context of a multiplex nucleic acid amplification reaction, differentiator tag sequences provide a basis for differentiating between multiple independent, primary amplicons of a target nucleic acid, for example. Thus, combinations of target nucleic acid and differentiator tag sequence (target:differentiator tag sequences) of an isolated nucleic acid of a preparative method provide a basis for identifying unique, independently isolated target nucleic acids. FIG. 4A-C depict various non-limiting examples of methods for combining differentiator tag sequence and target sequences.

It will be apparent to the skilled artisan that differentiator tags may be synthesized using any one of a number of different methods known in the art. For example, differentiator tags may be synthesized by random nucleotide addition.

Differentiator tag sequences are typically of a predefined length, which is selected to control the likelihood of producing unique target:differentiator tag sequences in a preparative reaction (e.g., amplification-based reaction, a circularization selection-based reaction, e.g., a MIP reaction). Differentiator tag sequences may be, up to 5, up to 6, up to 7 up to 8, up to 9, up to 10, up to 11, up to 12, up to 13, up to 14, up to 15, up to 16, up to 17, up to 18, up to 19, up to 20, up to 21, up to 22, up to 23, up to 24, up to 25, or more nucleotides in length. For purposes of

Petitioner Illumina, Inc., EX1004, p. 686

genotyping, isolated nucleic acids are identified as independently isolated if they comprise unique combinations of target nucleic acid and differentiator tag sequences, and observance of threshold numbers of unique combinations of target nucleic acid and differentiator tag sequences provide a certain statistical confidence in the genotype.

During a library preparation process, each nucleic acid molecule may be tagged with a unique differentiator tag sequence in a configuration that permits the differentiator tag sequence to be sequenced along with the target nucleic acid sequence of interest (the nucleic acid sequence for which the library is being prepared, e.g., a polymorphic sequence). The incorporation of the nucleic acid comprising a differentiator tag sequence at a particular step allows the detection and correction of biases in subsequent steps of the protocol.

A large library of unique differentiator tag sequences may be created by using degenerate, random-sequence polynucleotides of defined length. The differentiator tag sequences of the polynucleotides may be read at the final stage of the sequencing. The observations of the differentiator tag sequences may be used to detect and correct biases in the final sequencing read-out of the library. For example, the total possible number of differentiator tag sequences, which may be produced, e.g., randomly, is 4^N, where N is the length of the differentiator tag sequence. Thus, it is to be understood that the length of the differentiator tag sequence may be adjusted such that the size of the population of MIPs having unique differentiator tag sequences is sufficient to produce a library of MIP capture products in which identical independent combinations of target nucleic acid and differentiator tag sequence are rare. As used herein combinations of target nucleic acid and differentiator tag sequences, may also be referred to as "target:differentiator tag sequences".

In the final readout of a sequencing process, each read may have an additional unique differentiator tag sequence. In some embodiments, when differentiator tag sequences are distributed randomly in a library, all the unique differentiator tag sequences will be observed about an equal number of times. Accordingly, the number of occurrences of a differentiator tag sequence may follow a Poisson distribution.

In some embodiments, overrepresentation of target:differentiator tag sequences in a pool of preparative nucleic acids (e.g., amplified MIP capture products) is indicative of bias in the preparative process (e.g., bias in the amplification process). For example, target:differentiator tag sequence combinations that are statistically overrepresented are indicative of bias in the

Petitioner Illumina, Inc., EX1004, p. 687

protocol at one or more steps between the incorporation of the differentiator tag sequences into MIPs and the actual sequencing of the MIP capture products.

The number of reads of a given target:differentiator tag sequence may be indicative (may serve as a proxy) of the amount of that target sequence present in the originating sample. In some embodiments, the numbers of occurrence of sequences in the originating sample is the quantity of interest. For example, using the methods disclosed herein, the occurrence of differentiator tag sequences in a pool of MIPs may be predetermined (e.g., may be the same for all differentiator tag sequences). Accordingly, changes in the occurrence of differentiator tag sequences after amplification and sequencing may be indicative of bias in the protocol. Bias may be corrected to provide an accurate representation of the composition of the original MIP pool, e.g., for diagnostic purposes.

According to some aspects, a library of preparative nucleic acid molecules (e.g., MIPs, each nucleic acid in the library having a unique differentiator tag sequence, may be constructed such that the number of nucleic acid molecules in the library is significantly larger than the number prospective target nucleic acid molecules to be captured using the library. This ensures that products of the preparative methods include only unique target:differentiator tag sequence; e.g., in a MIP reaction the capture step would undersample the total population of unique differentiator tag sequences in the MIP library. For example, an experiment utilizing 1 ug of genomic DNA will contain about ~150,000 copies of a diploid genome. For a MIP library, each MIP in the library comprising a randomly produced 12-mer differentiator tag sequence (~1.6 million possible unique differentiator tag sequences), there would be more than 100 unique differentiator tag sequences per genomic copy. For a MIP library, each MIP in the library comprising a randomly produced 15-mer differentiator tag sequence (~1 billion possible unique differentiator tag sequences), there would be more than 7000 unique differentiator tag sequences per genomic copy. Therefore, the probability of the same differentiator tag sequence being incorporated multiple times is incredibly small. Thus, it is to be appreciated that the length of the differentiator tag sequence is to be selected based on the amount of target sequence in a MIP capture reaction and the desired probability for having multiple, independent occurrences of target:differentiator tag sequence combinations.

Petitioner Illumina, Inc., EX1004, p. 688

FIG. 5 depicts a non-limiting method for genotyping based on target and differentiator tag sequences. Sequencing reads of target and differentiator tags sequences are collapsed to make diploid genotype calls. FIG. 6 depicts non-limiting results of a simulation of a MIP capture reaction in which MIP probes, each having a differentiator tag sequence of 15 nucleotides, are combined with 10000 target sequence copies (e.g., genome equivalents). In this simulated reaction, the probability of capturing one or more copies of a target sequence having the same differentiator tag sequence is 0.05. The Y axis reflects the number of observations. The X axis reflects the number of independent occurrences of target:differentiator tag combinations. FIG. 7 depicts a non-limiting graph of sequencing coverage, which can help ensure that alleles are sampled to sufficient depth (e.g., either 10x or 20x minimum sampling per allele, assuming 1000 targets). In this non-limiting example, the X axis is total per-target coverage required, and the Y axis is the probability that a given total coverage will result in at least 10x or 20x coverage for each allele.

The skilled artisan will appreciate that as part of a MIP library preparation process, adapters may be ligated onto the ends of the molecules of interest. Adapters often contain PCR primer sites (for amplification or emulsion PCR) and/or sequencing primer sites. In addition, barcodes may be included, for example, to uniquely identify individual samples (e.g., patient samples) that may be mixed together. (See, e.g., USPTO Publication Number US 2007/0020640 Al (McCloskey et al.)

The actual incorporation of the random differentiator tag sequences can be performed through various methods known in the art. For example, nucleic acids comprising differentiator tag sequences may be incorporated by ligation. This is a flexible method, because molecules having differentiator tag sequence can be ligated to any blunt-ended nucleic acids. The sequencing primers must be incorporated subsequently such that they sequence both the differentiator tag sequence and the target sequence. Alternatively, the sequencing adaptors can be synthesized with the random differentiator tag sequences at their 3' end (as degenerate bases), so that only one ligation must be performed. Another method is to incorporate the differentiator tag sequence into a PCR primer, such that the primer structure is arranged with the common adaptor sequence followed by the random differentiator tag sequence followed by the PCR priming sequence (in 5' to 3' order). A differentiator tag sequence and adaptor sequence (which may contain the sequencing primer site) are incorporated as tags. Another method to incorporate

Petitioner Illumina, Inc., EX1004, p. 689

the differentiator tag sequences is to synthesize them into a padlock probe prior to performing a gene capture reaction. The differentiator tag sequence is incorporated 3' to the targeting arm but 5' to the amplification primer that will be used downstream in the protocol. Another method to incorporate the differentiator tag sequences is as a tag on a gene-specific or poly-dT reverse-transcription primer. This allows the differentiator tag sequence to be incorporated directly at the cDNA level.

In some embodiments, at the incorporation step, the distribution of differentiator tag sequences can be assumed to be uniform. In this case, bias in any part of the protocol would change the uniformity of this distribution, which can be observed after sequencing. This allows the differentiator tag sequence to be used in any preparative process where the ultimate output is sequencing of many molecules in parallel.

Differentiator tag sequences may be incorporated into probes (e.g., MIPs) of a plurality when they are synthesized on-chip in parallel, such that degeneracy of the incorporated nucleotides is sufficient to ensure near-uniform distribution in the plurality of probes. It is to be appreciated that amplification of a pool of unique differentiator tag sequences may itself introduce bias in the initial pool. However, in most practical cases, the scale of synthesis (e.g., by column synthesis, chip based synthesis, etc.) is large enough that amplification of an initial pool of differentiator tag sequences is not necessary. By avoiding amplification or selection steps on the pool of unique differentiator tag sequences, potential bias may be minimized.

One example of the use of the differentiator tag sequences is in genome re-sequencing. Considering that the raw accuracy of most next-generation sequencing instruments is relatively low, it is crucial to oversample the genomic loci of interest.

Furthermore, since there are two alleles at every locus, it is important to sample enough to ensure that both alleles have been observed a sufficient number of times to determine with a sufficient degree of statistical confidence whether the sample is homozygous or heterozygous. Indeed, the sequencing is performed to sample the composition of molecules in the originating sample. However, after multiple reads have been collected for a given locus, it is possible that due to bias (e.g., caused by PCR amplification steps), a large fraction of the reads are derived from a single originating molecule. This would skew the population of target sequences observed, and would affect the outcome of the genotype call. For example, it is possible that a locus that is heterozygous is called as homozygous, because there are only a few observations of

Petitioner Illumina, Inc., EX1004, p. 690

the second allele out of many observations of that locus. However, if information is available on differentiator tag sequences, this situation could be averted, because the over-represented allele would be seen to also have an over-represented differentiator tag sequence (i.e., the sequences with the overrepresented differentiator tag sequence all originated from the same single molecule). Therefore, the sequences and corresponding distribution of differentiator tag sequences can be used as an additional input to the genotype-calling algorithm to significantly improve the accuracy and confidence of the genotype calls.

In some aspects, the disclosure provides methods for analyzing a plurality of target sequences which are genetic loci or portions of genetic loci (e.g., a genetic locus of Table 1). The genetic loci may be analyzed by sequencing to obtain a genotype at one or more polymorphisms (e.g., SNPs). Exemplary polymorphisms are disclosed in Table 2. The skilled artisan will appreciate that other polymorphisms are known in the art and may

Petitioner Illumina, Inc., EX1004, p. 691

be identified, for example, by querying the Entrez Single Nucleotide Polymorphism database, for example, by searching with a GeneID from Table 1.

Table 1: Target Nucleic Acids

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
CYP21A2	1589	cytochrome P450, family 21, subfamily A, polypeptide 2	CAH1; CPS1; CA21H; CYP21; CYP21B; P450c21B; MGC150536; MGC150537; CYP21A2	201910	NG_008337.1	6p21.3
ABCC8	6833	ATP-binding cassette, sub-family C (CFTR/MRP), member 8	HI; SUR; HHF1; MRP8; PHHI; SUR1; ABC36; HRINS; TNDM2; ABCC8	600509	NG_008867.1	11p15.1
ATRX	546	alpha thalassemia/mental retardation syndrome X-linked (RAD54 homolog, S. cerevisiae)	SHS; XH2; XNP; ATR2; SFM1; RAD54; MRXHF1; RAD54L; ZNF-HX; MGC2094; ATRX	300032	NG_008838.1	Xq13.1-q21.1
ARSA	410	arylsulfatase A	MLD; ARSA	607574	NG_009260.1	22q13.31-qter; 22q13.33
PSAP	5660	Prosaposin	GLBA; SAP1; FLJ00245; MGC110993; PSAP	176801	NG_008835.1	10q21-q22
BTD	686	Biotinidase	BTD	609019	NG_008019.1	3p25

5

Petitioner Illumina, Inc., EX1004, p. 692

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
HLCS	3141	holocarboxylase synthetase (biotin-(proprionyl-Coenzyme A-carboxylase (ATP-hydrolysing)) ligase)	HCS; HLCS	609018	NC_000021.7	21q22.1; 21q22.13
BLM	641	Bloom syndrome, RecQ helicase-like	BS; RECQ2; RECQL2; RECQL3; MGC126616; MGC131618; MGC131620; BLM	604610	NG_007272.1	15q26.1
ASPA	443	aspartoacylase (Canavan disease)	ASP; ACY2; ASPA	608034	NG_008399.1	17pter-p13
CFTR	1080	cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7)	CF; MRP7; ABC35; ABCC7; CFTR/MRP; TNR-CFTR; dJ760C5.1; CFTR	602421	NC_000007.12	7q31.2
ASS1	445	argininosuccinate synthetase 1	ASS; CTLN1; ASS1	603470	NG_011542.1	9q34.1
MMACHC	25974	methylmalonic aciduria (cobalamin deficiency) cblC type, with homocystinuria	cblC; FLJ25671; DKFZp564I122; RP11-291L19.3; MMACHC	609831	NC_000001.9	1p34.1
IKBKAP	8518	inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase complex-associated protein	FD; DYS; ELP1; IKAP; IKI3; TOT1; FLJ12497; DKFZp781H1425; IKBKAP	603722	NG_008788.1	9q31
FANCC	2176	Fanconi anemia, complementation group C	FA3; FAC; FACC; FLJ14675; FANCC	227645	NG_011707.1	9q22.3
GALK1	2584	galactokinase 1	GK1; GALK; GALK1	604313	NG_008079.1	17q24
GALT	2592	galactose-1-phosphate uridylyltransferase	GALT	606999	NC_000009.10	9p13
GALE	2582	UDP-galactose-4-epimerase	SDR1E1; FLJ95174; FLJ97302; GALE	606953	NG_007068.1	1p36-p35
GBA	2629	glucosidase, beta; acid (includes glucosylceramidase)	GCB; GBA1; GLUC; GBA	606463	NG_009783.1	1q21
GJB2	2706	gap junction protein, beta 2, 26kDa	HID; KID; PPK; CX26; DFNA3; DFNB1; NSRD1; DFNA3A; DFNB1A; GJB2	121011	NG_008358.1	13q11-q12
GCDH	2639	glutaryl-Coenzyme A dehydrogenase	GCD; ACAD5; GCDH	608801	NG_009292.1	19p13.2
G6PC	2538	glucose-6-phosphatase, catalytic subunit	G6PT; GSD1; GSD1a; MGC163350; G6PC	232200	NG_011808.1	17q21
HBB	3043	hemoglobin, beta	CD113t-C; beta-globin; HBB	141900	NG_000007.3	11p15.5

Petitioner Illumina, Inc., EX1004, p. 693

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
BCKDHA	593	branched chain keto acid dehydrogenase E1, alpha polypeptide	MSUD1; OVD1A; BCKDE1A; FLJ45695; BCKDHA	608348	NC_000019.8	19q13.1-q13.2
BCKDHB	594	branched chain keto acid dehydrogenase E1, beta polypeptide	E1B; FLJ17880; dJ279A18.1; BCKDHB	248611	NG_009775.1	6q13-q15
DBT	1629	dihydrolipoamide branched chain transacylase E2	E2; E2B; BCATE2; MGC9061; DBT	248610	NG_011852.1	1p31
DLD	1738	dihydrolipoamide dehydrogenase	E3; LAD; DLDH; GCSL; PHE3; DLD	238331	NG_008045.1	7q31-q32
ACADM	34	acyl-Coenzyme A dehydrogenase, C-4 to C-12 straight chain	MCAD; ACAD1; MCADH; FLJ18227; FLJ93013; FLJ99884; ACADM	607008	NG_007045.1	1p31
MMAA	166785	methylmalonic aciduria (cobalamin deficiency) cblA type	cblA; MGC120010; MGC120011; MGC120012; MGC120013; MMAA	607481	NG_007536.1	4q31.22
MMAB	326625	methylmalonic aciduria (cobalamin deficiency) cblB type	ATR; cblB; MGC20496; MMAB	607568	NG_007096.1	12q24
MUT	4594	methylmalonyl Coenzyme A mutase	MCM; MUT	609058	NG_007100.1	6p12.3
MCOLN1	57192	mucolipin 1	ML4; MLIV; MST080; TRPML1; MSTP080; TRP-ML1; TRPM-L1; MCOLN1	605248	NC_000019.8	19p13.3-p13.2
ACTA1	58	actin, alpha 1, skeletal muscle	ACTA; ASMA; CFTD; MPFD; NEM1; NEM2; NEM3; CFTD1; CFTDM; ACTA1	102610	NG_006672.1	1q42.13
TPM3	7170	tropomyosin 3	TM3; TRK; NEM1; TM-5; TM30; TM30nm; TPMsk3; hscp30; MGC3261; FLJ41118; MGC14582; MGC72094; OK/SW-cl.5; TPM3	191030	NG_008621.1	1q21.2
TNNT1	7138	troponin T type 1 (skeletal, slow)	ANM; TNT; STNT; TNTS; FLJ98147; MGC104241; TNNT1	191041	NG_011829.1	19q13.4
NEB	4703	nebulin	NEM2; NEB177D; FLJ11305; FLJ36536; FLJ39568; FLJ39584; DKFZp686C1456; NEB	161650	NG_009382.1	2q22
SMPD1	6609	sphingomyelin phosphodiesterase 1, acid lysosomal	ASM; NPD; SMPD1	607608	NG_011780.1	11p15.4-p15.1
GLDC	2731	glycine dehydrogenase (decarboxylating)	GCE; NKH; GCSP; HYGN1; MGC138198; MGC138200; GLDC	238300	NC_000009.10	9p22

Petitioner Illumina, Inc., EX1004, p. 694

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
GCSH	2653	glycine cleavage system protein H (aminomethyl carrier)	GCE; NKH; GCSH	238330	NC_000016.8	16q23.2
AMT	275	aminomethyltransferase	GCE; NKH; GCST; AMT	238310	NC_000003.10	3p21.2-p21.1
OTC	5009	ornithine carbamoyltransferase	OCTD; MGC129967; MGC129968; MGC138856; OTC	300461	NG_008471.1	Xp21.1
PAH	5053	phenylalanine hydroxylase	PH; PKU; PKU1; PAH	612349	NG_008690.1	12q22-q24.2
DHPR	5860	quinoid dihydropteridine reductase	DHPR; PKU2; SDR33C1; FLJ42391; QDPR	612676	NG_008763.1	4p15.31
PTS	5805	6-pyruvoyltetrahydropterin synthase	PTPS; FLJ97081; PTS	261640	NG_008743.1	11q22.3-q23.3
PCCA	5095	propionyl Coenzyme A carboxylase, alpha polypeptide	PCCA	232000	NG_008768.1	13q32
PCCB	5096	propionyl Coenzyme A carboxylase, beta polypeptide	DKFZp451E113; PCCB	232050	NG_008939.1	3q21-q22
ACADS	35	acyl-Coenzyme A dehydrogenase, C-2 to C-3 short chain	SCAD; ACAD3; ACADS	606885	NG_007991.1	12q22-qter
DHCR7	1717	7-dehydrocholesterol reductase	SLOS; DHCR7	602858	NC_000011.8	11q13.2-q13.5
SMN1	6606	survival of motor neuron 1, telomeric	SMA; SMN; SMA1; SMA2; SMA3; SMA4; SMA@; SMNT; BCD541; T-BCD541; SMN1	600354	NG_008691.1	5q13
HEXA	3073	hexosaminidase A (alpha polypeptide)	TSD; MGC99608; HEXA	606869	NG_009017.1	15q23-q24
MYO7A	4647	myosin VIIA	DFNB2; MYU7A; NSRD2; USH1B; DFNA11; MYOVIIA; MYO7A	276903	NG_009086.1	11q13.5
USH1C	10083	Usher syndrome 1C (autosomal recessive, severe)	PDZ73; AIE-75; DFNB18; PDZ-45; PDZ-73; NY-CO-37; NY-CO-38; ush1cpst; PDZ-73/NY-CO-38; USH1C	605242	NC_000011.8	11p15.1-p14
CDH23	64072	cadherin-like 23	USH1D; DFNB12; FLJ00233; FLJ36499; KIAA1774; KIAA1812; MGC102761; DKFZp434P2350; CDH23	605516	NG_008835.1	10q21-q22
PCDH15	65217	protocadherin 15	USH1F; DFNB23; DKFZp667A1711; PCDH15	605514	NG_009191.1	10q21.1

Petitioner Illumina, Inc., EX1004, p. 695

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
SANS	124590	Usher syndrome 1G (autosomal recessive)	SANS; ANKS4A; FLJ33924; USH1G	607696	NG_007882.1	17q25.1
ARX	170302	aristaless related homeobox	ISSX; PRTS; MRX29; MRX32; MRX33; MRX36; MRX38; MRX43; MRX54; MRX76; MRX87; MRXS1; ARX	300382	NG_008281.1	Xp21
OPHN1	4983	oligophrenin 1	OPN1; MRX60; OPHN1	300127	NG_008960.1	Xq12
JARID1C	8242	lysine (K)-specific demethylase 5C	MRXJ; SMCX; MRXSJ; XE169; JARID1C; DXS1272E; KDM5C	314690	NG_008085.1	Xp11.22-p11.21
FTSJ1	24140	FtsJ homolog 1 (E. coli)	JM23; MRX9; SPB1; TRM7; CDLIV; MRX44; FTSJ1	300499	NG_008879.1	Xp11.23
SLC6A8	6535	solute carrier family 6 (neurotransmitter transporter, creatine), member 8	CRT; CT1; CRTR; MGC87396; SLC6A8	300036	NC_000023.9	Xq28
DLG3	1741	discs, large homolog 3 (Drosophila)	MRX; MRX90; NEDLG; NE-Dlg; SAP102; SAP-102; KIAA1232; DLG3	300189	NC_000023.9	Xq13.1
TM4SF2	7102	tetraspanin 7	AI5; MXS1; CD231; MRX58; CCG-B7; TM4SF2; TALLA-1; TM4SF2b; DXS1692E; TSPAN7	300096	NG_009160.1	Xp11.4
ZNF41	7592	zinc finger protein 41	MRX89; MGC8941; ZNF41	314995	NG_008238.1	Xp11.23
FACL4	2182	acyl-CoA synthetase long-chain family member 4	ACS4; FACL4; LACS4; MRX63; MRX68; ACSL4	300157	NG_008053.1	Xq22.3-q23
PQBP1	10084	polyglutamine binding protein 1	SHS; MRX55; MRXS3; MRXS8; NPW38; RENS1; PQBP1	300463	NC_000023.9	Xp11.23
PEX1	5189	peroxisomal biogenesis factor 1	ZWS1; PEX1	602136	NG_008341.1	7q21.2
PXMP3	5828	peroxisomal membrane protein 3, 35kDa	PAF1; PEX2; PMP3; PAF-1; PMP35; RNF72; PXMP3	170993	NG_008371.1	8q21.1
PEX6	5190	peroxisomal biogenesis factor 6	PAF2; PAF-2; PXAAA1; PEX6	601498	NG_008370.1	6p21.1
PEX10	5192	peroxisomal biogenesis factor 10	NALD; RNF69; MGC1998; PEX10	602859	NG_008342.1	1p36.32
PEX12	5193	peroxisomal biogenesis factor 12	PAF-3; PEX12	601758	NG_008447.1	17q12
PEX5	5830	peroxisomal biogenesis factor 5	PXR1; PTS1R; PTS1-BP; FLJ50634; FLJ50721; FLJ51948; PEX5	600414	NG_008448.1	12p13.31
PEX26	55670	peroxisomal biogenesis factor 26	FLJ20695; PEX26M1T; Pex26pM1T; PEX26	608666	NG_008339.1	22q11.21

Petitioner Illumina, Inc., EX1004, p. 696

The mutations listed in Table 2 are documented polymorphisms in several disease-associated genes (CFTR is mutated in cystic fibrosis, GBA is mutated in Gaucher disease, ASPA is mutated in Canavan disease, HEXA is mutated in Tay Sachs disease). The polymorphisms are of several types: insertion/deletion polymorphisms which will cause frameshifts (and thus generally interrupt protein function) unless the insertion/deletion length is a multiple of 3bp, and substitutions which can alter the amino acid sequence of the protein and in some cases cause complete inactivation by introduction of a stop codon.

Table 2: Non-limiting examples of polymorphisms

Gene name	GeneID	SNP ID	Mutation	SEQ ID NO:
CFTR	1080	rs63500661	TCACATCACCAAGTTAAAAAAAAAAA[A/G]G GGGCGGGGGGGCAGAATGAAAATT	1
CFTR	1080	rs63107760	AAACAAGGATGAATTAAGTTTTTTTT[-/T]AAAAAAGAAACATTTGGTAAGGGGA	2
CFTR	1080	rs62469443	ATCACCAAGTTAAAAAAAAAAAAGGG[A/G]C GGGGGGGCAGAATGAAAATTGCAT	3
CFTR	1080	rs62469442	CTATTGAACCAGAACCAAACAGGAAT[A/G]C CATAGCATTTTGTAAACTAAACTG	4
CFTR	1080	rs62469441	CAGGAGTTCAAGACCAGCCTACTAAA[A/C]C ACACACACACACACACACACACAC	5
CFTR	1080	rs62469439	GATTAAATAATAGTGTTTATGTACCC[C/G]GC TTATAGGAGAAGAGGGTGTGTGT	6
CFTR	1080	rs62469438	ATTGTTATCTTTTCATATAAGGTAAC[A/T]GA GGCCCAGAGAGATTAAATAACAT	7
CFTR	1080	rs62469437	TAATTTTAATTAAGTAAATTTAATTG[A/G]TA GATAAATAAGTAGATAAAAAATA	8
CFTR	1080	rs62469436	GTATAAAAAAAAAAAAAAAAAAAGTT[A/T]G AATGTTTTCTTGCATTCAGAGCCT	9
CFTR	1080	rs62469435	ATACTAAAAATTTAAAGTTCTCTTGC[A/G]AT ATATTTTCTTAATATCTTACATC	10
CFTR	1080	rs62469434	TGCTGGGATTACAGGCGTGAGCCACC[A/G]C GCCTGGCCTGATGGGACATATTTT	11
CFTR	1080	rs62469433	CTACAATATAAGTATAGTATTGCAAA[A/C]CC ATCAGGAAGGGTGTTAACTATTT	12
CFTR	1080	rs61763210	GTTGTCTCCAAACTTTTTTTCAGGTG[-/AGA]AGGTGGCCAACCGAGCTTCGGAAAG	13
CFTR	1080	rs61720488	TTTTTTCATAAAAGATTATATAAAGG[A/C]TA TTGCTTTTGAATCACAAACACTA	14
CFTR	1080	rs61481156	ATCTAGTGAGCAGTCAGGAAAGAGAA[C/T]T TCCAGATCCTGGAAATCAGGGTTA	15
CFTR	1080	rs61443875	TAGAGTATAAAAAAAAAAAAAAAAA[-/A]GTTTGAATGTTTTCTTGCATTCAGA	16

Petitioner Illumina, Inc., EX1004, p. 697

CFTR	1080	rs61312222	TGCAAATGCCAACTATCAAAGATATT[C/G]GA GTATACTGTCAATAAACTTCATA	17
CFTR	1080	rs61159372	TCCTCAACAGTTAGAAACAATATTTT[C/G]AG TGATTTCCCATGCCAACTTTACT	18
CFTR	1080	rs61094145	TTTTTGGTATTGTTGTTAAATAAGTG[A/G]GA ATTCAATACAGTATAATGTCTGT	19
CFTR	1080	rs61086387	CTTGAAATCGGATATATATATATATATA[- /TGTATATATATATATATATATATATATAT ACATATATATATATA]GTATTATCCCTGTTTTC ACAGTTTT	20
CFTR	1080	rs60996744	AGAGGGGCTGTGAAGGACACCAAGGA[A/G]G AGACTAAGAGCCAGGAGGGAAAAC	21
CFTR	1080	rs60960860	TAGAGTTTATTAGCTTTTACTACTCT[A/G]CTT AGTTACTTTGTGTTACAGAATA	22
CFTR	1080	rs60923902	ACTAGTGATGATGAGCTTCTTTTCAT[- /AT]GTTTGTTGGCTGCATAAATGTCTTC	23
CFTR	1080	rs60912824	GCAGAGAAAAGAGGGGCTGTGAAGGA[C/G]A CCAAGGAGGAGACTAAGAGCCAGG	24
CFTR	1080	rs60887846	TTCAGAGGTCTACCACTGGTGCATAC[G/T]CT AATCACAGTGTCGAAAATTTTAC	25
CFTR	1080	rs60793174	AAGAAAGAGCAAAAGAGGGCAAACTT[C/T]T CATACATTTTTGATGTCGAAACCA	26
CFTR	1080	rs60788575	CCTAAAGTTTAAAAAGAAAAAAAAAA[- /A]GGAAGAAGGAATTAAAAATCCAAAG	27
CFTR	1080	rs60760741	GTGTGTGTGTGTATATATATATATATAT[A/T]TA TATATTTTTTTTTTCCTGAGCCA	28
CFTR	1080	rs60456599	AAACTGTTGATGTTTTCATTTATTTA[C/G]ATC ATTGGAAAACTTTAGATTCTAG	29
CFTR	1080	rs60363249	TTTATCCATTCTTAACCAGAACAGAC[A/G]TT TTTTCAGAGCTGGTCCAGGAAAA	30
CFTR	1080	rs60355115	TTGAAATCGGATATATATATATATATAT[A/G]TA TATATATATATATATATATATAT	31
CFTR	1080	rs60308689	TAGTTTTTTATTTCCTCATATTATTT[- /T]CAGTGGCTTTTTCTTCCACATCTTT	32
CFTR	1080	rs60271242	ACATAGTTCTCAGTGGTACAACTACA[A/G]GT GATTTCTCTTTTCTTATTTCTGG	33
CFTR	1080	rs60010318	AGAGCAATGGCATCCCTTGTCTTGTG[C/T]TA TACAGGATGCAGCAATTTATAGG	34
CFTR	1080	rs59961323	TTCTGTCTACATAAGATGTCATACTA[A/G]AT TATCTTTTCCAGCATGCATTCAG	35
CFTR	1080	rs59961270	CAGGGTGGCATGTTAGGCAGTGCTTA[A/G]A ATAAATGAGTTGGTTATACAAGTA	36
CFTR	1080	rs59837506	AGGACACACACACACACACACACACA[- /CA]TGCACACACATTTAAATAGATGCAT	37
CFTR	1080	rs59572090	TAAAAAATTGGTATAATGAAATTGCA[C/T]TT GTAGTCTTTGGACATTTAAATCC	38
CFTR	1080	rs59548252	TTTCAATACTTAAGAGGTACGCAGAG[A/G]A AAGAGGGGCTGTGAAGGACACCAA	39
CFTR	1080	rs59519859	CAGCAATGAATATTTTGAGGCTGAGG[C/T]GC TGAGGGGTAAAATTGCAGCCTGG	40
CFTR	1080	rs59509837	TTATGGTTTATATTTTTGTGTCTTCT[- /CTTT]AACACATCTTTTCTAGCAGAATTCA	41

Petitioner Illumina, Inc., EX1004, p. 698

CFTR	1080	rs59417037	GTATTTTAGTTTTTTTTTTTTGTTTG[-/T]TTTGTTTTGTTTTGTTTTGTTTTTG	42
CFTR	1080	rs59159458	TGGGTGACTCCATTTTTACTTTTAGT[C/T]TGG TCTGTTGAGGCCTCGTGAGAGA	43
CFTR	1080	rs59048119	TATTTTCATGTATTTTAGTTTTTTTTT[-/TTTT]GTTTGTTTTGTTTTGTTTTGTTTTG	44
CFTR	1080	rs58970500	GTGTGTGTGTATATATATATATATAT[A/T]TA TATTTTTTTTTTCCTGAGCCAAA	45
CFTR	1080	rs58942292	AACCTATTAGCATGTCTGGCAGAAAA[-/A]TAGATACTTAATAAATTTCTTAAAT	46
CFTR	1080	rs58917054	GAGGCTTAGACAGTTTAAGTAACTCA[A/G]G CATGGTTACACAACTAGCTAGGGC	47
CFTR	1080	rs58837484	GTGTGAGTATTATGAGACCATATGTT[A/G]GG AGATTTTATTTGGTATTGAGGAT	48
CFTR	1080	rs58829491	GAAACCCCACCCCTTCTATAGTTTTC[C/T]CTT TAATATTTACAATGGAACCATT	49
CFTR	1080	rs58805195	CATATATATATAGTGTGTGTGTGTGT[A/G]TA TATATATATATATATATATTTTT	50
GBA	2629	rs60866785	CGAGCGAGAGAGAGAGAGAGAGAGAG[-/AG]GAGCCGGCGCGAGAACTACGCATGC	51
GBA	2629	rs60239603	GGCAGGTAATATCTAGTACCTTACTT[A/T]TA TTTCCTGAGCACATTCTACATTT	52
GBA	2629	rs56310840	GGCCAGGAATGGGAGTGCTTAGGTGC[A/G]G AGGTGGCACTGTTCCCGCAGCTGC	53
GBA	2629	rs41264927	TGAAAACTCCATCCCCTCAGGGTCAT[C/T]AG ATGAAGAGAAGACCACAGGGGTT	54
GBA	2629	rs41264925	TGTAGGTAAGGGTCACATGTGGGAGA[C/G]G CAGCTGTGGGTAGGTCAGCCCTGT	55
GBA	2629	rs36024691	CCAAGAAGGCGCCATTACACTCCAGC[-/C]TGGGCGACAGGGCGAGACTCCCTCA	56
GBA	2629	rs36024092	TGCCACACCCAGCTAATTTGTGTGTG[-/G]TATGTGTGTGTATGTATGTGTGTGT	57
GBA	2629	rs35682967	GTTCCTCCAGTAATTTTTTTTTTTTT[-/T]CGTTTTGAGACAGAGTCTTGCCCTG	58
GBA	2629	rs35033592	ATCATGCCCAGATAATTTTTTTTTTT[-/T]GTATTTTAGTAGACACAGGGTTTCA	59
GBA	2629	rs34732744	CGAGCGAGAGAGAGAGAGAGAGAGAG[-/AG]GAGCCGGCGCGAGAACTACGCATGC	60
GBA	2629	rs34620635	CCTGTGAGGGGCACATTCCTTAGTAG[-/C]TAAGGAGTTGGGGGTGTGAAGATCC	61
GBA	2629	rs34302637	ACAGGCTACTGGCTGGGCCCAGGCAA[-/A]GGGGGCCTTGGCAGGAAAAGTTCCT	62
GBA	2629	rs33949225	GCGAGAGAGAGAGAGAGAGAGAGAGG[-/AG]AGCCGGCGCGAGAACTACGCATGCG	63
GBA	2629	rs28678003	AAGAAGAAAAATAAAAAGAAAGTGGG[C/T]C AGACCGAGAGAACAGGAAGCCTGA	64
GBA	2629	rs28559737	AAGGACAAAGGCAAAGAGACAAAGGC[G/T]C AACACTGGGGGTCCCCAGAGAGTG	65
GBA	2629	rs28373017	TACCTAGTCACTTCCTGCCTCCATGG[C/T]GC AAAAGGGGATGGGTGTGCCTCTT	66
GBA	2629	rs12752133	CTCTTCCGAGGTTCCACCCTGAACAC[C/T]TT CCTGCTCCCTCGTGGTGTAGAGT	67

Petitioner Illumina, Inc., EX1004, p. 699

GBA	2629	rs12747811	TTCTGACTGGCAACCAGCCCCACTCT[C/T]TG GGAGCCCTCAGGAATGAACTTGC	68
GBA	2629	rs12743554	gctcagcctcccaggctggagtgcag[A/T]ggcgcgatctcggctcac cgcaacc	69
GBA	2629	rs12041778	CATGAACCACATCAAATGAGATTTAG[C/T]GG GAGTGGCACACACAGTCATGACC	70
GBA	2629	rs12034326	AAGCAGCCCTGGGGAGTCGGGGCGGG[A/G]C CTGGATTGGAAAAGAGACGGTCAC	71
GBA	2629	rs11558184	CTCCAAGTTCTGGGAGCAGAGTGTGC[A/G]G CTAGGCTCCTGGGATCGAGGGATG	72
GBA	2629	rs11430678	GTTCCTCCAGTAAttttttttttt[- /G/T]gtttgagacagagtcttgccctgt	73
GBA	2629	rs11264345	CTAGTACCTTACTTCCCTCAAGTTCA[A/T]TC ATCTCACAGATATTTCCTGAGCA	74
GBA	2629	rs10908459	aattagccgtgcggtggtggcgggtgc[C/T]gtaatcccacgtacttgg gaggct	75
GBA	2629	rs10796940	CCATGGCCAGCCGGGGAGGGGACGGG[A/C]A CACACAGACCCACACAGAGACTCA	76
GBA	2629	rs10668496	agcgagagagagagagagagagagag[- /AG]gagCCGGCGCGAGAACTACGCATGC	77
GBA	2629	rs7416991	CGTAGCAGTTAGCAGATGATAGGCGG[C/G/T] GAAATCTTATTTCACAGGGCATTAA	78
GBA	2629	rs4024049	CTGGCCCTGGTGACAGTGGGGCTGTG[C/T]GT GGGGCCAGAGCCTTCTCAGAGGT	79
GBA	2629	rs4024048	CAGATACTGGCCCTGGTGACAGTGGG[A/G]C TGTGCGTGGGGCCAGAGCCTTCTC	80
GBA	2629	rs4024047	GACAGATACTGGCCCTGGTGACAGTG[G/T]G GCTGTGCGTGGGGCCAGAGCCTTC	81
GBA	2629	rs3841430	GGCTCctctctctctctctctctc[- /TC]gctcgctctctcgctctctcgctct	82
GBA	2629	rs3754485	GTTTCAGACCAGCCTGGCCAACATAG[C/T]GA AACCCCATCTCTACTAAAAATAA	83
GBA	2629	rs3205619	AGTGGGCGATTGGATGGAGCTGAGTA[C/T]G GGGCCCATCCAGGCTAATCACACC	84
GBA	2629	rs2990227	CCGGGCTCCGTGAATGTTTGTCACAT[C/G]TC TGAAGAACGTATGAATTACATAA	85
GBA	2629	rs2990226	GAATCCCAACCCCGACGCTCGTCGCC[C/G]G GCTCCGTGAATGTTTGTCACATGT	86
GBA	2629	rs2990225	GCGAATCCCAACCCCGACGCTCGTCG[C/T]CG GGCTCCGTGAATGTTTGTCACAT	87
GBA	2629	rs2990224	TGGGCAGAAGTCAGGGTCCAAAGAAA[G/T]G GCAAAGAAAAGTGTcagtggctca	88
ASPA	443	rs63751297	TAAGAAAGACGTTTTTGATTTTTTTC[A/G]GA CTTCTCTGGCTCCACTACCCTGC	89
ASPA	443	rs62071301	CTGATTCCTGGCCAGGAGCGGTGGCT[C/T]AC GCCTGTAATCCCAGCGCTTTGGG	90
ASPA	443	rs62071300	TAAAAATGCTGATTCCTGGCCAGGAG[C/T]GG TGGCTCACGCCTGTAATCCCAGC	91
ASPA	443	rs62071299	TTTAAAAATGCTGATTCCTGGCCAGG[A/C]GC GGTGGCTCACGCCTGTAATCCCA	92
ASPA	443	rs62071297	CAAGACCTGTCAAAGATCTGAGAAAT[A/T]TT ACCCGACTTACAAGCTAACCATT	93

Petitioner Illumina, Inc., EX1004, p. 700

ASPA	443	rs61697033	ACTGTAATAAGTGCTGTAAAAGAAAT[A/G]C ACAAAATAATATAGCAGAGGGTAT	94
ASPA	443	rs60743592	CTTGAGGTCAGGAGTTCAAGACCAGT[C/T]TG GGCAACATGGGGAAAACCTTGTC	95
ASPA	443	rs60666840	AGGTTGCAGTGAGCCGAGATCATGCC[A/G]TT GCACTCCAGCCGGGGCAACAAAA	96
ASPA	443	rs60147514	ACAAGTGTCTTGAAATTATCTGTGAT[C/T]TG CTATAGAGCAATACTTTTGTAAA	97
ASPA	443	rs59930743	GTGGGTATATGCAGCTCTATGCACTA[C/T]CT GCTCATTTATTTGGTAAATCTAA	98
ASPA	443	rs59690349	TGTGTGTGTGTGCGTGTGTGTGTGTG[-/TGTGTGTG]ATCATAAGAGTGGCTGCAGCAA ACT	99
ASPA	443	rs59676360	AGTCTGGAGTGCAATGGTGCAATCTC[A/G]GC TCACTGCAGCCTCCACCTCCGGG	100
ASPA	443	rs59335404	CTCCTAATGGATATTTCCTAAATTTT[G/T]CTG AACAGAATTTAACTTGAGCTGG	101
ASPA	443	rs58879097	ATTTAAAAATGGATTTCTAGAAAAAC[A/G]AT CACATACTTGAATATTTTAGCAA	102
ASPA	443	rs58686774	CTATAAATGGGTAGCATGAGGGATTC[A/G]A GGAGGTGGCTGAAAGAAGCACGTA	103
ASPA	443	rs57511162	AAGAAACCAAGCATAGTAGAGTGTTA[A/G]A AAACCAAAGCAACTAAACAACTGT	104
ASPA	443	rs55859596	CGGGGCTCAGAACTTGTAACAGAAAA[A/T]T AAAATATACTCCACTCAAGGGAAT	105
ASPA	443	rs55742972	TACTACACTTCACGGATACTGTACTT[-/GTACTT]TTTTTCCAAATTGAAGGTTTTTGGC	106
ASPA	443	rs55640436	TTGTTTTTGTTTTTGTTTTTGTTTTT[-/GTTTTTGTTTTT]TGAGATGGAGTCTCGCTCT GTCGCC	107
ASPA	443	rs36225687	TTTGCCTTACTACACTTCACGGATAC[-/TGTACT]TGTACTTTTTTTCCAAATTGAAGGT	108
ASPA	443	rs36051310	GAGGTGGCTGAAAGAAGCACGTATCC[-/C]TGATGGCATGGTTGCGGGTTATATG	109
ASPA	443	rs36034906	GAGAAAAGCAGTTCCTGGAACACCCC[-/C]ACCCCTTAACCCCTTATCTCTGCTT	110
ASPA	443	rs36033666	TTACATATGTATACATGTGCCATGTT[-/T]GGTGTGCCGCACCCATTAACTCGTC	111
ASPA	443	rs35730123	CTTTTTCCAGATTTTTTTTTTTTTTT[-/T]GAGACAGAGTTTCACTCTTGTTGCC	112
ASPA	443	rs35629100	TTTGGAAATCTTAAGCTTTTATTTGG[-/G]TGTCACAGAGAAACAGGATCTGTAT	113
ASPA	443	rs35614631	TACTTTAAGTTTTAGGGTACATGTGC[-/A]CCATGTGCAGGTTTGTTACATATGT	114
ASPA	443	rs35225782	ATTCATGACCAGCCACATAAATGCAC[-/A]GTATTACTTCGCAAGCATGCCAATG	115
ASPA	443	rs35178659	GTGCACTAGAATTAGCTAAAGTGGGG[-/G]AAAAAAAGATGCATTTGATGGTCTA	116
ASPA	443	rs35095578	AACCTCCACCTCCCAGGTTCAAGAGA[-/A]TTCTCCTGCCTCAGCCTCCCAAGTA	117
ASPA	443	rs35002210	CCTCCCTGTGATCCGAAGTAGCAGAC[A/G]TA CTTAACTTCCATGGTGGATTGTT	118

Petitioner Illumina, Inc., EX1004, p. 701

ASPA	443	rs34744839	AAAACATTATTATATCTAGAAAAAAA[-/A]TGTATCTTAACCATTGTGGGAAGTG	119
ASPA	443	rs34680506	TTGAAGGTAAAATCATAGGGAGTTGG[-/G]AGCTGTCCTCTTGCGCTGAATCAGT	120
ASPA	443	rs34365618	ACTTGTGGCCTTTTTGGAGAGGTTAG[-/CA]ACTCTGAAAACTCTGTCCCTGGACC	121
ASPA	443	rs34275920	GAAGGAGAAAAAGAGAGGAAATAAGT[-/T]AAAATAATAAACACAATTAATAAAG	122
ASPA	443	rs34109510	TGTATACATGTGCCATGTTGGTGTGC[C/T]GC ACCCATTAACTCGTCATTTAGCA	123
ASPA	443	rs34054576	TCACCTGTCACCTCCTATAGAACTTT[-/C]CCCTGACCCTCCTCTATAGCATTAA	124
ASPA	443	rs34015272	ATAAATGATCATCATTCACAGTAGGG[-/G]TTTTGTTTTGTTTTTTTTCTGGAAA	125
ASPA	443	rs34002091	ACAGACATATCTACAAACACACTTTT[-/T]CACATATTTGTGTAAGTCATTTATG	126
ASPA	443	rs28940574	AAAGACAACTAAACTAACGCTCAATG[A/C]A AAAAGTATTCGCTGCTGTTTACAT	127
ASPA	443	rs28940279	TACCGTGTACCCCGTGTTTGTGAATG[A/C]GG CCGCATATTACGAAAAGAAAGAA	128
ASPA	443	rs17850703	CAGGGCTGGAGGTAAAACCATTTATT[A/G]CT AACCCCAGAGCAGTGAAGAAGTG	129
ASPA	443	rs17222495	TTCTTCATTGCCTATTGAAGAGAGAG[C/T]GG AATGCTTTGGTTGCCAGATATGG	130
ASPA	443	rs17175228	CACAAGATCTCATTACTCAGGAGCTG[C/T]CC AAGTGTCTAATGTACTTAGTTAA	131
ASPA	443	rs16953074	TTCTGTGTAACATTTCATTTAAGCAA[A/G]GG ATTCGGCAAATCAAAAATTGTCA	132
ASPA	443	rs16953070	TAAAACGTATTGAAGGTATTATTGAC[G/T]CT GTTGAAGCAAAGAGAACAAAACA	133
HEXA	3073	rs62022858	ATCTGCTCTTCCAGTTGGATGACAAG[C/T]CT TGCTGTCTAACACCTGCTGCAGA	134
HEXA	3073	rs62022857	CCATTTTTTGTTGTATTTTTTTTTTC[C/T]TGAA TACTTTTTATCGCAGTTGGTT	135
HEXA	3073	rs62017872	CCCTGTCTCTAAAAGAAAAAAAAAAA[A/G]A AAAAAAAAAGAAAACAAAACCCAA	136
HEXA	3073	rs62017871	AGTGGCTCCAAAAAGGTCATGGAACC[C/T]CT TGAGGATGATGCAAATTGACTCT	137
HEXA	3073	rs61662730	TAAAGTTACTTTTCTTTTATTGACTT[C/T]CCC TTATTTTTTAACCTTATGCTTT	138
HEXA	3073	rs61329913	CAGAGTTAAAAAAAAAAAAAAAAAAA[-/A]GGAAGTAGCAGCAACAGCTTGGAAA	139
HEXA	3073	rs60920713	GTTGCCCAGGGTTGAGTGCAGAGGCA[C/T]AT CTTGGCTCACAGCAACCTCTGCC	140
HEXA	3073	rs60783213	AAGGCTTTTTTTTTTTTTTTTTTTT[-/TTTT]GAGACAGAGTCTTGCTGTGTCACCC	141
HEXA	3073	rs60644867	GCCTACATTCTGCAAAGAGGAGGGAA[C/G]A TTCACAGCTCCATACTTGAACCCT	142
HEXA	3073	rs60288568	CCAAAGGAGAATAGCTCTAGGGGAGG[C/G]A GGTGGATGAGTATGCATGGGGGAG	143
HEXA	3073	rs59888548	GACTCCATCTCAAAAAAAAAAAAAAA[-/A]TGCAGTCTAATGGCAGAATTAGACT	144

Petitioner Illumina, Inc., EX1004, p. 702

HEXA	3073	rs59733856	TTATTTATTTATTTATTTATTTTTGA[A/G]ACA GGGTCTCTGTTGTCCAGGCTGG	145
HEXA	3073	rs59427837	TTTTGAGGCAGGGTCTCACTCTGTTG[C/T]CC AGGGTTGAGTGCAGAGGCACATC	146
HEXA	3073	rs59171976	CGCCTTGCGAAGGCCCCACAGCTTGC[C/T]TG TGACAAACGTTCATAGGCAAATG	147
HEXA	3073	rs58706602	GGAGGTCTGTACAAAGCACCACCTAC[C/T]TC ATGGGTCAGTTTCCACAGCAGAA	148
HEXA	3073	rs58696963	GAATCTTATAATTCACTGTGTACCTC[-/CCTC]TGTTTCATATTTTCGCAATTGAACT	149
HEXA	3073	rs58610850	AACATAGTATCTAATATAGCTTTACA[C/T]CC AAAGCCAAAATATGAATACACTG	150
HEXA	3073	rs58016062	TTGTTTTGTTTTGTTTGGGGGGGGGG[-/G]TTGTTTTTCTGAGAGGGAGTCTTGC	151
HEXA	3073	rs57733983	CATACCAAAGGGCAGCTGGAGGGATA[C/T]A GACGGAAGTCATGTGGAGAGTGAA	152
HEXA	3073	rs57476645	CAGGTGTGAGCCACCACGACCACCAA[A/T]T TAGCTCTTTTTACTCCTTCCCTTC	153
HEXA	3073	rs56870003	AGTGGTAGCTGATTTTGCTTCTGGAT[A/C]CT TTGCCACCTTCCCACTCTTTAAT	154
HEXA	3073	rs56338339	AAAGACCTGTTTCTTAAAAAAAAAAA[-/AGAAAAAAAAAAA]GAAAGAAAAGAAAAG AAAAAAACAG	155
HEXA	3073	rs55995352	TAAAAAATCTTTCAATGAGGAGATGT[C/T]CC CAGAGCAAGACAGCTGTAGGATG	156
HEXA	3073	rs55860138	AAAAGAAAAAAAAAAAAAAAAAAAA[-/A]GAAAACAAAACCCAAACCCATAAAG	157
HEXA	3073	rs55743646	CCTGTCTCTAAAAGAAAAAAAAAAAA[A/G]A AAAAAAAAGAAAACAAAACCCAAA	158
HEXA	3073	rs55665666	GTTATCATAGAAAAATATCACACTCT[-/GT]CTGTATCCCCACTTCCAGAAACTGT	159
HEXA	3073	rs36106892	CAGGAGCTCATAGAATTACATACAAT[-/C]TTTTTTTTTTTTTTTTGAGACAGCG	160
HEXA	3073	rs36091525	TTGAGAATCTTATAATTCACTGTGTA[-/CCTC]CCTCTGTTTCATATTTTCGCAATTG	161
HEXA	3073	rs35949555	CCACTACCACAGTGCCTAGAGAACAA[C/T]A TGTGTTTAATAATATTTAAATAAT	162
HEXA	3073	rs35827424	CCCTGTCTCTAAAAGAAAAAAAAAAA[-/A]AAAAAAAAAAGAAAACAAAACCCAA	163
HEXA	3073	rs35729578	CCATTATATCATTCATTTCCCACTCA[-/T]TTTCTTCATTCCAACCAAGATATAT	164
HEXA	3073	rs35649102	TCCGTCTCAAAAAAAAAAAAAAAAAG[-/A]GAAAGGAATTATTCTCATGTATACA	165
HEXA	3073	rs35118677	CTGGGGCAGTTAAAAAGAAAAACAAA[-/C]CCCTGGTCCCTGCCCTTGAGGAGAT	166
HEXA	3073	rs35005352	CTCCAGGGTCCCATTCCAGGACCACA[-/C]GCCTGCTACCTCTGCAGCTCACTCA	167
HEXA	3073	rs34736306	GGATTGACATATACCAGTTAGACGGA[-/T]TTTTTTTTTCCATAAACCAGGCTCA	168
HEXA	3073	rs34607939	ACAAATAATTACTACATATCTACAAC[A/G]TT CCAGATACAGAAGAAATGGCCAA	169

Petitioner Illumina, Inc., EX1004, p. 703

HEXA	3073	rs34496117	TAAACACACTTGAAACATCATATAAA[-/ATG]ATATTACTACAAGACTTAACCGTAA	170
HEXA	3073	rs34300017	ACACAGGTAATCCATGTTTATTATAG[-/A]AAAATGCCACATTACTCTTTATTGA	171
HEXA	3073	rs34206496	AGTTATCATAGAAAAATATCACACTC[-/TG]TCTGTATCCCCACTTCCAGAAACTG	172
HEXA	3073	rs34110830	AATGAACTTACAGGAAGGTAATATAT[-/G]GGAAATAAACATCTTATTGAATTTA	173
HEXA	3073	rs34093438	GGACCCCTGAAAGGCACAAGACACCC[-/T]TTCAGGTTCACACTTCCTGAAAGCT	174
HEXA	3073	rs34085965	CCACCAATCACCAGAGCCTTCTGCTC[A/G]GG GGTACCTGAGGGAAAACAAGCAA	175
HEXA	3073	rs34004907	AAAGACTGAAAAAACATTCATAACTA[-/T]TTTTCTTGTTATCCTCGGAAATGTC	176
HEXA	3073	rs28942072	TATCTTCATCTTGGAGGAGATGAGGT[C/T]GA TTTCACCTGCTGGAAGTCCAACC	177
HEXA	3073	rs28942071	TTGCCTATGAACGTTTGTCACACTTC[C/T]GCT GTGAGTTGCTGAGGCGAGGTGT	178
HEXA	3073	rs28941771	GCTTGCTGTTGGATACATCTCGCCAT[C/T]AC CTGCCACTCTCTAGCATCCTGGA	179
HEXA	3073	rs28941770	CCGGGGCTTGCTGTTGGATACATCTC[G/T]CC ATTACCTGCCACTCTCTAGCATC	180

3. Nucleic acid target length evaluation:

In some embodiments, aspects of the invention relate to methods for detecting nucleic acid deletions or insertions in regions containing nucleic acid sequence repeats.

Genomic regions that contain nucleic acid sequence repeats are often the site of genetic instability due to the amplification or contraction of the number of sequence repeats (e.g., the insertion or deletion of one or more units of the repeated sequence). Instability in the length of genomic regions that contain high numbers of repeat sequences has been associated with a number of hereditary and non hereditary diseases and conditions.

For example, "Fragile X syndrome, or Martin-Bell syndrome, is a genetic syndrome which results in a spectrum of characteristic physical, intellectual, emotional and behavioral features which range from severe to mild in manifestation. The syndrome is associated with the expansion of a single trinucleotide gene sequence (CGG) on the X chromosome, and results in a failure to express the FMR-1 protein which is required for normal neural development. There are four generally accepted forms of Fragile X syndrome which relate to the length of the repeated CGG sequence; Normal (29-31 CGG repeats), Premutation (55-200 CGG repeats), Full

Petitioner Illumina, Inc., EX1004, p. 704

Mutation (more than 200 CGG repeats), and Intermediate or Gray Zone Alleles (40 - 60 repeats)."

Other examples include cancer, which has been associated with microsatellite instability (MSI) involving an increase or decrease in the genomic copy number of nucleic acid repeats at one or more microsatellite loci (e.g., BAT-25 and/or BAT-26). The are currently many sequencing-based assays for determining the number of nucleic acid sequence repeats at a particular locus and identifying the presence of nucleic acid insertions or deletions. However, such techniques are not useful in a high throughput multiplex analysis where the entire length of a region may not be sequenced.

In contrast, in some embodiments, aspects of the invention relate to detecting the presence of an insertion or deletion at a genomic locus without requiring the locus to be sequenced (or without requiring the entire locus to be sequenced). Aspects of the invention are particularly useful for detecting an insertion or deletion in a nucleic acid region that contains high levels of sequence repeats. The presence of sequence repeats at a genetic locus is often associated with relatively high levels of polymorphism in a population due to insertions or deletions of one or more of the sequence repeats at the locus. The polymorphisms can be associated with diseases or predisposition to diseases (e.g., certain polymorphic alleles are recessive alleles associated with a disease or condition). However, the presence of sequence repeats often complicates the analysis of a genetic locus and increases the risk of errors when using sequencing techniques to determine the precise sequence and number of repeats at that locus.

In some embodiments, aspects of the invention relate to determining the size of a genetic locus by evaluating the capture frequency of a portion of that locus suspected of containing an insertion or deletion (e.g., due to the presence of sequence repeats) using a nucleic acid capture technique (e.g., a nucleic acid sequence capture technique based on molecular inversion probe technology). According to aspects of the invention, a statistically significant difference in capture efficiency for a genetic locus of interest in different biological samples (e.g., from different subjects) is indicative of different relative lengths in those samples. It should be appreciated that the length differences may be at one or both alleles of the genetic locus. Accordingly, aspects of the invention may be used to identify polymorphisms regardless of whether biological samples being interrogated at heterozygous or homozygous for the

Petitioner Illumina, Inc., EX1004, p. 705

polymorphisms. According to aspects of the invention, subjects that contain one or more loci with an insertion or deletion can be identified by analyzing capture efficiencies for nucleic acids obtained from one or more biological samples using appropriate controls (e.g., capture efficiencies for known nucleic acid sizes, capture efficiencies for other regions that are not suspected of containing an insertion or deletion in the biological sample(s), or predetermined reference capture efficiencies, or any combination thereof. However, it should be appreciated that aspects of the invention are not limited by the nature or presence of the control. In some embodiments, if a statistically significant variation in capture efficiency is detected, a subject may be identified as being at risk for a disease or condition associated with insertions or deletions at that genetic locus. In some embodiments, the subject may be analyzed in greater detail in order to determine the precise nature of the insertion or deletion and whether the subject is heterozygous or homozygous for one or more insertions or deletions. For example, gel electrophoresis of an amplification (e.g., PCR) product of the locus, or Southern blotting, or any combination thereof can be used as an orthogonal approach to verify the length of the locus. In some embodiments, a more exhaustive and detailed sequence analysis of the locus can be performed to identify the number and types of insertions and deletions. However, other techniques may be used to further analyze a locus identified as having an abnormal length according to aspects of the invention.

Accordingly, aspects of the invention relate to detecting abnormal nucleic acid lengths in genomic regions of interest. In some embodiments, the invention aims to estimate the size of genomic regions that are hard to be accessed, such as repetitive elements. However, it should be appreciated that methods of the invention do not require that the precise length be estimated. In some embodiments, it is sufficient to determine that one or more alleles with abnormal lengths are present at a locus of interest (e.g., based on the detection of abnormal capture efficiencies).

In a non-limiting example, fragile X can be used to illustrate aspects of the invention where the size of trinucleotide repeats (genotype) is linked to a symptom (phenotype). However, it should be appreciated that fragile X is a non-limiting example and similar analyses may be performed for other genetic loci (e.g., independently or simultaneously in multiplex analyses).

Use of molecular inversion probes (MIPs) has been demonstrated for detection of single nucleotide polymorphisms (Hardenbol et al. 2005 Genome Res 15:269-75) and for preparative amplification of large sets of exons (Porreca et al. 2007 Nat Methods 4:931-6, Krishnakumar et

Petitioner Illumina, Inc., EX1004, p. 706

al. 2008 Proc Natl Acad Sci USA 105:9296-301). In both cases, oligonucleotide probes are designed which have ends ('targeting arms') that hybridize up-stream and down-stream of the locus that is to be amplified.

In some embodiments, aspects of the invention are based on the recognition that the effect of length on probe capturing efficiency can be used in the context of an assay (e.g., a high throughput and/or multiplex assay) to allow the length of sequences to be determined without requiring sequencing of the entire region being evaluated. This is particularly useful for repeat regions that are prone to changes in size. As illustrated in FIG. 8, which is reproduced from Deng et al., Nature Biotech. 27:353-60, (see Supplemental Fig. 1G of Deng et al.,) illustrates that shorter sequences are captured with higher efficiency that longer sequences using MIPs. The statistical package R and its effects module were used for this analysis. A linear model was used, and each individual factor was assumed to be independent. The dashed lines represent a 95% confidence interval. Shorter target sequences were captured with higher efficiency than long target sequences ($p<2x10^{-16}$). However, the use of this differential capture efficiency for systematic sequence length analysis was not previously recognized.

In some embodiments, following probe hybridization, polymerase fill-in and ligation reactions are performed to convert the hybridized probe to a covalently-closed, circular molecule containing the desired target. PCR or rolling circle amplification plus exonuclease digestion of non-circularized material is performed to isolate and amplify the circular targets from the starting nucleic acid pool. Since one of the main benefits of the method is the potential for a high degree of multiplexing, generally thousands of targets are captured in a single reaction containing thousands of probes.

According to aspects of the invention, repetitive regions are surrounded by non repetitive unique sequences, which can be used to amplify the repeat-containing regions using, for example, PCR or padlock (MIP)-based method.

In addition to the repetitive regions, a probe (e.g., a MIP or padlock probe) can be designed to include at least a sequence that is sufficient to be uniquely identified in the genome (or target pool). After the probe is circularized and amplified, the amplicon can be end-sequenced so that the unique sequence can be identified and served as the "representative" of the repetitive region as illustrated in FIG. 9. FIG. 9 illustrates a non-limiting scheme of padlock (MIP) capture of a region that includes both repetitive regions (thick wavy line) and the adjacent

Petitioner Illumina, Inc., EX1004, p. 707

unique sequence (thick strait line). The regions of the probe are indicated with the targeting arms shown as regions "1" and "3." An intervening region that may be, or include, a sequencing primer binding site is shown as "2." After the padlock is circularized and amplified, it can be end-sequenced to obtain the sequence of the unique sequence, which represents the repetitive region of interest.

Although capturing efficiency is overall negatively correlated with target length, different probe sequences may have unique features. Therefore, multiple probes could be designed and tested so that an optimal one is chosen to be sensitive enough to differentiate repetitive sizes of roughly 0-150bp, 150-600bp, and beyond, which represent normal, premutation and full mutation of fragile X syndrome, respectively. However, it should be appreciated that other probe sizes and sequences can be designed, and optionally optimized, to distinguish a range of repeat region size differences (e.g., length differences of about 3-30 bases, about 30-60 bases, about 60-90 bases, about 90-120 bases, about 120-150 bases, about 150-300 bases, about 300-600 bases, about 600-900 bases, or any intermediate or longer length difference). It should be appreciated that a length difference may be an increase in size or a decrease in size.

In some embodiments, an initial determination of an unexpected capture frequency is indicative of the presence of size difference. In some embodiments, an increase in capture frequency is indicative of a deletion. In some embodiments, a decrease in capture frequency is indicative of an insertion. However, it should be appreciated that depending on specific sequence parameters and the relative sizes of the capture probes, the target region, and the deletions or insertions, a change in capture frequency can be associated with either an increase or decrease in target region length. In some embodiments, the precise nature of the change can be determined using one or more additional techniques as described herein.

Accordingly, in some aspects a MIP probe includes a linear nucleic acid strand that contains two hybridization sequences or targeting arms, one at each end of the linear probe, wherein each of the hybridization sequences is complementary to a separate sequence on a the same strand of a target nucleic acid, and wherein these sequences on the target nucleic acid flank the two ends of the target nucleic acid sequence of interest. It should be appreciated that upon hybridization, the two ends of the probe are inverted with respect to each other in the sense that both 5' and 3' ends of the probe hybridize to the same strand to separate regions flanking the target region (as illustrated in FIG. 9 for example).

Petitioner Illumina, Inc., EX1004, p. 708

In some embodiments, the hybridization sequences are between about 10-100 nucleotides long, for example between about 10-30, about 30-60, about 60-90, or about 20, about 30, about 40, or about 50 nucleotides long. However, other lengths may be used depending on the application. In some embodiments, the hybridization Tms of both targeting arms of a probe are designed or selected to be similar. In some embodiments, the hybridization Tms of the targeting arms of a plurality of probes designed to capture different target regions are selected or designed to be similar so that they can be used together in a multiplex reaction. Accordingly, a typical size of a MIP probe prior to fill in is about 60-80 nucleotides long. However, other sizes can be used depending on the sizes of the targeting arms and any other sequences (e.g., primer binding or tag sequences) that are present in the MIP probe. In some embodiments, MIP probes are designed to avoid sequence-dependent secondary structures. In some embodiments, MIP probes are designed such that the targeting arms do not overlap with known polymorphic regions. In some embodiments, targeting arms that can be used for capturing the repeat region of the Fragile X locus can have the following sequences or complementary to these sequences depending on the strand that is captured.

left: CTCCUITYCGGTITCACTTC (SEQ ID NO: 181)

right: ATCTTCTCTTCAGCCCTGCT (SEQ ID NO: 182)

The typical captured size using these targeting arms is about 100 nucleotides in length (e.g., about 30 repeats of a tri-nucleotide repeat).

In some embodiments, the number of reads obtained for the "representative" of the repetitive region is not informative to estimate the target length because it is dependent on the total number of reads obtained. To overcome this, it is useful to include one or more probes that target other "control" regions where no or minimal polymorphism exists among populations. Because of the systematic consistency of capturing efficiency (see, e.g., FIG. 9), the ratio of reads obtained for the repetitive "representative" to reads obtained for the control region(s) will be tuned using DNA with defined numbers of repeats. Ultimately, the ratio can serve as a measure of the repeat length as illustrated in FIG. 10. FIG. 10 illustrates a non-limiting hypothetical relationship between target gap size and the relative number of reads of the repetitive region, which is measured by the ratio of the repeat "representative" reads vs. the "control" region reads. The unit of y-axis is arbitrary.

Petitioner Illumina, Inc., EX1004, p. 709

In some embodiments, to better tell targets with similar size range apart, the whole repetitive region can be sequenced by making a shotgun library (e.g., by making a shotgun library from a captured sequence, for example a sequence captured using a MIP probe). The longer the repeat is, the more short reads of repeats will be obtained. Therefore, the target length will contribute twice to the relative number of "repetitive" reads, which will gain better resolution of differentiating targets. In some embodiments, the expectation is that the number of reads from any given repeat will be a direct function of the number of repeats present. However, in some embodiments, a Poisson sampling induced spread may need to be considered and in some embodiments may be sufficiently large to limit the resolution.

When a precise measurement of the length of both alleles from a diploid sample is desired, further manipulations may be required. This is because the capture efficiency measured will actually be the average efficiency of the two alleles. To effectively achieve separate measurements for each allele, barcodes (e.g., sequence tags) can be used that allow the efficiency of individual capture events (from individual genomic loci) to be followed. FIG. 11A-C shows the approach. For a given locus, MIPs are synthesized to contain one of a large number differentiator tags in their backbone such that the probability of any two MIPs in a reaction having the same differentiator tag sequence is low. MIP capture is performed on the sample; the reaction will be biased for shorter target lengths, and therefore the reaction product will be comprised of more 'short' circles than 'long' circles. Each circle should bear a unique differentiator tag sequence. Then, linear RCA (1RCA) is performed on the circles. In the 1RCA reaction, circles are converted into long, linear concatemers of themselves. The 1RCA reaction for a given circle stops when the concatemer has reached a 'fixed' length (based on the processivity/error rate of the polymerase). Concatemers derived from smaller circles will therefore contain more copies of the differentiator tag, and concatemers derived from larger circles will contain fewer copies of the differentiator tag. The number of each differentiator tag sequence is counted, for example, by next-generation sequencing.

When number of occurrences is plotted against differentiator tag ID, the data will naturally cluster into two groups reflecting the lengths of the two alleles in the diploid sample. The allele lengths can therefore be read directly off this graph, after absolute length calibration using known standards. In some embodiments, a sequencing technique (e.g., a next-generation sequencing technique) is used to sequence part of one or more captured targets (e.g., or

Petitioner Illumina, Inc., EX1004, p. 710

amplicons thereof) and the sequences are used to count the number of different barcodes that are present. Accordingly, in some embodiments, aspects of the invention relate to a highly-multiplexed qPCR reaction.

Other non-limiting examples of loci at which insertions or deletions or repeat sequences may be associated with a disease or condition are provided in Tables 3 and 4. It should be appreciated that the presence of an abnormal length at any one or more of these loci may be evaluated according to aspects of the invention. In some embodiments, two or more of these loci or other loci may be evaluated in a single multiplex reaction using different probes designed to hybridize under the same reaction conditions to different target nucleic acid in a biological sample.

Table 3.

Polyglutamine (PolyQ) Diseases

Type	Gene	Normal/wildtype	Pathogenic
DRPLA (Dentatorubropallidoluysian atrophy)	ATN1 or DRPLA	6 - 35	49 - 88
HD (Huntington's disease)	HTT (Huntingtin)	10 - 35	35+
SBMA (Spinobulbar muscular atrophy or Kennedy disease)	Androgen receptor on the X chromosome.	9 - 36	38 - 62
SCA1 (Spinocerebellar ataxia Type 1)	ATXN1	6 - 35	49 - 88
SCA2 (Spinocerebellar ataxia Type 2)	ATXN2	14 - 32	33 - 77
SCA3 (Spinocerebellar ataxia Type 3 or Machado-Joseph disease)	ATXN3	12 - 40	55 - 86
SCA6 (Spinocerebellar ataxia Type 6)	CACNA1A	4 - 18	21 - 30
SCA7 (Spinocerebellar ataxia Type 7)	ATXN7	7 - 17	38 - 120
SCA17 (Spinocerebellar ataxia Type 17)	TBP	25 - 42	47 - 63

Table 4.

Petitioner Illumina, Inc., EX1004, p. 711

Non-Polyglutamine Diseases

Type	Gene	Codon	Normal/wildtype	Pathogenic
FRAXA (Fragile X syndrome)	*FMR1*, on the X-chromosome	CGG	6 - 53	230+
FXTAS (Fragile X-associated tremor/ ataxia syndrome)	*FMR1*, on the X-chromosome	CGG	6 - 53	55-200
FRAXE (Fragile XE mental retardation)	AFF2 or FMR2, on the X-chromosome	GCC	6 - 35	200+
FRDA (Friedreich's ataxia)	FXN or X25, (frataxin)	GAA	7 - 34	100+
DM (Myotonic dystrophy)	DMPK	CTG	5 - 37	50+
SCA8 (Spinocerebellar ataxia Type 8)	OSCA or SCA8	CTG	16 - 37	110 – 250
SCA12 (Spinocerebellar ataxia Type 12)	PPP2R2B or SCA12	CAG On 5' end	7 - 28	66 - 78

The following examples illustrate aspects and embodiments of the invention and are not intended to be limiting or restrictive. Many variations of the invention will become apparent to those skilled in the art upon review of this specification. The full scope of the invention should be determined by reference to the claims, along with their full scope of equivalents, and the specification, along with such variations.

4. Increasing detection sensitivity:

In some embodiments, aspects of the invention relate to methods for increasing the sensitivity of nucleic acid detection assays.

There are currently many genomic assays that utilize next-generation (e.g., polony-based) sequencing to generate data, including genome resequencing, RNA-seq for gene expression, bisulphite sequencing for methylation, and Immune-seq, among others. In order to make

Petitioner Illumina, Inc., EX1004, p. 712

quantitative measurements (including genotype calling), these methods utilize the counts of sequencing reads of a given genomic locus as a proxy for the representation of that sequence in the original sample of nucleic acids. The majority of these techniques require a preparative step to construct a high-complexity library of DNA molecules that is representative of a sample of interest. Current assays use one of several alternative nucleic acid preparative techniques (e.g., amplification, for example PCR-based amplification; sequence-specific capture, for example, using immobilized capture probes; or target capture into a circularized probe followed by a sequence analysis step. In order to reduce errors associated with the unpredictability (stochastic nature) of nucleic acid isolation and sequence analysis techniques, current methods involve oversampling a target nucleic acid preparation in order to increase the likelihood that all sequences that are present in the original nucleic acid sample will be represented in the final sequence data. For example, a genomic sequencing library may contain an over- or under-representation of particular sequences from a source nucleic acid sample (e.g., genome preparation) as a result of stochastic variations in the library construction process. Such variations can be particularly problematic when they result in target sequences from a genome being absent or undetectable in a sequencing library. For example, an under-representation of particular allelic sequences (e.g., heterozygotic alleles) from a genome in a sequencing library can result in an apparent homozygous representation in a sequencing library.

In contrast, aspects of the invention relate to basing a nucleic acid sequence analysis on results from two or more different nucleic acid preparatory techniques that have different systematic biases in the types of nucleic acids that they sample rather than simply oversampling the target nucleic acid. According to some embodiments, different techniques have different sequence biases that are systematic and not simply due to stochastic effects during nucleic acid capture or amplification. Accordingly, in some embodiments, the degree of oversampling required to overcome variations in nucleic acid preparation needs to be sufficient to overcome the biases. In some embodiments, the invention provides methods that reduce the need for oversampling by combining nucleic acid and/or sequence results obtained from two or more different nucleic acid preparative techniques that have different biases.

According to the invention, different techniques have different characteristic or systematic biases. For example, one technique may bias a sample analysis towards one particular allele at a genetic locus of interest, whereas a different technique would bias the

Petitioner Illumina, Inc., EX1004, p. 713

sample analysis towards a different allele at the same locus. Accordingly, the same sample may be identified as being different depending on the type of technique that is used to prepare nucleic acid for sequence analysis. This effectively represents a sensitivity issue, because each technique has a different relative sensitivities for polymorphic sequences of interest.

According to aspects of the invention, the sensitivity of a nucleic acid analysis can be increased by combining the sequences from different nucleic acid preparative steps and using the combined sequence information for a diagnostic assay (e.g., for a making a call as to whether a subject is homozygous or heterozygous at a genetic locus of interest).

Currently, the ability of DNA sequencing to detect mutations is limited by the ability of the upstream sample isolation (e.g., by amplification, immobilization enrichment, circularization capture, etc.) methods to reliably isolate the locus of interest.

If one wishes to make heterozygote base-calls for a diploid genome (e.g. a human sample presented for molecular diagnostic sequencing), it is important in some embodiments that the isolation method produces near- or perfectly- uniform amounts of the two alleles to be sequenced (at least sufficiently uniform to be "called" unambiguously as a heterozygote or a homozygote for a locus of interest).

Sample preparative methods may fall into three classes: 1) single- or several target amplification (e.g., uniplex PCR, 'multiplex' PCR), 2) multi-target hybridization enrichment (e.g., Agilent SureSelect 'hybrid capture' [Gnirke et al 2009, *Nature methods* 27:182-9], Roche/Nimblegen 'sequence capture' [Hodges et al 2007, *Nature genetics* 39:1522-7], and 3) multi-target circularization selection (e.g. molecular inversion probes or padlock probes, [Porreca et al 2007, *Nature methods* 4:931-6, Turner et al 2009, *Nature methods* 6:315-6], 'selectors' [Dahl et al 2005, *Nucleic acids research* 33:e71]). Each of these methods can result in a pool of isolated product that does not adequately represent the input abundance distribution. For example, the two alleles at a heterozygous position can become skewed far from their input 50:50 ratio to something that results in a missed basecall during downstream sequencing. For example, if the ratio was skewed from 50:50 to 10:90, and the sample was sequenced to 10X average coverage, there is a high probability that one of the two alleles would not be observed once in the ten sequencing reads. This would reduce the sensitivity of the sequencing method by converting a heterozygous position to homozygous (where potentially the 'mutant' allele was the one not observed). In some embodiments, a skewed ratio is a particular issue that decreases the

Petitioner Illumina, Inc., EX1004, p. 714

sensitivity of detecting mutations present in a heterogeneous tumor tissue. For example, if only 10% of the cells analyzed in a heterogeneous sample harbored a heterozygous mutation, the mutation would be expected to be present in 5% of sequence reads, not 50%. In this scenario, the need for robust, sensitive detection may be even more acute.

The methods disclosed herein are based, in part, on the discovery that certain classes of isolation methods have different modes of bias. The disclosure provide methods for increasing the sensitivity of the downstream sequencing by using a combination of multiple isolation methods (e.g., one or more from at least two of the classes disclosed herein) for a sample. This is particularly important in molecular diagnostics where high sensitivity is required to minimize the chances of 'missing' a disease-associated mutation. For example, given a nominal false-negative error rate of $1x10^{-3}$ for sequencing following circularization selection, and a false-negative error rate of $1x10^{-3}$ for sequencing following hybridization enrichment, one can achieve a final false-negative rate of $1x10^{-6}$ by performing both techniques on the sample (assuming failures in each method are fully independent). For a recessive disease with carrier frequency of 0.1, caused by a single fully-penetrant mutant allele, the number of missed carrier diagnoses would decrease from 1000 per million patients tested to 1 per million patients tested. Furthermore, if the testing was used in the context of prenatal carrier screening, the number of affected children born as a result of missing the carrier call in one parent would decrease from 25 per million to 25 per billion born.

Additionally, the disclosure provides combinations of preparative methods to effectively increase sequencing coverage in regions containing disease-associated alleles. Since heterozygote error rate is largely tied to both deviations from 50:50 allele representation, and in the case of next-generation DNA sequencing deviations from average abundance (such that less abundant isolated targets are more likely to be undersampled at one or both alleles), selectively increasing coverage in these regions will also selectively increase sensitivity. Furthermore, MIPs that detect presence or absence of specific known disease-associated mutations can be used to increase sensitivity selectively. In some embodiments, these MIPs would have a targeting arm whose 3' -most region is complementary to the expected mutation, and has a fill-in length of 0 or more bp. Thus, the MIP will form only if the mutation is present, and its presence will be detected by sequencing.

Petitioner Illumina, Inc., EX1004, p. 715

Additionally, algorithms disclosed herein may be used to determine base identity with varying levels of stringency depending on whether the given position has any known disease-associated alleles. Stringency can be reduced in such positions by decreasing the minimum number of observed mutant reads necessary to make a consensus base-call. This will effectively increase sensitivity for mutant allele detection at the cost of decreased specificity.

An embodiment of the invention combines MIPs plus hybridization enrichment, plus optionally extra MIPs targeted to specific known, common disease-associated loci, e.g., to detect the presence of a polymorphism in a target nucleic acid. A non-limiting example is illustrated in FIG. 12 that illustrates a schematic using MIPs plus hybridization enrichment, plus optionally extra MIPs targeted to specific known, common disease-associated loci, e.g., to detect the presence of a polymorphism in a target nucleic acid.

FIGs. 13 and 14 illustrate different capture efficiencies for MIP-based captures. FIG. 13 shows a graph of per-target abundance with MIP capture. In this graph, bias largely drives the heterozygote error rate, since targets which are less abundant here are less likely to be covered in sufficient depth during sequencing to adequately sample both alleles. This is from Turner et al 2009, Nature methods 6:315-6. Hybridization enrichment results in a qualitatively similar abundance distribution, but the abundance of a given target is likely not correlated between the two methods. FIG. 14 shows a graph of correlation between two MIP capture reactions from Ball et al 2009, Nature biotechnology 27:361-8. Each point represents the target abundance in replicate 1 and replicate 2. Pearson correlation r=0.956. This indicates that MIP capture reproducibly biases targets to specific abundances. Hybridization enrichment is similarly correlated from one capture to the next.

According to aspects of the invention, such biases can be detected or overcome by systematically combining different capture and/or analytical techniques in an assay that interrogates a plurality of loci in a plurality of subject samples.

Accordingly, it should be appreciated that in any of the embodiments described herein (e.g., tiling/staggering, tagging, size-detection, sensitivity enhancing algorithms, or any combination thereof), aspects of the invention involve preparing genomic nucleic acid and/or contacting them with one or more different probes (e.g., capture probes, hybridization probes, MIPs, others etc.). In some embodiments, the amount of genomic nucleic acid used per subject ranges from 1 ng to 10 micrograms (e.g., 500 ng to 5 micrograms). However, higher or lower

Petitioner Illumina, Inc., EX1004, p. 716

amounts (e.g., less than 1 ng, more than 10 micrograms, 10-50 micrograms, 50-100 micrograms or more) may be used. In some embodiments, for each locus of interest, the amount of probe used per assay may be optimized for a particular application. In some embodiments, the ratio (molar ratio, for example measured as a concentration ratio) of probe to genome equivalent (e.g., haploid or diploid genome equivalent, for example for each allele or for both alleles of a nucleic acid target or locus of interest) ranges from 1/100, 1/10, 1/1, 10/1, 100/1, 1000/1. However, lower, higher, or intermediate ratios may be used.

In some embodiments, the amount of target nucleic acid and probe used for each reaction is normalized to avoid any observed differences being caused by differences in concentrations or ratios. In some embodiments, in order to normalize genomic DNA and probe, the genomic DNA concentration is read using a standard spectrophotometer or by fluorescence (e.g., using a fluorescent intercalating dye). The probe concentration may be determined experimentally or using information specified by the probe manufacturer.

Similarly, once a locus has been captured (e.g., on a MIP or other probe or in another form), it may be amplified and/or sequenced in a reaction involving one or more primers. The amount of primer added for each reaction can range from 0.1 pmol to 1 nmol, 0.15 pmol to 1.5 nmol (for example around 1.5 pmol). However, other amounts (e.g., lower, higher, or intermediate amounts) may be used.

In some embodiments, it should be appreciated that one or more intervening sequences (e.g., sequence between the first and second targeting arms on a MIP capture probe), identifier or tag sequences, or other probe sequences that are not designed to hybridize to a target sequence (e.g., a genomic target sequence) should be designed to avoid excessive complementarity (to avoid cross-hybridization) to target sequences or other sequences (e.g., other genomic sequences) that may be in a biological sample. For example, these sequences may be designed have a sufficient number of mismatches with any genomic sequence (e.g., at least 5, 10, 15, or more mismatches out of 30 bases) or as having a Tm (e.g., a mismatch Tm) that is lower (e.g., at least 5, 10, 15, 20, or more degrees C lower) than the hybridization reaction temperature.

It should be appreciated that a targeting arm as used herein may be designed to hybridize (e.g., be complementary) to either strand of a genetic locus of interest if the nucleic acid being analyzed is DNA (e.g., genomic DNA). However, in the context of MIP probes, whichever strand is selected for one targeting arm will be used for the other one. However, in the context of

Petitioner Illumina, Inc., EX1004, p. 717

RNA analysis, it should be appreciated that a targeting arm should be designed to hybridize to the transcribed RNA. It also should be appreciated that MIP probes referred to herein as "capturing" a target sequence are actually capturing it by template-based synthesis rather than by capturing the actual target molecule (other than for example in the initial stage when the arms hybridize to it or in the sense that the target molecule can remain bound to the extended MIP product until it is denatured or otherwise removed).

It should be appreciated that in some embodiments a targeting arm may include a sequence that is complementary to one allele or mutation (e.g., a SNP or other polymorphism, a mutation, etc.) so that the probe will preferentially hybridize (and capture) target nucleic acids having that allele or mutation. However, in many embodiments, each targeting arm is designed to hybridize (e.g., be complementary) to a sequence that is not polymorphic in the subjects of a population that is being evaluated. This allows target sequences to be captured and/or sequenced for all alleles and then the differences between subjects (e.g., calls of heterozygous or homozygous for one or more loci) can be based on the sequence information and/or the frequency as described herein.

It should be appreciated that sequence tags (also referred to as barcodes) may be designed to be unique in that they do not appear at other positions within a probe or a family of probes and they also do not appear within the sequences being targeted. Thus they can be used to uniquely identify (e.g., by sequencing or hybridization properties) particular probes having other characteristics (e.g., for particular subjects and/or for particular loci).

It also should be appreciated that in some embodiments probes or regions of probes or other nucleic acids are described herein as comprising or including certain sequences or sequence characteristics (e.g., length, other properties, etc.). However, it should be appreciated that in some embodiments, any of the probes or regions of probes or other nucleic acids consist of those regions (e.g., arms, central regions, tags, primer sites, etc., or any combination thereof) of consist of those sequences or have sequences with characteristics that consist of one or more characteristics (e.g., length, or other properties, etc.) as described herein in the context of any of the embodiments (e.g., for tiled or staggered probes, tagged probes, length detection, sensitivity enhancing algorithms or any combination thereof).

It should be appreciated that probes, primers, and other nucleic acids designed or used herein may be synthetic, natural, or a combination thereof. Accordingly, as used herein, the term

Petitioner Illumina, Inc., EX1004, p. 718

"nucleic acid" refers to multiple linked nucleotides (i.e., molecules comprising a sugar (e.g., ribose or deoxyribose) linked to an exchangeable organic base, which is either a pyrimidine (e.g., cytosine (C), thymidine (T) or uracil (U)) or a purine (e.g., adenine (A) or guanine (G)). "Nucleic acid" and "nucleic acid molecule" may be used interchangeably and refer to oligoribonucleotides as well as oligodeoxyribonucleotides. The terms shall also include polynucleosides (i.e., a polynucleotide minus a phosphate) and any other organic base containing nucleic acid.

The organic bases include adenine, uracil, guanine, thymine, cytosine and inosine. Unless otherwise stated, nucleic acids may be single or double stranded. The nucleic acid may be naturally or non-naturally occurring. Nucleic acids can be obtained from natural sources, or can be synthesized using a nucleic acid synthesizer (i.e., synthetic).

Harvest and isolation of nucleic acids are routinely performed in the art and suitable methods can be found in standard molecular biology textbooks. (See, for example, Maniatis' Handbook of Molecular Biology.) The nucleic acid may be DNA or RNA, such as genomic DNA, mitochondrial DNA, mRNA, cDNA, rRNA, miRNA, or a combination thereof. Non-naturally occurring nucleic acids such as bacterial artificial chromosomes (BACs) and yeast artificial chromosomes (YACs) can also be used.

The invention also contemplates the use of nucleic acid derivatives. As will be described herein, the use of certain nucleic acid derivatives may increase the stability of the nucleic acids of the invention by preventing their digestion, particularly when they are exposed to biological samples that may contain nucleases. As used herein, a nucleic acid derivative is a non-naturally occurring nucleic acid or a unit thereof. Nucleic acid derivatives may contain non-naturally occurring elements such as non-naturally occurring nucleotides and non-naturally occurring backbone linkages.

Nucleic acid derivatives may contain backbone modifications such as but not limited to phosphorothioate linkages, phosphodiester modified nucleic acids, phosphorothiolate modifications, combinations of phosphodiester and phosphorothioate nucleic acid, methylphosphonate, alkylphosphonates, phosphate esters, alkylphosphonothioates, phosphoramidates, carbamates, carbonates, phosphate triesters, acetamidates, carboxymethyl esters, methylphosphorothioate, phosphorodithioate, p-ethoxy, and combinations thereof. The backbone composition of the nucleic acids may be homogeneous or heterogeneous.

Petitioner Illumina, Inc., EX1004, p. 719

Nucleic acid derivatives may contain substitutions or modifications in the sugars and/or bases. For example, they include nucleic acids having backbone sugars which are covalently attached to low molecular weight organic groups other than a hydroxyl group at the 3' position and other than a phosphate group at the 5' position (e.g., an 2'-0-alkylated ribose group). Nucleic acid derivatives may include non-ribose sugars such as arabinose. Nucleic acid derivatives may contain substituted purines and pyrimidines such as C-5 propyne modified bases, 5-methylcytosine, 2-aminopurine, 2-amino-6-chloropurine, 2,6-diaminopurine, hypoxanthine, 2-thiouracil and pseudoisocytosine. In some embodiments, substitution(s) may include one or more substitutions/modifications in the sugars/bases, groups attached to the base, including biotin, fluorescent groups (fluorescein, cyanine, rhodamine, etc), chemically-reactive groups including carboxyl, NHS, thiol, etc., or any combination thereof.

A nucleic acid may be a peptide nucleic acid (PNA), locked nucleic acid (LNA), DNA, RNA, or co-nucleic acids of the same such as DNA-LNA co-nucleic acids. PNA are DNA analogs having their phosphate backbone replaced with 2-aminoethyl glycine residues linked to nucleotide bases through glycine amino nitrogen and methylenecarbonyl linkers. PNA can bind to both DNA and RNA targets by Watson-Crick base pairing, and in so doing form stronger hybrids than would be possible with DNA or RNA based oligonucleotides in some cases.

PNA are synthesized from monomers connected by a peptide bond (Nielsen, P.E. et al. Peptide Nucleic Acids, Protocols and Applications, Norfolk: Horizon Scientific Press, p. 1-19 (1999)). They can be built with standard solid phase peptide synthesis technology. PNA chemistry and synthesis allows for inclusion of amino acids and polypeptide sequences in the PNA design. For example, lysine residues can be used to introduce positive charges in the PNA backbone. All chemical approaches available for the modifications of amino acid side chains are directly applicable to PNA. Several types of PNA designs exist, and these include single strand PNA (ssPNA), bisPNA and pseudocomplementary PNA (pcPNA).

The structure of PNA/DNA complex depends on the particular PNA and its sequence. ssPNA binds to single stranded DNA (ssDNA) preferably in antiparallel orientation (i.e., with the N-terminus of the ssPNA aligned with the 3' terminus of the ssDNA) and with a Watson-Crick pairing. PNA also can bind to DNA with a Hoogsteen base pairing, and thereby forms triplexes with double stranded DNA (dsDNA) (Wittung, P. et al., *Biochemistry* 36:7973 (1997)).

Petitioner Illumina, Inc., EX1004, p. 720

A locked nucleic acid (LNA) is a modified RNA nucleotide. An LNA form hybrids with DNA, which are at least as stable as PNA/DNA hybrids (Braasch, D.A. et al., *Chem & Biol.* 8(1):1-7(2001)). Therefore, LNA can be used just as PNA molecules would be. LNA binding efficiency can be increased in some embodiments by adding positive charges to it. LNAs have been reported to have increased binding affinity inherently.

Commercial nucleic acid synthesizers and standard phosphoramidite chemistry are used to make LNAs. Therefore, production of mixed LNA/DNA sequences is as simple as that of mixed PNA/peptide sequences. The stabilization effect of LNA monomers is not an additive effect. The monomer influences conformation of sugar rings of neighboring deoxynucleotides shifting them to more stable configurations (Nielsen, P.E. et al. Peptide Nucleic Acids, Protocols and Applications, Norfolk: Horizon Scientific Press, p. 1-19 (1999)). Also, lesser number of LNA residues in the sequence dramatically improves accuracy of the synthesis. Most of biochemical approaches for nucleic acid conjugations are applicable to LNA/DNA constructs.

These and other aspects of the invention are illustrated by the following non-limiting examples.

EXAMPLES

The following examples illustrate non-limiting embodiments of the invention.

Example 1: Design a set of capture probes for a human target exon

All targets are captured as a set of partially-overlapping subtargets. For example, in the tiling approach, a 200 bp target exon might be captured as a set of 12 subtargets, each 60bp in length (FIG. 1). Each subtarget is chosen such that it partially overlaps two or three other targets.

In some embodiments, all probes are composed of three regions: 1) a 20 bp 'targeting arm' comprised of sequence which hybridizes immediately upstream from the sub-target, 2) a 30 bp 'constant region' comprised of sequence used as a pair of amplification priming sites, and 3) a second 20 bp 'targeting arm' comprised of sequence which hybridizes immediately downstream from the sub-target. Targeting arm sequences will be different for each capture probe in a set, while constant region sequence will be the same for all probes in the set, allowing all captured targets to be amplified with a single set of primers. Targeting arm sequences should be designed

Petitioner Illumina, Inc., EX1004, p. 721

such that any given pair of 20 bp sequences is unique in the target genome (to prevent spurious capture of undesired sites). Additionally, melting temperatures should be matched for all probes in the set such that hybridization efficiency is uniform for all probes at a constant temperature (e.g., 60 C). Targeting arm sequences should be computationally screened to ensure they do not form strong secondary structure that would impair their ability to basepair with the genomic target.

Hybridize capture probes to human genomic sample

Assemble hybridization reaction:

- 1.0 ul capture probe mix (~2.5 pmol)
- 2.0 ul 10X Ampligase buffer (Epicentre)
- 6.0 ul 500 ng/ul human genomic DNA (~16.7 fmol)
- 11 ul dH20

In a thermal cycler, heat reaction to 95C for 5 min to denature genomic DNA, then cool to 60C. Allow to incubate at 60C for 40 hours.

Convert hybridized probes into covalently-closed circular products containing subtargets

Prepare fill-in / ligation reaction mixture:

- 0.25 ul 2 mM dNTP mix (Invitrogen)
- 2.5 ul 10X Ampligase buffer (Epicentre)
- 5.0 ul 5 U/ul Taq Stoffel fragment (Applied Biosystems)
- 12.5 ul 5 U/ul Ampligase (Epicentre)
- 4.75 ul dH20

Add 1.0 ul of this mix to the hybridized probe reaction, and incubate at 60C for 10 hours.

Purify circularized probe / subtarget products from un-reacted probes and genomic DNA

Prepare exonuclease reaction mixture:

- 21 ul fill-in / ligation reaction product
- 2.0 ul 10X exonuclease I buffer (New England Biolabs)

Petitioner Illumina, Inc., EX1004, p. 722

- 2.0 ul 20 U/ul exonuclease I (New England Biolabs)
- 2.0 ul 100 U/ul exonuclease III (New England Biolabs)

Incubate at 37C for 60 min, then heat-inactivate by incubating at 80C for 15 min. Immediately cool to 4C for storage.

Amplify circular material by PCR using primers specific to the 'constant region' of the probes

Prepare PCR mixture:

- 5.0 ul 10X Accuprime reaction buffer (Invitrogen)
- 1.5 ul 10 uM CP-2-FA (5'-GCACGATCCGACGGTAGTGT-3') (SEQ ID NO:183)
- 1.5 ul 10 uM CP-2-RA (5'-CCGTAATCGGGAAGCTGAAG-3') (SEQ ID NO:184)
-
- 0.4 ul 25 mM dNTP mix (Invitrogen)
- 2.0 ul heat-inactivated exonuclease reaction mix
- 1.5 ul 10X SybrGreen (Invitrogen)
- 0.4 ul 2.5 U/ul Accuprime Pfx polymerase (Invitrogen)
- 37.7 ul dH20

Thermal cycle in real-time thermal cycler according to the following protocol, but stop cycling before amplification yield plateaus (generally 8-12 cycles):

1. 95C for 5 min
2. 95C for 30 sec
3. 58C for 60 sec
4. 72C for 60 sec
5. goto 2, N more times

Prepare a shotgun next-generation sequencing library for analysis

- Purify desired amplicon population from non-specific amplification products by gel extraction.

Petitioner Illumina, Inc., EX1004, p. 723

- Concatemerize amplicons into high-molecular weight products suitable for shearing
- Mechanically shear, using either a nebulizer, BioRuptor, Hydroshear, Covaris, or similar instrument. DNA should be sheared into fragments several hundred basepairs in length.
- Ligate adapters required for amplification by the sequencing platform used. If necessary, purify ligated product from unligated product and adapters.

Example 2: Use of differentiator tag sequences to detect and correct bias in a MIP-capture reaction of a set of exon targets

The first step in performing the detection/correction is to determine how many differentiator tag sequences are necessary for the given sample. In this example, 1000 genomic targets corresponding to 1000 exons were captured. Since the differentiator tag sequence is part of the probe, it will measure/report biases that occur from the earliest protocol steps. Also, being located in the backbone, the differentiator tag sequence can easily be sequenced from a separate priming site, and therefore not impact the total achievable read-length for the target sequence. MIP probes are synthesized using standard column-based oligonucleotide synthesis by any number of vendors (e.g. IDT), and differentiator tag sequences are introduced as 'degenerate' positions in the backbone. Each degenerate position increases the total number of differentiator tag sequences synthesized by a factor of 4, so a 10 nt degenerate region implies a differentiator tag sequence complexity of ~1e6 species.

Hybridize capture probes to human genomic sample

Assemble hybridization reaction:

- 1.0 ul capture probe mix (~2.5 pmol)
- 2.0 ul 10X Ampligase buffer (Epicentre)
- 6.0 ul 500 ng/ul human genomic DNA (~16.7 fmol)
- 11 ul dH20

In a thermal cycler, heat reaction to 95C for 5 min to denature genomic DNA, then cool to 60C. Allow to incubate at 60C for 40 hours.

Convert hybridized probes into covalently-closed circular products containing subtargets

Petitioner Illumina, Inc., EX1004, p. 724

Prepare fill-in / ligation reaction mixture:

- 0.25 ul 2 mM dNTP mix (Invitrogen)
- 2.5 ul 10X Ampligase buffer (Epicentre)
- 5.0 ul 5 U/ul Taq Stoffel fragment (Applied Biosystems)
- 12.5 ul 5 U/ul Ampligase (Epicentre)
- 4.75 ul dH20

Add 1.0 ul of this mix to the hybridized probe reaction, and incubate at 60C for 10 hours.

Purify circularized probe / subtarget products from un-reacted probes and genomic DNA

Prepare exonuclease reaction mixture:

- 21 ul fill-in / ligation reaction product
- 2.0 ul 10X exonuclease I buffer (New England Biolabs)
- 2.0 ul 20 U/ul exonuclease I (New England Biolabs)
- 2.0 ul 100 U/ul exonuclease III (New England Biolabs)

Incubate at 37C for 60 min, then heat-inactivate by incubating at 80C for 15 min. Immediately cool to 4C for storage.

Amplify circular material by PCR using primers specific to the 'constant region' of the probes

Prepare PCR mixture:

- 5.0 ul 10X Accuprime reaction buffer (Invitrogen)
- 1.5 ul 10 uM CP-2-FA (5'-GCACGATCCGACGGTAGTGT-3') (SEQ ID NO: 183)
-
- 1.5 ul 10 uM CP-2-RA (5'-CCGTAATCGGGAAGCTGAAG-3') (SEQ ID NO: 184)
-
- 0.4 ul 25 mM dNTP mix (Invitrogen)
- 2.0 ul heat-inactivated exonuclease reaction mix
- 1.5 ul 10X SybrGreen (Invitrogen)
- 0.4 ul 2.5 U/ul Accuprime Pfx polymerase (Invitrogen)

Petitioner Illumina, Inc., EX1004, p. 725

- 37.7 ul dH20

Thermal cycle in real-time thermal cycler according to the following protocol, but stop cycling before amplification yield plateaus (generally 8-12 cycles):

6. 95C for 5 min
7. 95C for 30 sec
8. 58C for 60 sec
9. 72C for 60 sec
10. goto 2, N more times

Prepare a shotgun next-generation sequencing library for analysis

- Purify desired amplicon population from non-specific amplification products by gel extraction.
- Concatemerize amplicons into high-molecular weight products suitable for shearing
- Mechanically shear, using either a nebulizer, BioRuptor, Hydroshear,
- Covaris, or similar instrument. DNA should be sheared into fragments several hundred basepairs in length.
- Ligate adapters required for amplification by the sequencing platform used. If necessary, purify ligated product from unligated product and adapters.

Perform sequencing of library according to manufacturer's directions (e.g. Illumina,ABI, etc), reading both the target sequence and the differentiator tag sequence.

Analyze data by correcting for any biases detected by quantitation of differentiator tag sequence abundance

Construct a table of target: differentiator tag abundances from the read data, e.g.:

Target ID	Differentiator Tag sequence ID	Count
1	3547	1
2	4762	1
1	9637	1
1	1078	5
3	4762	1
1	2984	1

Petitioner Illumina, Inc., EX1004, p. 726

All 'count' entries should be '1', since any particular target:differentiator tag mapping will not occur more than once by chance, and therefore will only be observed if bias was present somewhere in the sample preparation process. For any target:differentiator tag combination observed more than once, all such reads are 'collapsed' into a single read before consensus basecalls are determined. This will cancel the effect of bias on consensus basecall accuracy. FIG. 5 depicts a method for making diploid genotype calls in which repeat target:differentiator tag combination are collapsed.

Example 3: Differentiator tag sequence design for MIP capture reactions

For a set of targets, the number of differentiator tag sequences necessary to be confident (within some statistical bounds) that a certain differentiator tag sequence will not be observed more than once by chance in combination with a certain target sequence was determined. The total number of unique differentiator tag sequences for a certain differentiator tag sequence length is determined as $4^{(\text{Length in nucleotides of the differentiator tag sequence})}$. For a molecular inversion probe capture reaction that uses MIP probes having differentiator tag sequences, the probability of performing the capture reaction and capturing one or more copies of a target sequence having the same differentiator tag sequence is calculated as: $p = 1 - [N! / (N-M)!] / [NAM]$, wherein N is the total number of possible unique differentiator tag sequences and M is the number of target sequence copies in the capture reaction. Thus, by varying the differentiator tag sequence length it is possible to perform a MIP capture reaction in which the probability of capturing one or more copies of a target sequence having the same differentiator tag sequence is set at a predetermined probability value.

For example, for a differentiator tag sequence of 15 nucleotides in length, there are 1,073,741,824 possible differentiator tag sequences. A MIP capture reaction in which MIP probes, each having a differentiator tag sequence of 15 nucleotides, are combined with 10000 target sequence copies (e.g., genome equivalents), the probability of capturing one or more copies of a target sequence having the same differentiator tag sequence is 0.05. In this example, the MIP reaction will produce very few (usually 0, but occasionally 1 or more) targets where multiple copies are tagged with the same differentiator tag sequence. FIG. 6 depicts results of a simulation for 100000 capture reactions having 15 nucleotide differentiator tag sequences and 10000 target sequences.

Petitioner Illumina, Inc., EX1004, p. 727

Example 4: Assessment of the probability for obtaining enough sequencing reads to make accurate base-calls at multiple independent loci, as a function of sequencing coverage.

Monte Carlo simulations were performed to determine sequencing coverage requirements. The simulations assume 10000 genomic copies of a given locus (target) half mom alleles and half dad alleles. The simulations further assume 1% efficiency of capture for the MIP reaction. The simulation samples from a capture mix 100 times without replacement to create a set of 100 capture products. The simulation then samples from the set of 100 capture products with replacement (assuming unbiased amplification) to generate 'reads' from either mom or dad. The number of reads sampled depends on the coverage. The number of independent reads from both mom and dad necessary to make a high-quality base-call (assumed to be 10 or 20 reads) were then determined. The process was repeated 1000 times for each coverage level, and the fraction of times that enough reads from both parents were successfully obtained was determined. This fraction was raised to the power 1000, assuming we have 1000 independent loci that must obtain successful base-calls, plotted (See FIG. 7). Result show that roughly 50x coverage is required to capture each allele >= 10x with > 0.95 probability.

Example 5: MIP capture of 'target' locus and 'control' loci

In some embodiments, to accurately quantify the efficiency of target locus capture, at least three sets of control loci are captured in parallel that have a priori been shown to serve as proxies for various lengths of target locus. For example, if the target locus is expected to have a length between 50 and 1000 bp, then sets of control loci having lengths of 50, 250, and 1000 bp could be captured (e.g. 20 loci per set should provide adequate protection from outliers), and their abundance digitally measured by sequencing. These loci should be chosen such that minimal variation in efficiency between samples and on multiple runs of the same sample is observed (and are therefore 'efficiency invariant'). These will serve as 'reference' points that define the shape of the curve of abundance-vs-length. Determining the length of the target is then simply a matter of 'reading' the length from the appropriate point on the calibration curve.

In some embodiments, the statistical confidence one has in the estimate of target length from this method is driven largely by three factors: 1) reproducibility / variation of the abundance data used to generate the calibration curve; 2) goodness of fit of the regression to the 'control' datapoints; 3) reproducibility of abundance data for the target locus being measured. Statistical bounds on 1) and 2) will be known in advance, having been measured during

Petitioner Illumina, Inc., EX1004, p. 728

development of the assay. Additionally, statistical bounds on 3) will be known in general in advance, since assay development should include adequate population sampling and measure of technical reproducibility. Standard statistical methods should be used to combine these three measures into a single P value for any given experimental measure of target abundance.

In some embodiments, given the set of calibration observations, and a linear regression fit to that data, the regression can be used to predict the length value for n observations of the target locus whose length is unknown. First, choose an acceptable range for the confidence interval of the length estimate. For example, in the case of distinguishing "normal" (87-93bp) from "premutation" (165-600bp) potential cases of Fragile X, the goal is to measure length to sufficient precision to distinguish 93 bp from 165 bp. The predicted response value, computed when n observations is substituted into the equation for the regressed line, will have arbitrary precision. However, if for example a 95% confidence level is desired, that 95% confidence interval must be sufficiently short that it does not overlap both the "normal" and "premutation" length ranges. Continuing the example, if one calculates a length of 190 from n=400 MIP observations, and based on the regression from calibration data, the 95% confidence interval is 190 +/- 20 bp, one can conclude the sample represents a "premutation" length with 95% certainty. Conversely, if the calibration data were less robust, error estimates of the regression would be higher, leading to larger confidence intervals on the predicted response value. In some embodiments, if the 95% CI were calculated as 190 +/- 100 bp from n=400, one could not determine whether the predicted response value corresponds to a "normal" or "premutation" length.

In some embodiments, the confidence interval for a predicted response is calculated as:

The estimate for the response $\hat{y}$ *is identical to the estimate for the mean of the response:* $\hat{y} = b_0 + b_1 x^*$. *The confidence interval for the predicted value is given by* $\hat{y} \pm t^* s_{\hat{y}}$, *where* $\hat{y}$ *is the fitted value corresponding to* x^*. *The value* t^* *is the upper (1 - C)/2 critical value for the t(n - 2) distribution.*

In some embodiments, a technique for analyzing a locus of interest can involve the following steps.

Petitioner Illumina, Inc., EX1004, p. 729

Convert hybridized probes into covalently-closed circular products containing subtargets

Prepare fill-in / ligation reaction mixture:

- 0.25 ul 2 mM dNTP mix (Invitrogen)
- 2.5 ul 10X Ampligase buffer (Epicentre)
- 5.0 ul 5 U/ul Taq Stoffel fragment (Applied Biosystems)
- 12.5 ul 5 U/ul Ampligase (Epicentre)
- 4.75 ul dH20

Add 1.0 ul of this mix to the hybridized probe reaction, and incubate at 60C for 10 hours.

Purify circularized probe / subtarget products from un-reacted probes and 30 genomic DNA

Prepare exonuclease reaction mixture:

- 21 ul fill-in / ligation reaction product
- 2.0 ul 10X exonuclease I buffer (New England Biolabs)
- 2.0 ul 20 U/ul exonuclease I (New England Biolabs)
- 2.0 ul 100 U/ul exonuclease III (New England Biolabs)

Incubate at 37C for 60 min, then heat-inactivate by incubating at 80C for 15 min. Immediately cool to 4C for storage.

Amplify circular material by PCR using primers specific to the 'constant region' of the probes

Prepare PCR mixture:

- 5.0 ul 10X Accuprime reaction buffer (Invitrogen)
- 1.5 ul 10 uM CP-2-FA-Ilmn (platform-specific amplification sequence plus 'circle constant region'-specific sequence)
- 1.5 ul 10 uM CP-2-RA-Ilmn (platform-specific amplification sequence plus 'circle constant region'-specific sequence)
- 0.4 ul 25 mM dNTP mix (Invitrogen)
- 2.0 ul heat-inactivated exonuclease reaction mix
- 1.5 ul 10X SybrGreen (Invitrogen)
- 0.4 ul 2.5 U/ul Accuprime Pfx polymerase (Invitrogen)

Petitioner Illumina, Inc., EX1004, p. 730

- 37.7 ul dH20

Thermal cycle in real-time thermal cycler according to the following protocol, but stop cycling before amplification yield plateaus (generally 8-12 cycles):

11. 95C for 5 min
12. 95C for 30 sec
13. 58C for 60 sec
14. 72C for 60 sec
15. goto 2, N more times

Perform sequencing (e.g., next-generation sequencing) on sample for digital quantitation according to manufacturer's instructions (e.g., Illumina, ABI)

Example 6: MIP-capture reaction of a set of exon target nucleic acids

MIP probes are synthesized using standard column-based oligonucleotide synthesis by any number of vendors (e.g. IDT).

Hybridize capture probes to human genomic sample

Assemble hybridization reaction:

- 1.0 ul capture probe mix (-2.5 pmol)
- 2.0 ul 10X Ampligase buffer (Epicentre)
- 6.0 ul 500 ng/ul human genomic DNA (-16.7 fmol)
- 11 ul dH20

In a thermal cycler, heat reaction to 95C for 5 min to denature genomic DNA, then cool to 60C. Allow to incubate at 60C for 40 hours.

Convert hybridized probes into covalently-closed circular products containing target nucleic acids

Prepare fill-in / ligation reaction mixture:

- 0.25 ul 2 mM dNTP mix (Invitrogen)
- 2.5 ul 10X Ampligase buffer (Epicentre)
- 5.0 ul 5 U/ul Taq Stoffel fragment (Applied Biosystems)
- 12.5 ul 5 U/ul Ampligase (Epicentre)
- 4.75 ul dH20

Add 1.0 ul of this mix to the hybridized probe reaction, and incubate at 60C for 10 hours.

Petitioner Illumina, Inc., EX1004, p. 731

Purify, circularized probe/target nucleic acid products from un-reacted probes and genomic DNA

Prepare exonuclease reaction mixture:

- 21 ul fill-in / ligation reaction product
- 2.0 ul 10X exonuclease I buffer (New England Biolabs)
- 2.0 ul 20 U/ul exonuclease I (New England Biolabs)
- 2.0 ul 100 U/ul exonuclease III (New England Biolabs)

Incubate at 37C for 60 min, then heat-inactivate by incubating at 80C for 15 min. Immediately cool to 4C for storage.

Amplify circular material by PCR using primers specific to the 'constant region' of the probes

Prepare PCR mixture:

- 5.0 ul 10X Accuprime reaction buffer (Invitrogen)
- 1.5 ul 10 uM CP-2-FA (5'-GCACGATCCGACGGTAGTGT-3') (SEQ ID NO: 183)
-
- 1.5 ul 10 uM CP-2-RA (5'-CCGTAATCGGGAAGCTGAAG-3') (SEQ ID NO: 184)
- 0.4 ul 25 mM dNTP mix (Invitrogen)
- 2.0 ul heat-inactivated exonuclease reaction mix
- 1.5 ul 10X SybrGreen (Invitrogen)
- 0.4 ul 2.5 U/ul Accuprime Pfx polymerase (Invitrogen)
- 37.7 ul dH20

Thermal cycle in real-time thermal cycler according to the following protocol, but stop cycling before amplification yield plateaus (generally 8-12 cycles):

16. 95C for 5 min
17. 95C for 30 sec
18. 58C for 60 sec
19. 72C for 60 sec
20. goto 2, N more times

Petitioner Illumina, Inc., EX1004, p. 732

Prepare a shotgun next-generation sequencing library for analysis

- Purify desired amplicon population from non-specific amplification products by gel extraction.
- Concatemerize amplicons into high-molecular weight products suitable for shearing
- Mechanically shear, using either a nebulizer, BioRuptor, Hydroshear, Covaris, or similar instrument. DNA should be sheared into fragments several hundred basepairs in length.
- Ligate adapters required for amplification by the sequencing platform used. If necessary, purify ligated product from unligated product and adapters.

Perform sequencing of library according to manufacturer's directions (e.g. Illumina, ABI, etc), reading the target sequence to determine abundance of the target nucleic acid.

Example 7: Use of MIPs, hybridization, and mutation-detection MIPs to genotype a set of 1000 targets

MIPs, hybridization, and mutation-detection MIPs are used to genotype a set of 1000 targets. The protocol permits detection of any of 50 specific known point mutations

First, separate MIP, hybridization, and mutation-detection MIP reactions are performed on a biological sample. A MIP capture reaction is performed essentially as described in Turner et al 2009, Nature methods 6:315-6. A set of MIPs is designed such that each probe in the set flanks one of the 1000 targets. Separately, a hybridization enrichment reaction is performed using the Agilent SureSelect procedure. Prior to selection, the genomic DNA to be enriched is converted into a shotgun sequencing library using Illumina's 'Fragment Library' kit and protocol. Agilent's web interface is used to design a set of probes which will hybridize to the target nucleic acids. Separately, a set of probes are designed (mutation-detection MIPs) which will form MIPs only if mutations (e.g., specific polymorphisms) are present. Each mutation-detection MIP has a 3'-most base identity that is specific for a single known mutation. A reaction with this set of mutation-detection MIPs is performed to selectively detect the presence of any mutant alleles.

Once all three reactions have been performed, the two MIP reactions are combined (e.g., at potentially non-equimolar ratios to further increase sensitivity of mutation detection) into a

Petitioner Illumina, Inc., EX1004, p. 733

single tube, and run as one sample on the next-generation DNA sequencing instrument. The hybridization-enriched reaction is run as a separate sample on the next-generation DNA sequencing instrument. Reads from each 'sample' are combined by a software algorithm which forms a consensus diploid genotype at each position in the target set by evaluating the total coverage at each position, the origin of each read in that total coverage, the quality score of each individual read, and the presence (or absence) of any reads derived from mutation-specific MIPs overlapping the region.

It should be appreciated that the preceding examples are non-limiting and aspects of the invention may be implemented as described herein using alternative techniques and/or protocols that are available to one or ordinary skill in the art.

It will be clear that the methods may be practiced other than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present disclosure are possible in light of the above teachings and, therefore, are within the scope of the claims. Preferred features of each aspect of the Disclosure are as for each of the other aspects mutatis mutandis. The documents including patents, patent applications, journal articles, or other disclosures mentioned herein are hereby incorporated by reference in their entirety. In the event of conflict, the disclosure of present application controls, other than in the event of clear error.

Doc Code: CRF.TRNS.REQ
Document Description: Request for Transfer of a Computer Readable Form

PTO/SB/93 (11-14)
Approved for use through 09/30/2022. OMB 0651-0024
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

REQUEST FOR TRANSFER OF A COMPUTER READABLE FORM UNDER 37 CFR 1.821(e)

Application No.:	16/952,764	First Named Inventor:	Gregory Porreca
Filing Date:	November 19, 2020	Practitioner Docket No.:	MLB-002/06US 34468/40
Title of the Invention:	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS		

The sequence information in the paper copy or PDF file of the Sequence Listing submitted in this application in accordance with 37 CFR 1.821(c),

- ■ as part of the originally-filed specification of this application
- ☐ as a separate amendment of the specification filed
  - ☐ herewith
  - ☐ on _____

is identical to the sequence information in the

- ☐ only computer readable form
- ■ last filed computer readable form
- ☐ other computer readable form, which was filed on _____

in application number 15/231,687, which was filed on August 8, 2016.

This computer readable form was compliant with 37 CFR 1.821-1.825, and applicant hereby requests that it be used as the computer readable form for the present application, in accordance with 37 CFR 1.821(e).

The above referenced paper copy or PDF file of the Sequence Listing contains no new matter.

A sequence listing text file submitted via EFS-Web that complies with the requirements of 37 CFR 1.824(a) (2)-(6) and (b) (i.e., is a compliant sequence listing ASCII text file), serves as both the paper copy required by 37 CFR 1.821(c) and the CRF required by 37 CFR 1.821(e). If a user submits a compliant sequence listing ASCII text file via EFS-Web, the U.S. Patent and Trademark Office will not carry out a request to use a compliant computer readable "Sequence Listing" that is already on file for another application pursuant to 37 CFR 1.821(e) but will use the sequence listing submitted with the application as originally filed via EFS-Web.

It is understood that upon the transfer of a copy of the computer readable form to this application, the U.S. Patent and Trademark Office will update the copy of the computer readable form to reflect the application number and filing date for this application.

Signature /Thomas C. Meyers/	Date December 3, 2020
Name (Print/Typed) Thomas C. Meyers	Practitioner Registration Number 36,989

Note: *This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4(d) for signature requirements and certifications. Submit multiple forms if more than one signature is required.**

1 *Total of 1 forms are submitted.

This collection of information is required by 35 U.S.C. 22 and 37 CFR 1.821-1.825. The information is required to obtain or retain a benefit by the public, which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 6 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS.

Privacy Act Statement

The **Privacy Act of 1974 (P.L. 93-579)** requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether disclosure of these records is required by the Freedom of Information Act.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (*i.e.*, GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspection or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING or 371(c) DATE	GRP ART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
16/952,764	11/19/2020		830	MLB-002/06US 34468/40	9	2

CONFIRMATION NO. 7812

UPDATED FILING RECEIPT

21710
BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

OC000000122006058

Date Mailed: 12/15/2020

Receipt is acknowledged of this non-provisional utility patent application. The application will be taken up for examination in due course. Applicant will be notified as to the results of the examination. Any correspondence concerning the application must include the following identification information: the U.S. APPLICATION NUMBER, FILING DATE, NAME OF FIRST INVENTOR, and TITLE OF INVENTION. Fees transmitted by check or draft are subject to collection.

Please verify the accuracy of the data presented on this receipt. If an error is noted on this Filing Receipt, please submit a written request for a corrected Filing Receipt, including a properly marked-up ADS showing the changes with strike-through for deletions and underlining for additions. If you received a "Notice to File Missing Parts" or other Notice requiring a response for this application, please submit any request for correction to this Filing Receipt with your reply to the Notice. When the USPTO processes the reply to the Notice, the USPTO will generate another Filing Receipt incorporating the requested corrections provided that the request is grantable.

Inventor(s)

Gregory Porreca, Cambridge, MA;
Uri Laserson, Boston, MA;

Applicant(s)

Molecular Loop Biosolutions, LLC, Cambridge, MA;

Power of Attorney: None

Domestic Priority data as claimed by applicant

This application is a CON of 15/231,687 08/08/2016
which is a CON of 13/266,862 03/13/2012
which is a 371 of PCT/US10/01293 04/30/2010
which claims benefit of 61/182,089 05/28/2009
and claims benefit of 61/179,358 05/18/2009
and claims benefit of 61/178,923 05/15/2009
and claims benefit of 61/174,470 04/30/2009

Foreign Applications for which priority is claimed (You may be eligible to benefit from the **Patent Prosecution Highway** program at the USPTO. Please see http://www.uspto.gov for more information.) - None.
Foreign application information must be provided in an Application Data Sheet in order to constitute a claim to foreign priority. See 37 CFR 1.55 and 1.76.

Permission to Access Application via Priority Document Exchange: Yes

Petitioner Illumina, Inc., EX1004, p. 737

Permission to Access Search Results: Yes

Applicant may provide or rescind an authorization for access using Form PTO/SB/39 or Form PTO/SB/69 as appropriate.

If Required, Foreign Filing License Granted: 11/30/2020

The country code and number of your priority application, to be used for filing abroad under the Paris Convention, is **US 16/952,764**

Projected Publication Date: Request for Non-Publication Acknowledged

Non-Publication Request: Yes

**********WARNING**********

Applicant has requested non-publication, certifying that the invention disclosed in this application has not been and will not be the subject of an application filed in another country, or under a multilateral agreement, that requires publication at eighteen months after filing; **and has also claimed** the benefit of an application filed in another country or under a multilateral agreement. Because the non-publication request and corresponding certification appears to be inconsistent with the priority claim, applicant is reminded that he/she has an obligation to promptly notify the Office and **rescind the non-publication request** if the invention disclosed in this application is the subject of an application filed in another country, or under a multilateral agreement, that requires eighteen-month publication.

**********WARNING**********

Early Publication Request: No
**** SMALL ENTITY ****
Title

METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS

Preliminary Class

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications: No

PROTECTING YOUR INVENTION OUTSIDE THE UNITED STATES

Since the rights granted by a U.S. patent extend only throughout the territory of the United States and have no effect in a foreign country, an inventor who wishes patent protection in another country must apply for a patent in a specific country or in regional patent offices. Applicants may wish to consider the filing of an international application under the Patent Cooperation Treaty (PCT). An international (PCT) application generally has the same effect as a regular national patent application in each PCT-member country. The PCT process **simplifies** the filing of patent applications on the same invention in member countries, but **does not result** in a grant of "an international patent" and does not eliminate the need of applicants to file additional documents and fees in countries where patent protection is desired.

Almost every country has its own patent law, and a person desiring a patent in a particular country must make an application for patent in that country in accordance with its particular laws. Since the laws of many countries differ in various respects from the patent law of the United States, applicants are advised to seek guidance from specific foreign countries to ensure that patent rights are not lost prematurely.

Applicants also are advised that in the case of inventions made in the United States, the Director of the USPTO must issue a license before applicants can apply for a patent in a foreign country. The filing of a U.S. patent application

serves as a request for a foreign filing license. The application's filing receipt contains further information and guidance as to the status of applicant's license for foreign filing.

Applicants may wish to consult the USPTO booklet, "General Information Concerning Patents" (specifically, the section entitled "Treaties and Foreign Patents") for more information on timeframes and deadlines for filing foreign patent applications. The guide is available either by contacting the USPTO Contact Center at 800-786-9199, or it can be viewed on the USPTO website at http://www.uspto.gov/web/offices/pac/doc/general/index.html.

For information on preventing theft of your intellectual property (patents, trademarks and copyrights), you may wish to consult the U.S. Government website, http://www.stopfakes.gov. Part of a Department of Commerce initiative, this website includes self-help "toolkits" giving innovators guidance on how to protect intellectual property in specific countries such as China, Korea and Mexico. For questions regarding patent enforcement issues, applicants may call the U.S. Government hotline at 1-866-999-HALT (1-866-999-4258).

LICENSE FOR FOREIGN FILING UNDER

Title 35, United States Code, Section 184

Title 37, Code of Federal Regulations, 5.11 & 5.15

GRANTED

The applicant has been granted a license under 35 U.S.C. 184, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" followed by a date appears on this form. Such licenses are issued in all applications where the conditions for issuance of a license have been met, regardless of whether or not a license may be required as set forth in 37 CFR 5.15. The scope and limitations of this license are set forth in 37 CFR 5.15(a) unless an earlier license has been issued under 37 CFR 5.15(b). The license is subject to revocation upon written notification. The date indicated is the effective date of the license, unless an earlier license of similar scope has been granted under 37 CFR 5.13 or 5.14.

This license is to be retained by the licensee and may be used at any time on or after the effective date thereof unless it is revoked. This license is automatically transferred to any related applications(s) filed under 37 CFR 1.53(d). This license is not retroactive.

The grant of a license does not in any way lessen the responsibility of a licensee for the security of the subject matter as imposed by any Government contract or the provisions of existing laws relating to espionage and the national security or the export of technical data. Licensees should apprise themselves of current regulations especially with respect to certain countries, of other agencies, particularly the Office of Defense Trade Controls, Department of State (with respect to Arms, Munitions and Implements of War (22 CFR 121-128)); the Bureau of Industry and Security, Department of Commerce (15 CFR parts 730-774); the Office of Foreign AssetsControl, Department of Treasury (31 CFR Parts 500+) and the Department of Energy.

NOT GRANTED

No license under 35 U.S.C. 184 has been granted at this time, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" DOES NOT appear on this form. Applicant may still petition for a license under 37 CFR 5.12, if a license is desired before the expiration of 6 months from the filing date of the application. If 6 months has lapsed from the filing date of this application and the licensee has not received any indication of a secrecy order under 35 U.S.C. 181, the licensee may foreign file the application pursuant to 37 CFR 5.15(b).

Petitioner Illumina, Inc., EX1004, p. 739

SelectUSA

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The U.S. offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to promote and facilitate business investment. SelectUSA provides information assistance to the international investor community; serves as an ombudsman for existing and potential investors; advocates on behalf of U.S. cities, states, and regions competing for global investment; and counsels U.S. economic development organizations on investment attraction best practices. To learn more about why the United States is the best country in the world to develop technology, manufacture products, deliver services, and grow your business, visit http://www.SelectUSA.gov or call +1-202-482-6800.

Petitioner Illumina, Inc., EX1004, p. 740

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/952,764	11/19/2020	Gregory Porreca	MLB-002/06US 34468/40	7812

21710 7590 01/12/2021

BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

EXAMINER	
ART UNIT	PAPER NUMBER
1634	

NOTIFICATION DATE	DELIVERY MODE
01/12/2021	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

ip@brownrudnick.com
usactions@brownrudnick.com

Decision Granting Request for Prioritized Examination (Track I)	**Application No.** 16/952,764	**Applicant(s)** Porreca et al.	
	Examiner DIANE C GOODWYN	**Art Unit** OPET	**AIA (FITF) Status** No

1. THE REQUEST FILED 19 November 2020 IS **GRANTED** .

The above-identified application has met the requirements for prioritized examination

A. ☑ for an original nonprovisional application (Track I).
B. ☐ for an application undergoing continued examination (RCE).

2. **The above-identified application will undergo prioritized examination**. The application will be accorded special status throughout its entire course of prosecution until one of the following occurs:

A. filing a **petition for extension of time** to extend the time period for filing a reply;

B. filing an **amendment to amend the application to contain more than four independent claims, more than thirty total claims** , or a multiple dependent claim;

C. filing a **request for continued examination** ;

D. filing a notice of appeal;

E. filing a request for suspension of action;

F. mailing of a notice of allowance;

G. mailing of a final Office action;

H. completion of examination as defined in 37 CFR 41.102; or

I. abandonment of the application.

Telephone inquiries with regard to this decision should be directed to DIANE GOODWYN at (571) 272-6735. In his/her absence, calls may be directed to Petition Help Desk at (571) 272-3282.

/DIANE C GOODWYN/ Paralegal Specialist, OPET	

U.S. Patent and Trademark Office
PTO-2298 (Rev. 02-2012)

Electronic Acknowledgement Receipt

EFS ID:	41659306
Application Number:	16952764
International Application Number:	
Confirmation Number:	7812
Title of Invention:	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS
First Named Inventor/Applicant Name:	Gregory Porreca
Customer Number:	21710
Filer:	Thomas C. Meyers/Kelley Warren
Filer Authorized By:	Thomas C. Meyers
Attorney Docket Number:	MLB-002/06US 34468/40
Receipt Date:	15-JAN-2021
Filing Date:	19-NOV-2020
Time Stamp:	16:41:41
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Transmittal Letter	Information_Disclosure_Statement.pdf	25932 3df62bbd4322820d3fb58a98b2f8b6a7a7676adb	no	2

Warnings:

Information:					
2	Information Disclosure Statement (IDS) Form (SB08)	34468_40US_-_Information_Disclosure_Statement_Fillable_PDF_1_of_6.pdf	1046574 880e987eb90f3b7caaeb03b3be54bae569f268be	no	40
Warnings:					
Information:					
3	Information Disclosure Statement (IDS) Form (SB08)	34468_40US_-_Information_Disclosure_Statement_Fillable_PDF_2_of_6.pdf	1039597 d597d5d6a5b3cb3e8251f4989a105a856241c3f1	no	14
Warnings:					
Information:					
4	Information Disclosure Statement (IDS) Form (SB08)	34468_40US_-_Information_Disclosure_Statement_Fillable_PDF_3_of_6.pdf	1037799 54255f3b2aad185f694c3b393b4c1fb99e438fee	no	8
Warnings:					
Information:					
A U.S. Patent Number Citation or a U.S. Publication Number Citation is required in the Information Disclosure Statement (IDS) form for autoloading of data into USPTO systems. You may remove the form to add the required data in order to correct the Informational Message if you are citing U.S. References. If you chose not to include U.S. References, the image of the form will be processed and be made available within the Image File Wrapper (IFW) system. However, no data will be extracted from this form. Any additional data such as Foreign Patent Documents or Non Patent Literature will be manually reviewed and keyed into USPTO systems.					
5	Information Disclosure Statement (IDS) Form (SB08)	34468_40US_-_Information_Disclosure_Statement_Fillable_PDF_4_of_6.pdf	1037988 a9a24dfeb424ad323a468c07b91207a7624e9a7e	no	8
Warnings:					
Information:					
A U.S. Patent Number Citation or a U.S. Publication Number Citation is required in the Information Disclosure Statement (IDS) form for autoloading of data into USPTO systems. You may remove the form to add the required data in order to correct the Informational Message if you are citing U.S. References. If you chose not to include U.S. References, the image of the form will be processed and be made available within the Image File Wrapper (IFW) system. However, no data will be extracted from this form. Any additional data such as Foreign Patent Documents or Non Patent Literature will be manually reviewed and keyed into USPTO systems.					
6	Information Disclosure Statement (IDS) Form (SB08)	34468_40US_-_Information_Disclosure_Statement_Fillable_PDF_5_of_6.pdf	1038092 8e9906c3d50b779bc684bf510aab15064e3f0a86	no	8
Warnings:					
Information:					
A U.S. Patent Number Citation or a U.S. Publication Number Citation is required in the Information Disclosure Statement (IDS) form for autoloading of data into USPTO systems. You may remove the form to add the required data in order to correct the Informational Message if you are citing U.S. References. If you chose not to include U.S. References, the image of the form will be processed and be made available within the Image File Wrapper (IFW) system. However, no data will be extracted from this form. Any additional data such as Foreign Patent Documents or Non Patent Literature will be manually reviewed and keyed into USPTO systems.					

7	Information Disclosure Statement (IDS) Form (SB08)	34468_40US_-_Information_Disclosure_Statement_Fillable_PDF_6_of_6.pdf	1038034 81ae7082d277e2ed0dcaafaf23b5c6f2a3748848	no	8
Warnings:					
Information:					
A U.S. Patent Number Citation or a U.S. Publication Number Citation is required in the Information Disclosure Statement (IDS) form for autoloading of data into USPTO systems. You may remove the form to add the required data in order to correct the Informational Message if you are citing U.S. References. If you chose not to include U.S. References, the image of the form will be processed and be made available within the Image File Wrapper (IFW) system. However, no data will be extracted from this form. Any additional data such as Foreign Patent Documents or Non Patent Literature will be manually reviewed and keyed into USPTO systems.					
8	Non Patent Literature	Chennagiri-2013-ASHGmtg.pdf	156838 5f25b7a42361a22f3197ffe8c3744c47c38d9f22	no	2
Warnings:					
Information:					
9	Non Patent Literature	Chevreux-1999-ProcGCB.pdf	665185 ebb573b3007f33a9a5878d9dc2c6302487236d7f	no	13
Warnings:					
Information:					
10	Non Patent Literature	Deng-2012-NB.PDF	887689 ddf1bd81f62283cec7994909f52a9b43ed8a6737	no	12
Warnings:					
Information:					
11	Non Patent Literature	Frey-2006-StatHacks.pdf	850589 c3b76c169810a3ac78ff7a13aee12d249cde397a	no	9
Warnings:					
Information:					
12	Non Patent Literature	Goto-2010-Bioinform_.pdf	261554 9cfce92738b0b8d483d3f2051346d6aab78a2150	no	3
Warnings:					
Information:					
13	Non Patent Literature	Guerrero-Fernandez-2013-Granada.pdf	406234 16b9f3afa920527b86036e6ce34af8bc21e04cea	no	8

Warnings:					
Information:					
14	Non Patent Literature	Gustincich-Biotechniques-1991.pdf	1830923 8c5c119be0616a6c03f74095dd8191ab6837b04d	no	4
Warnings:					
Information:					
15	Non Patent Literature	Homer-2008-PLoSOne.pdf	1991422 17b5069588491f0d29c6e15bfaf92bcb23e13ebf	no	9
Warnings:					
Information:					
16	Non Patent Literature	29168_244US_-_Kambara-1988-BioTechnol_11.pdf	621210 2cf3332011ffb01f16b98f26009ec98f8564df35	no	6
Warnings:					
Information:					
17	Non Patent Literature	Kennedy-2013-ASHGmtg.pdf	144805 68b3e8e0a071207ea92fa1e57e9ac0acfc670355	no	2
Warnings:					
Information:					
18	Non Patent Literature	Larkin-2007-Bioinformatics.pdf	353957 08b973bd020a1420163ba85cc80d3c4c77d495be	no	2
Warnings:					
Information:					
19	Non Patent Literature	May1988.pdf	958586 5dee85932e7bb13c531451af931551868508 35a2	no	9
Warnings:					
Information:					
20	Non Patent Literature	McDonnell.pdf	60837 009fc900237fef2078faead37164854d2b87dc4b	no	1
Warnings:					
Information:					

21	Non Patent Literature	Miesenbock-1998a.pdf	60877 94d158a2236f5d87cdb94a731198801e062b85d5	no	1
Warnings:					
Information:					
22	Non Patent Literature	Moudrianakis-1965-PNAS.pdf	1325856 c72874d4ac699c7ee27daa5b6cf4374633703e0e	no	8
Warnings:					
Information:					
23	Non Patent Literature	Rosendahl-2013-Gut.pdf	345735 bfedac8c59713770bf0ee971711d5a4980736304	no	12
Warnings:					
Information:					
24	Non Patent Literature	Sauro-How-Do-You-Calculate-2004.pdf	142346 00d24abbda607cafbde88c8c684a9713180199ef	no	3
Warnings:					
Information:					
25	Non Patent Literature	Sunnucks-1996-Genetics.pdf	987812 3b48daeaf487176b50f7669baf23047fee123fcc	no	10
Warnings:					
Information:					
			Total Files Size (in bytes):	18316471	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Docket No.: MLB-002/06US
34468/40
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of:
Molecular Loop Biosolutions, LLC

Application No.: 16/952,764 | Confirmation No.: 7812

Filed: November 19, 2020 | Art Unit: N/A

For: METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS | Examiner: Not Yet Assigned

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

INFORMATION DISCLOSURE STATEMENT (IDS)

Pursuant to 37 C.F.R. §§ 1.56, 1.97, and 1.98, the attention of the United States Patent and Trademark Office is hereby directed to the references listed on the attached PTO/SB/08. It is respectfully requested that the information be expressly considered during the prosecution of the above-identified application, and that the references be made of record therein and appear among the "References Cited" on any patent to issue therefrom.

Copies of the references listed below are attached:

Non-Patent Literature or Other Documents
Chennagiri, 2013, A generalized scalable database model for storing and exploring genetic variations detected using sequencing data, American Society of Human Genetics 63rd Annual Meeting, Abstract, October 22, 2013
Chevreux, 1999, Genome sequence assembly using trace signals and additional sequence information, Proc GCB 99:45-56
Deng et. al., 2012, Supplementary Material, Nature Biotechnology, S1-1 - S1-1 1, Retrieved from the Internet on 2012-10-24
Frey, 2006, Statistics Hacks 108-115.
Goto, 2010, BioRuby: bioinformatics software for the Ruby programming language, Bioinformatics 26(20):2617-2619
Guerrero-Fernandez, 2013, FQbin: a compatible and optimize dformat for storing and managing sequence data, IWBBIO Proceedings, Granada 337-344
Gustincich, 1991, A fast method for high-quality genomic DNA extraction from whole human blood, BioTechniques 11(3):298-302

63948143

Homer, 2008, Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays. PLoS One 4(8):e1000167
Kambara, 1988, Optimization of Parameters in a DNA Sequenator Using Fluorescence Detection, Nature Biotechnology 6:816-821
Kennedy, 2013, Accessing more human genetic variation with short sequencing reads, American Society of Human Genetics 63rd Annual Meeting, Abstract, October 22, 2013
Larkin, 2007, Clustal W and Clustal X version 2.0, Bioinformatics, 23(21):2947-2948
May, 1988, How Many Species Are There on Earth?, Science 241(4872):1441-9
McDonnell, 2007, Antisepsis, disinfection, and sterilization: types, action, and resistance, page 239
Miesenbock, 1998, Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins, Nature 394(6689):192-95
Moudrianakis, 1965, Base Sequence Determination in Nucleic Acids with the Electron Microscope, III. Chemistry and Microscopy of Guanine-Labeled DNA, PNAS, 53:564-71
Rosendahl, 2013, CFTR, SPINK1, CTRC and PRSS1 variants in chronic pancreatitis: is the role of mutated CFTR overestimated?, Gut 62:582-592
Sauro, 2004, How Do You Calculate a Z-Score/ Sigma Level?, https://www.measuringusability.com/zcalc.htm (online publication)
Sunnucks, 1996, Microsatellite and chromosome evolution of parthenogenetic sitobion aphids in Australia, Genetics 144:747-756

In accordance with 37 C.F.R. § 1.98(d)(1) the remaining references are not supplied because they were previously cited by or submitted to the Office in prior application numbers 15/231687 filed August 8, 2016 and 13/266862 filed March 13, 2012 and relied on in the above-identified application for an earlier effective filing date under 35 U.S.C. § 120.

It is submitted that the Information Disclosure Statement is in compliance with 37 C.F.R. § 1.98, and the Examiner is respectfully requested to consider the listed references.

Applicant believes no fee is due with this response.

Dated: January 15, 2021

Respectfully submitted,

Electronic signature: /Thomas C. Meyers/
Thomas C. Meyers
Registration No.: 36,989
BROWN RUDNICK LLP
Brown Rudnick LLP
One Financial Center
Boston, Massachusetts 02111
(617) 856-8200
Attorney for Applicant

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (**Not for submission under 37 CFR 1.99)**		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

See attached certification statement.

The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/Thomas C. Meyers/	Date (YYYY-MM-DD)	2021-01-15
Name/Print	Thomas C. Meyers	Registration Number	36,989

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

Doc code: IDS
Doc description: Information Disclosure Statement (IDS) Filed

PTO/SB/08a (02-18)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (**Not for submission under 37 CFR 1.99)**		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

U.S.PATENTS Remove

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	4683195		1987-07-28	Mullis et al.	
	2	4683202		1987-07-28	Mullis	
	3	4988617		1991-01-29	Landegren et al.	
	4	5060980		1991-10-29	Johnson et al.	
	5	5210015		1993-05-11	Gelfand et al.	
	6	5234809		1993-08-10	Boom et al.	
	7	5242794		1993-09-07	Whiteley et al.	
	8	5348853		1994-09-20	Wang et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	9	5434049	A	1995-07-18	Okano et al.	
	10	5459307		1995-10-17	Klotz, Jr.	
	11	5486686		1996-01-23	Zdybel, Jr. et al.	
	12	5491224	A	1996-02-13	Bittner et al.	
	13	5494810		1996-02-27	Barany et al.	
	14	5567583		1996-10-22	Wang et al.	
	15	5583024		1996-12-10	McElroy, deceased et al.	
	16	5604097		1997-02-18	Brenner	
	17	5636400		1997-06-10	Young	
	18	5674713		1997-10-07	McElroy, deceased et al.	
	19	5695934		1997-12-09	Brenner	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	20	5700673		1997-12-23	McElroy, deceased et al.	
	21	5701256		1997-12-23	Marr et al.	
	22	5830064		1998-11-03	Bradish et al.	
	23	5846719		1998-12-08	Brenner et al.	
	24	5863722		1999-01-26	Brenner	
	25	5866337		1999-02-02	Schon	
	26	5869252		1999-02-09	Bouma et al.	
	27	5869717		1999-02-09	Frame et al.	
	28	5871921		1999-02-16	Landegren et al.	
	29	5888788		1999-03-30	De Miniac	
	30	5942391		1999-08-24	Zhang et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	31	5971921		1999-10-26	Timbel	
	32	5993611		1999-11-30	Moroney, III et al.	
	33	5994056		1999-11-30	Higuchi	
	34	6020127	A	2000-02-01	MacKenzie et al.	
	35	6033854		2000-03-07	Kumit et al.	
	36	6033872		2000-03-07	Bergsma et al.	
	37	6100099		2000-08-08	Gordon et al.	
	38	6138077		2000-10-24	Brenner	
	39	6150516		2000-11-21	Brenner et al.	
	40	6171785		2001-01-09	Higuchi	
	41	6172214		2001-01-09	Brenner	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	42	6172218		2001-01-09	Brenner	
	43	6197508		2001-03-06	Stanley	
	44	6197574		2001-03-06	Miyamoto et al.	
	45	6210891		2001-04-03	Nyren et al.	
	46	6223128	B1	2001-04-24	Allex et al.	
	47	6235472		2001-05-22	Landegren et al.	
	48	6235475		2001-05-22	Brenner et al.	
	49	6235501		2001-05-22	Gautsch et al.	
	50	6235502		2001-05-22	Weissman et al.	
	51	6258568		2001-07-10	Nyren	
	52	6274320		2001-08-14	Rothberg et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	53	6306597		2001-10-23	Macevicz	
	54	6352828		2002-03-05	Brenner	
	55	6360235		2002-03-19	Tilt et al.	
	56	6361940		2002-03-26	Van Ness et al.	
	57	6403320		2002-06-11	Read et al.	
	58	6462254		2002-10-08	Vernachio et al.	
	59	6489105		2002-12-03	Matlashewski et al.	
	60	6558928		2003-05-06	Landegren	
	61	6569920		2003-05-27	Wen et al.	
	62	6582938		2003-06-24	Su et al.	
	63	6585938		2003-07-01	Machida et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	64	6613516		2003-09-02	Christians et al.	
	65	6714874		2004-03-30	Myers et al.	
	66	6716580		2004-04-06	Gold et al.	
	67	6719449		2004-04-13	Laugharn, Jr. et al.	
	68	6818395	B1	2004-11-16	Quake et al.	
	69	6828100		2004-12-07	Ronaghi	
	70	6833246		2004-12-21	Balasubramanian	
	71	6858412		2005-02-22	Willis et al.	
	72	6911345		2005-06-28	Quake et al.	
	73	6913879		2005-07-05	TeleChem International Inc.	
	74	6927024		2005-08-09	Dodge et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	75	6941317		2005-09-06	Chamberlin et al.	
	76	6948843		2005-09-27	Laugharn, Jr. et al.	
	77	7034143		2006-04-25	Preparata et al.	
	78	7041481		2006-05-09	Anderson et al.	
	79	7049077		2006-05-23	Yang	
	80	7057026		2006-06-06	Barnes et al.	
	81	7071324		2006-07-04	Preparata et al.	
	82	7074564		2006-07-11	Landegren	
	83	7074586		2006-07-11	Cheronis et al.	
	84	7115400		2006-10-03	Adessi et al.	
	85	7169560		2007-01-30	Lapidus et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	86	7211390		2007-05-01	Rothberg et al.	
	87	7232656		2007-06-19	Balasubramanian et al.	
	88	7244559		2007-07-17	Rothberg et al.	
	89	7264929		2007-09-04	Rothberg et al.	
	90	7282337		2007-10-16	Harris	
	91	7297518		2007-11-20	Quake et al.	
	92	7320860		2008-01-22	Landegren et al.	
	93	7323305		2008-01-29	Leamon et al.	
	94	7335762		2008-02-26	Rothberg et al.	
	95	7351528		2008-04-01	Landegren	
	96	7393665		2008-07-01	Brenner	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	97	7510829		2009-03-31	Faham et al.	
	98	7523117		2009-04-21	Zhang et al.	
	99	7537889		2009-05-26	Sinha et al.	
	100	7537897		2009-05-26	Brenner et al.	
	101	7544473		2009-06-09	Brenner	
	102	7582431		2009-09-01	Drmanac et al.	
	103	7598035		2009-10-06	Macevicz	
	104	7629151		2009-12-08	Gold et al.	
	105	7642056		2010-01-05	Ahn et al.	
	106	7666593		2010-02-23	Lapidus	
	107	7700323		2010-04-20	Willis et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	108	7774962		2010-08-17	Ladd	
	109	7776616		2010-08-17	Heath et al.	
	110	7790388		2010-09-07	Landegren et al.	
	111	7809509		2010-10-05	Milosavljevic	
	112	7835871		2010-11-16	Kain et al.	
	113	7862999		2011-01-04	Zheng et al.	
	114	7865534		2011-01-04	Genstruct	
	115	7883849		2011-02-08	Dahl	
	116	7957913		2011-06-07	Chinitz et al.	
	117	7960120		2011-06-14	Rigatti et al.	
	118	7985716		2011-07-26	Yershov et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	119	7993880		2011-08-09	Willis et al.	
	120	8024128		2011-09-20	Rabinowitz et al.	
	121	8114027		2012-02-14	Triva	
	122	8165821		2012-04-24	Zhang	
	123	8209130		2012-06-26	Kennedy et al.	
	124	8283116		2012-10-09	Bhattacharyya et al.	
	125	8462161		2013-06-11	Barber	
	126	8463895		2013-06-11	Arora et al.	
	127	8474228		2013-07-02	Adair et al.	
	128	8496166		2013-07-30	Burns et al.	
	129	8529744		2013-09-10	Marziali et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	130	8738300		2014-05-27	Porreca et al.	
	131	8778609		2014-07-15	Umbarger	
	132	8812422		2014-08-19	Nizzari et al.	
	133	8847799		2014-09-30	Kennedy et al.	
	134	8976049		2015-03-10	Kennedy et al.	
	135	9074244		2015-07-07	Sparks et al.	
	136	9115387		2015-08-25	Umbarger	
	137	9228233		2016-01-05	Kennedy et al.	
	138	9292527		2016-03-22	Kennedy et al.	
	139	9535920		2017-01-03	Kennedy et al.	
	140	9567639		2017-02-14	Oliphant et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	141	9677124		2017-06-13	Umbarger	
	142	10066259	B2	2018-09-04	Gore et al.	
	143	10202637		2019-02-12	Umbarger	
	144	10227635		2019-03-12	Umbarger et al.	
	145	10604799		2020-03-31	Porreca et al.	
	146	10683533		2020-06-16	Umbarger et al.	
	147	RE39793		2007-08-21	Brenner	
	148	RE41780		2010-09-28	Anderson et al.	

If you wish to add additional U.S. Patent citation information please click the Add button. Add

U.S.PATENT APPLICATION PUBLICATIONS Remove

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	20010007742	A1	2001-07-12	LANDERGREN	

EFS Web 2.1.18

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	2	20010046673	A1	2001-11-29	French et al.	
	3	20020001800	A1	2002-01-03	LAPIDUS	
	4	20020040216	A1	2002-04-04	Dumont et al.	
	5	20020042052	A1	2002-04-11	NILSEN et al.	
	6	20020091666	A1	2002-07-11	Rice et al.	
	7	20020164629	A1	2002-11-07	Quake et al.	
	8	20020172954	A1	2002-11-21	Mao et al.	
	9	20020182609	A1	2002-12-05	Arcot	
	10	20020187496	A1	2002-12-12	Andersson et al.	
	11	20020190663	A1	2002-12-19	Rasmussen	
	12	20030166057	A1	2003-09-04	Hildebrand et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	13	20030175709	A1	2003-09-18	Murphy et al.	
	14	20030177105	A1	2003-09-18	Xiao et al.	
	15	20030203370	A1	2003-10-30	Yakhini et al.	
	16	20030208454	A1	2003-11-06	Rienhoff et al.	
	17	20030224384	A1	2003-12-04	Sayood et al.	
	18	20040029264	A1	2004-02-12	Robbins	
	19	20040053275	A1	2004-03-18	Shafer	
	20	20040106112	A1	2004-06-03	Nilsson et al.	
	21	20040121373	A1	2004-06-24	Friedlander et al.	
	22	20040142325	A1	2004-07-22	Mintz et al.	
	23	20040152108	A1	2004-08-05	Keith et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	24	20040161773	A1	2004-08-19	Rogan et al.	
	25	20040170965	A1	2004-09-02	Scholl et al.	
	26	20040171051	A1	2004-09-02	Holloway	
	27	20040197813	A1	2004-10-07	Hoffman et al.	
	28	20040209299	A1	2004-10-21	Pinter et al.	
	29	20050003369	A1	2005-01-06	Christians et al.	
	30	20050026204	A1	2005-02-03	Landegren	
	31	20050032095	A1	2005-02-10	Wigler et al.	
	32	20050048505	A1	2005-03-03	Fredrick et al.	
	33	20050059048	A1	2005-03-17	Gunderson et al.	
	34	20050100900	A1	2005-05-12	Kawashima et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	35	20050112590	A1	2005-05-26	Boom et al.	
	36	20050186589	A1	2005-08-25	Kowalik et al.	
	37	20050214811	A1	2005-09-29	Margulies et al.	
	38	20050244879	A1	2005-11-03	Schumm et al.	
	39	20050250147	A1	2005-11-10	Macevicz	
	40	20050272065	A1	2005-12-08	Lakey et al.	
	41	20060008824	A1	2006-01-12	Ronaghi et al.	
	42	20060019304	A1	2006-01-26	Hardenbol et al.	
	43	20060024681	A1	2006-02-02	Smith et al.	
	44	20060030536	A1	2006-02-09	Yu et al.	
	45	20060078894	A1	2006-04-13	Winkler et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	46	20060149047	A1	2006-07-06	Nanduri et al.	
	47	20060177837	A1	2006-08-10	Borozan et al.	
	48	20060183132	A1	2006-08-17	Fu et al.	
	49	20060192047	A1	2006-08-31	Goossen	
	50	20060195269	A1	2006-08-31	Yeatman et al.	
	51	20060246500	A1	2006-11-02	Browne	
	52	20060263789	A1	2006-11-23	Kincaid	
	53	20060281098	A1	2006-12-14	Miao et al.	
	54	20060286577	A1	2006-12-21	Jia	
	55	20060292585	A1	2006-12-28	Nautiyal et al.	
	56	20060292611	A1	2006-12-28	Berka et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	57	20070009925	A1	2007-01-11	Fang et al.	
	58	20070020640	A1	2007-01-25	McCloskey et al.	
	59	20070042369	A1	2007-02-22	Reese et al.	
	60	20070092883	A1	2007-04-26	Schouten et al.	
	61	20070114362	A1	2007-05-24	Feng et al.	
	62	20070128624	A1	2007-06-07	Gormley et al.	
	63	20070161013	A1	2007-07-12	Hantash	
	64	20070162983	A1	2007-07-12	Hesterkamp et al.	
	65	20070166705	A1	2007-07-19	Milton et al.	
	66	20070212704	A1	2007-09-13	Dong et al.	
	67	20070225487	A1	2007-09-27	Nilsson et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	68	20070238122	A1	2007-10-11	Allbritton et al.	
	69	20070244675	A1	2007-10-18	Shai et al.	
	70	20070264653	A1	2007-11-15	Berlin et al.	
	71	20080003142	A1	2008-01-03	Link et al.	
	72	20080014589	A1	2008-01-17	Link et al.	
	73	20080076118	A1	2008-03-27	Tooke et al.	
	74	20080081330	A1	2008-04-03	Kahvejian	
	75	20080085836	A1	2008-04-10	Kearns et al.	
	76	20080090239	A1	2008-04-17	SHOEMAKER et al.	
	77	20080125324	A1	2008-05-29	Petersdorf et al.	
	78	20080176209	A1	2008-07-24	Muller et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	79	20080269068	A1	2008-10-30	CHURCH et al.	
	80	20080280955	A1	2008-11-13	McCamish	
	81	20080293589	A1	2008-11-27	Shapero	
	82	20090009904	A1	2009-01-08	Yasuna et al.	
	83	20090019156	A1	2009-01-15	Mo et al.	
	84	20090026082	A1	2009-01-29	Rothberg et al.	
	85	20090029385	A1	2009-01-29	Christians et al.	
	86	20090035777	A1	2009-02-05	Kokoris et al.	
	87	20090042206	A1	2009-02-12	Schneider et al.	
	88	20090098551	A1	2009-04-16	Landers et al.	
	89	20090099041	A1	2009-04-16	CHURCH et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	90	20090105081	A1	2009-04-23	Rodesch et al.	
	91	20090119313	A1	2009-05-07	Pearce	
	92	20090127589	A1	2009-05-21	Rothberg et al.	
	93	20090129647	A1	2009-05-21	Dimitrova et al.	
	94	20090156412	A1	2009-06-18	Boyce, IV et al.	
	95	20090163366	A1	2009-06-25	Nickerson et al.	
	96	20090181389	A1	2009-07-16	Li et al.	
	97	20090191565	A1	2009-07-30	Lapidus et al.	
	98	20090192047	A1	2009-07-30	Parr et al.	
	99	20090202984	A1	2009-08-13	Cantor	
	100	20090203014	A1	2009-08-13	WU et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	101	20090220955	A1	2009-09-03	Verrant	
	102	20090226975	A1	2009-09-10	Sabot et al.	
	103	20090233814	A1	2009-09-17	Bashkirov et al.	
	104	20090298064	A1	2009-12-03	Batzoglou et al.	
	105	20090301382	A1	2009-12-10	Patel	
	106	20090318310	A1	2009-12-24	Liu et al.	
	107	20100035243	A1	2010-02-11	Muller et al.	
	108	20100035252	A1	2010-02-11	Rothberg et al.	
	109	20100063742	A1	2010-03-11	Hart et al.	
	110	20100069263	A1	2010-03-18	Shendure et al.	
	111	20100076185	A1	2010-03-25	Adey et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	112	20100086914	A1	2010-04-08	BENTLEY et al.	
	113	20100086926	A1	2010-04-08	Craig et al.	
	114	20100105107	A1	2010-04-29	Hildebrand et al.	
	115	20100137143	A1	2010-06-03	Rothberg et al.	
	116	20100137163	A1	2010-06-03	Link et al.	
	117	20100143908	A1	2010-06-10	Gillevet	
	118	20100159440	A1	2010-06-24	Messier et al.	
	119	20100188073	A1	2010-07-29	Rothberg et al.	
	120	20100196911	A1	2010-08-05	Hoffman et al.	
	121	20100197507	A1	2010-08-05	Rothberg et al.	
	122	20100216151	A1	2010-08-26	LAPIDUS et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	123	20100216153	A1	2010-08-26	LAPIDUS et al.	
	124	20100248984	A1	2010-09-30	Shaffer et al.	
	125	20100282617	A1	2010-11-11	Rothberg et al.	
	126	20100285578	A1	2010-11-11	Selden et al.	
	127	20100297626	A1	2010-11-25	MCKERNAN et al.	
	128	20100300559	A1	2010-12-02	Schultz et al.	
	129	20100300895	A1	2010-12-02	Nobile et al.	
	130	20100301042	A1	2010-12-02	Kahlert	
	131	20100301398	A1	2010-12-02	Rothberg et al.	
	132	20100304982	A1	2010-12-02	Hinz et al.	
	133	20100311061	A1	2010-12-09	KORLACH et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	134	20100330619	A1	2010-12-30	Willis et al.	
	135	20110004413	A1	2011-01-06	Carnevali et al.	
	136	20110009278	A1	2011-01-13	Kain et al.	
	137	20110015863	A1	2011-01-20	Pevzner et al.	
	138	20110021366	A1	2011-01-27	Chinitz et al.	
	139	20110034342	A1	2011-02-10	Fox	
	140	20110092375	A1	2011-04-21	Zamore et al.	
	141	20110098193	A1	2011-04-28	Kingsmore et al.	
	142	20110117544	A1	2011-05-19	Lexow	
	143	20110159499	A1	2011-06-30	HINDSON et al.	
	144	20110166029	A1	2011-07-07	Margulies et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	145	20110224105	A1	2011-09-15	Kurn et al.	
	146	20110230365	A1	2011-09-22	Rohlfs et al.	
	147	20110257889	A1	2011-10-20	Klammer et al.	
	148	20110288780	A1	2011-11-24	Rabinowitz et al.	
	149	20110301042	A1	2011-12-08	Steinmann et al.	
	150	20120015050	A1	2012-01-19	Abkevich et al.	
	151	20120021930	A1	2012-01-26	Schoen et al.	
	152	20120046877	A1	2012-02-23	HYLAND et al.	

If you wish to add additional U.S. Published Application citation information please click the Add button. Add

FOREIGN PATENT DOCUMENTS Remove

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2]	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1	1 321 477	EP	A1	2003-06-25	Grawunder, Ulf		

EFS Web 2.1.18

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	2	1 564 306	EP	A2	2005-08-17	Affymetrix, Inc		
	3	10770071.8	EP		2010-11-04	Good Start Genetics, Inc.		
	4	2 437 191	EP	A2	2012-04-04	Gene Security Network Llc		
	5	2425240	EP	A2	2012-03-07	Good Start Genetics, Inc		
	6	2716766	EP	A1	2014-04-09	Berry Genomics Co., Ltd		
	7	00/018957	WO	A1	2000-04-06	Applied Research Systems		
	8	02/093453	WO	A2	2002-11-21	X-mine, Inc		
	9	1996/019586	WO	A1	1996-06-27	Visible Genetics Inc		
	10	2004/018497	WO	A2	2004-03-04	Solexa Ltd		
	11	2004/083819	WO	A2	2004-09-30	Trace Genetics, Inc		
	12	2005/003304	WO	A2	2005-01-13	Illumina, Inc		

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	13	2006/084132	WO	A2	2006-08-10	Agencourt Bioscience Corp		
	14	2007/010251	WO	A2	2007-01-25	Solexa Ltd		
	15	2007/061284	WO	A1	2007-05-31	Plant Res Int BV		
	16	2007/107717	WO	A1	2007-09-27	Ucl Business Plc		
	17	2007/123744	WO	A2	2007-11-01	Solexa, Inc		
	18	2007/135368	WO	A2	2007-11-29	Solexa Ltd		
	19	2008067551	WO	A2	2008-06-05	Navigenics Inc		
	20	2009/036525	WO	A2	2009-03-26	Univ Leuven Kath		
	21	2009/076238	WO	A2	2009-06-18	Complete Genomics, Inc		
	22	2010/024894	WO	A1	2010-03-04	23andme, Inc		
	23	2010/115154	WO	A1	2010-10-07	Fluidigm Corp		

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	24	2010/126614	WO	A2	2010-11-04	Good Start Genetics, Inc		
	25	2011/006020	WO	A1	2011-01-13	Qualcomm Inc		
	26	2011/102998	WO	A2	2011-08-25	Helicos Biosciences Corp		
	27	2011/155833	WO	A2	2011-12-15	Keygene Nv		
	28	2011066476	WO	A1	2011-06-03	Quantalife, Inc		
	29	2011067378	WO	A1	2011-06-09	Olink Genomics Ab		
	30	2012/006291	WO	A2	2012-01-12	Life Technologies Corp		
	31	2012/051208	WO	A2	2012-04-19	Complete Genomics, Inc		
	32	2012/087736	WO	A1	2012-06-28	Good Start Genetics, Inc		
	33	2012/109500	WO	A2	2012-08-16	Bio-rad Laboratories, Inc		
	34	2012/134884	WO	A1	2012-10-04	Good Start Genetics, Inc		

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	35	2012/149171	WO	A1	2012-11-01	Univ California		
	36	2012/170725	WO	A2	2012-12-13	Sinai School Medicine		
	37	2012040387	WO	A1	2012-03-29	Univ Leland Stanford Junior		
	38	2013/058907	WO	A1	2013-04-25	Good Start Genetics, Inc		
	39	2013/148496	WO	A1	2013-10-03	Univ Johns Hopkins		
	40	2013/177086	WO	A1	2013-11-28	Sequenom, Inc		
	41	2013/191775	WO	A2	2013-12-27	Nugen Technologies, Inc		
	42	2014/052909	WO	A2	2014-04-03	Childrens Mercy Hospital		
	43	2014/074246	WO	A1	2014-05-15	Good Start Genetics, Inc		
	44	2015/119941	WO	A2	2015-08-13	Igenomx Internat Genomics Corp		
	45	95/011995	WO	A1	1995-05-04	Affymax Tech Nv		

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	46	98/014275	WO	A1	1998-04-09	Intronn Llc		
	47	98/044151	WO	A1	1998-10-08	Glaxo Group Ltd		

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T5
	1	Abravaya, 1995, Detection of point mutations with a modified ligase chain reaction (Gap-LCR), Nucleic Acids Research, 23(4): 675-682	
	2	Adey, 2010, Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition, Genome Biol 11:R119	
	3	Ageno, 1969, The alkaline denaturation of DNA, Biophys J 9(11):1281-1311	
	4	Agrawal, 1990, Site-specific functionalization of oligodeoxynucleotides for non-radioactive labelling, Tetrahedron Let 31:1543-1546	
	5	Akhras, 2007, Connector inversion probe technology: A powerful one-primer multiplex DNA amplification system for numerous scientific applications, PLoSOne 9:e915	
	6	Akhras, 2007, PathogenMip Assay: a multiplex pathogen detection assay, PLOS One 2:e2230	
	7	Alazard, 2002, Sequencing of production-scale synthetic oligonucleotides by enriching for coupling failures using matrix-assisted laser desorption/ ionization time-of-flight mass spectrometry, Anal Biochem 301:57-64	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	8	Alazard, 2006, Sequencing oligonucleotides by enrichment of coupling failures using matrix-assisted laser desorption/ ionization time-of-flight mass spectrometry, Curr Protoc Nucleic Acid Chem, Chapter 10, Unit 10:1-7	
	9	Albert, 2007, Direct selection of human genomic loci by microarray hybridization, Nature Methods 4(11):903-5	
	10	Aljanabi, 1997, Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques, Nucl. Acids Res 25:4692-4693	
	11	Antonarakis and the Nomenclature Working Group, 1998, Recommendations for a nomenclature system for human gene mutations, Human Mutation 11:1-3	
	12	Archer, 2014, Selective and flexible depletion of problematic sequences from RNA-seq libraries at the cDNA stage, BMC Genomics 15(1):401	
	13	Ball, 2009, Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells, Nat Biotech 27:361-8	
	14	Balzer, 2013, Filtering duplicate reads from 454 pyrosequencing data, Bioinformatics 29(7):830-836	
	15	Barany, 1991, Genetic disease detection and DNA amplification using cloned thermostable ligase, PNAS 88:189-193	
	16	Barany, 1991, The Ligase Chain Reaction in a PCR World, Genome Research 1:5-16	
	17	Bau, 2008, Targeted next-generation sequencing by specific capture of multiple genomic loci using low-volume microfluidic DNA arrays, Analytical and Bioanal Chem 393(1):171-5	
	18	Beer, 1962, Determination of base sequence in nucleic acids with the electron microscope: visibility of a marker, PNAS 48(3):409-416	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	19	Bell, 2011, Carrier testing for severe childhood recessive diseases by next-generation sequencing, Sci Trans Med 3 (65ra4)	
	20	Benner, 2001, Evolution, language and analogy in functional genomics, Trends Genet 17:414-8	
	21	Bentzley, 1996, Oligonucleotide sequence and composition determined by matrix-assisted laser desorption/ionization, Anal Chem 68:2141-2146	
	22	Bentzley, 1998, Base specificity of oligonucleotide digestion by calf spleen phosphodiesterase with matrix-assisted laser desorption ionization analysis, Anal Biochem 258:31-37	
	23	Bhangale, 2006, Automating resequencing-based detection of insertion-deletion polymorphisms, Nature Genetics 38:1457-1462	
	24	Bickle, 1993, Biology of DNA Restriction, Microbiol Rev 57(2):434-50	
	25	Bonfield, 2013, Compression of FASTQ and SAM format sequencing data, PLoS One 8(3):e59190	
	26	Bose, 2012, BIND-An algorithm for loss-less compression of nucleotide sequence data, J Biosci 37(4):785-789	
	27	Boyden, 2013, High-throughput screening for SMN1 copy number loss by next-generation sequencing, American Society of Human Genetics 63rd Annual Meeting, Abstract, October 22, 2013	
	28	Boyer, 1971, DNA restriction and modification mechanisms in bacteria, Ann Rev Microbiol 25:153–76	
	29	Braasch, 2001, Locked nucleic acid (LNA): fine-tuning the recognition of DNA and RNA, Chemistry & Biology 8(1):1-7	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	30	Braslavsky, 2003, Sequence information can be obtained from single DNA molecules, PNAS 100:3960-4.	
	31	Brinkman, 2004, Splice Variants as Cancer Biomarkers, Clin Biochem 37:584	
	32	Brison, 1982, General method for cloning amplified DNA by differential screening, Mol Cell Biol 2(5):578-587	
	33	Brown, 1979, Chemical synthesis and cloning of a tyrosine tRNA gene, Methods Enzymol 68:109-51	
	34	Browne, 2002, Metal ion-catalyzed nucleic Acid alkylation and fragmentation, J Am Chem Soc 124(27):7950-7962	
	35	Brownstein, 2014, An international effort towards developing standards for best practices in analysis, interpretation and reporting of clinical genome sequencing results in the CLARITY Challenge, Genome Biol 15:R53	
	36	Bunyan, 2004, Dosage analysis of cancer predisposition genes by multiplex ligation-dependent probe amplification, British Journal of Cancer, 91(6):1155-59.	
	37	Burrow, 1994, A block-sorting lossless data compression algorithm, Technical Report 124, Digital Equipment Corporation, CA. (24 pages)	
	38	Carpenter, 2013, Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries, Am J Hum Genet 93(5):852-864	
	39	Caruthers, 1985, Gene synthesis machines: DNA chemistry and its uses, Science 230:281-285	
	40	Castellani, 2008, Consenses on the use of and interpretation of cystic fibrosis mutation analysis in clinical practice, J Cyst Fib 7:179-196	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	41	Challis, 2012, An integrative variant analysis suite for whole exome next-generation sequencing data, BMC Informatics 13(8):1-12	
	42	Chan, 2011, Natural and engineered nicking endonucleases-from cleavage mechanism to engineering of strand-specificity, Nucl Acids Res 39(1):1-18	
	43	Chen, 2010, Identification of racehorse and sample contamination by novel 24-plex STR system, Forensic Sci Int: Genetics 4:158-167	
	44	Chennagiri, 2013, A generalized scalable database model for storing and exploring genetic variations detected using sequencing data, American Society of Human Genetics 63rd Annual Meeting, Abstract, October 22, 2013	
	45	Chevreux, 1999, Genome sequence assembly using trace signals and additional sequence information, Proc GCB 99:45-56	
	46	Chirgwin, 1979, Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease, Biochemistry, 18:5294-99.	
	47	Choe, 2010, Novel CFTR Mutations in a Korean Infant with Cystic Fibrosis and Pancreatic Insufficiency, J Korean Med Sci 25:163-5.	
	48	Ciotti, 2004, Triplet repeat prmied PCR (TP PCR) in molecular diagnostic testing for Friedrich ataxia, J Mol Diag 6 (4):285-9	
	49	Cock, 2010, The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants, Nucleic Acids Res 38(6):1767-1771	
	50	Collins, 2004, Finishing the euchromatic sequence of the human genome, Nature 431(7011):931-45	

If you wish to add additional non-patent literature document citation information please click the Add button Add

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (**Not for submission under 37 CFR 1.99)**		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

EXAMINER SIGNATURE			
Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (**Not for submission under 37 CFR 1.99)**		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

See attached certification statement.

The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/Thomas C. Meyers/	Date (YYYY-MM-DD)	2021-01-15
Name/Print	Thomas C. Meyers	Registration Number	36,989

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

Doc code: IDS
Doc description: Information Disclosure Statement (IDS) Filed

PTO/SB/08a (02-18)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (**Not for submission under 37 CFR 1.99)**		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

U.S.PATENTS Remove

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Patent citation information please click the Add button. Add

U.S.PATENT APPLICATION PUBLICATIONS Remove

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	20120059594	A1	2012-03-08	Hatchwell et al.	
	2	20120074925	A1	2012-03-29	Oliver	
	3	20120079980	A1	2012-04-05	TAYLOR et al.	
	4	20120115736	A1	2012-05-10	Bjornson et al.	
	5	20120164630	A1	2012-06-28	Porreca et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	6	20120165202	A1	2012-06-28	Porreca et al.	
	7	20120179384	A1	2012-07-12	Kuramitsu et al.	
	8	20120214678	A1	2012-08-23	RAVA et al.	
	9	20120216151	A1	2012-08-23	Sarkar et al.	
	10	20120236861	A1	2012-09-20	Ganeshalingam et al.	
	11	20120245041	A1	2012-09-27	Brenner et al.	
	12	20120252020	A1	2012-10-04	Shuber	
	13	20120252684	A1	2012-10-04	Selifonov et al.	
	14	20120258461	A1	2012-10-11	Weisbart	
	15	20120270212	A1	2012-10-25	RABINOWITZ et al.	
	16	20120270739	A1	2012-10-25	RAVA et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	17	20130129755	A1	2013-05-23	Song	
	18	20130130921	A1	2013-05-23	Gao et al.	
	19	20130178378	A1	2013-07-11	HATCH et al.	
	20	20130183672	A1	2013-07-18	de Laat et al.	
	21	20130222388	A1	2013-08-29	McDonald	
	22	20130268206	A1	2013-10-10	Porreca et al.	
	23	20130268474	A1	2013-10-10	Nizzari et al.	
	24	20130274146	A1	2013-10-17	Umbarger et al.	
	25	20130275103	A1	2013-10-17	Struble et al.	
	26	20130288242	A1	2013-10-31	Stoughton et al.	
	27	20130323730	A1	2013-12-05	CURRY et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	28	20130332081	A1	2013-12-12	Reese et al.	
	29	20130337447	A1	2013-12-19	Porreca et al.	
	30	20130344096	A1	2013-12-26	Chiang et al.	
	31	20140129201	A1	2014-05-08	Kennedy et al.	
	32	20140136120	A1	2014-05-15	COLWELL et al.	
	33	20140206552	A1	2014-07-24	Rabinowitz et al.	
	34	20140222349	A1	2014-08-07	Higgins et al.	
	35	20140228226	A1	2014-08-14	Yin et al.	
	36	20140255931	A1	2014-09-11	Porreca et al.	
	37	20140308667	A1	2014-10-16	Umbarger	
	38	20140318274	A1	2014-10-30	Zimmerman et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	39	20140342354	A1	2014-11-20	Evans et al.	
	40	20140361022	A1	2014-12-11	Finneran	
	41	20150051085	A1	2015-02-19	Vogelstein et al.	
	42	20150056613	A1	2015-02-26	Kural	
	43	20150111208	A1	2015-04-23	Umbarger et al.	
	44	20150178445	A1	2015-06-25	CIBULSKIS et al.	
	45	20150258170	A1	2015-09-17	McCabe et al.	
	46	20150299767	A1	2015-10-22	Armour et al.	
	47	20150310163	A1	2015-10-29	Kingsmore et al.	
	48	20150354003	A1	2015-12-10	Umbarger	
	49	20160003812	A1	2016-01-07	Porreca et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	50	20160034638	A1	2016-02-04	Spence et al.	
	51	20160068889	A1	2016-03-10	Gole et al.	
	52	20160210486	A1	2016-07-21	Porreca et al.	
	53	20160251719	A1	2016-09-01	Umbarger	
	54	20170044610	A1	2017-02-16	Johnson	
	55	20170129964	A1	2017-05-11	CHEUNG	
	56	20170183731	A1	2017-06-29	Mann et al.	
	57	20170275676	A1	2017-09-28	Umbarger	
	58	20180371533	A1	2018-12-27	Gore et al.	
	59	20190233881	A1	2019-08-01	Umbarger et al.	
	60	20200181696	A1	2020-06-11	Porreca et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

If you wish to add additional U.S. Published Application citation information please click the Add button. Add

FOREIGN PATENT DOCUMENTS Remove

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1							

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]
	1	Craig, 1997, Removal of repetitive sequences from FISH probes, Hum Genet 100:472	
	2	Cremers, 1998, Autosomal Recessive Retinitis Pigmentosa and Cone-Rod Dystrophy Caused by Splice Site Mutations in the Stargardt's Disease Gene ABCR, Hum Mol Gen 7(3):355	
	3	Cronin, 1996, Cystic Fibrosis Mutation Detection by Hybridization to Light-Generated DNA Probe Arrays" Human Mutation 7:244	
	4	Dahl, 2005, Multiplex amplification enabled by selective circularization of large sets of genomic DNA fragments, Nucleic Acids Res 33(8):e71	
	5	Danecek, 2011, The variant call format and VCFtools, Bioinformatics 27(15):2156-2158	
	6	de la Bastide, 2007, Assembling genome DNA sequences with PHRAP, Current Protocols in Bioinformatics 17:11.4.1-11.4.15	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	7	Delcher, 1999, Alignment of whole genomes, Nuc Acids Res 27(11):2369-2376	
	8	den Dunnen, 2003, Mutation Nomenclature, Curr Prot Hum Genet 7.13.1-7.13.8	
	9	Deng et. al., 2012, Supplementary Material, Nature Biotechnology, S1-1 - S1-1 1, Retrieved from the Internet on 2012-10-24	
	10	Deng, 2009, targeted bisulfite sequencing reveals changes in DNA methylation, Nat Biotech 27(4):353-360	
	11	Deorowicz, 2013, Data compression for sequencing data, Alg for Mole Bio 8:25	
	12	Diep, 2012, Library-free methylation sequencing with bisulfite padlock probes, Nature Methods 9:270-272 (and supplemental information)	
	13	DiGuistini, 2009, De novo sequence assembly of a filamentous fungus using Sanger, 454 and Illumina sequence data, Genome Biology, 10:R94	
	14	Dolinsek, 2013, Depletion of unwanted nucleic acid templates by selection cleavage: LNAzymes, catalytically active oligonucleotides containing locked nucleic acids, open a new window for detecting rare microbial community members, App Env Microbiol 79(5):1534-1544	
	15	Dong, 2011, Mutation surveyor: An in silico tool for sequencing analysis, Methods Mol Biol 760:223-37	
	16	Drmanac, 1992, Sequencing by hybridization: towards an automated sequencing of one million M13 clones arrayed on membranes, Elctrophoresis 13:566-573	
	17	Dudley, 2009, A quick guide for developing effective bioinformatics programming skills, PLoS Comp Biol 5(12): e1000589	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	18	Ericsson, 2008, A dual-tag microarray platform for high-performance nucleic acid and protein analyses, Nucl Acids Res 36:e45	
	19	Fares, 2008, Carrier frequency of autosomal-recessive disorders in the Ashkenazi Jewish population: should the rationale for mutation choice for screening be reevaluated?, Prenatal Diagnosis 28:236-41.	
	20	Faulstich, 1997, A sequencing method for RNA oligonucleotides based on mass spectrometry, Anal Chem 69:4349-4353	
	21	Faust, 2014, SAMBLASTER: fast duplicate marking and structural variant read extraction, Bioinformatics published online May 7, 2014	
	22	Fitch, 1970, Distinguishing homologs from analogous proteins, Syst Biol 19(2):99-113	
	23	Flaschker, 2007, Description of the mutations in 15 subjects with variant forms of maple syrup urine disease, J Inherit Metab Dis 30:903-909	
	24	Frey, 2006, Statistics Hacks 108-115.	
	25	Friedenson, 2005, BRCA1 and BRCA2 Pathways and the Risk of Cancers Other Than Breast or Ovarian, Medscape General Medicine 7(2):60.	
	26	Furtado, 2011, Characterization of large genomic deletions in the FBN1 gene using multiplex ligation-dependent probe amplification, BMC Med Gen 12:119-125	
	27	Garber, 2008, Fixing the front end, Nat Biotech 26(10):1101-1104	
	28	Gemayel, 2010, Variable tandem repeats accelerate evolution of coding and regulatory sequences, Ann Rev Genet 44:445-77	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	29	Giusti, 1993, Synthesis and Characterization of f'-Fluorescent-dye-labeled Oligonucleotides, PCR Meth Appl 2:223-227	
	30	Glover, 1995, Sequencing of oligonucleotides using high performance liquid chromatography and electrospray mass spectrometry, Rapid Com Mass Spec 9:897-901	
	31	Gnirke, 2009, Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing, nature biotechnology 27:182-9.	
	32	Goto, 1994, A Study on Development of a Deductive Object-Oriented Database and Its Application to Genome Analysis, PhD Thesis, Kyushu University, Kyushu, Japan (106 pages)	
	33	Goto, 2010, BioRuby: bioinformatics software for the Ruby programming language, Bioinformatics 26(20):2617-2619	
	34	Green, 2005, Suicide polymerase endonuclease restriction, a novel technique for enhancing PCR amplification of minor DNA template, Appl Env Microbiol 71(8):4721-4727	
	35	Guerrero-Fernandez, 2013, FQbin: a compatible and optimize dformat for storing and managing sequence data, IWBBIO Proceedings, Granada 337-344	
	36	Gupta, 1991, A general method for the synthesis of 3'-sulfhydryl and phosphate group containing oligonucleotides, Nucl Acids Res 19(11):3019-3025	
	37	Gupta, 2014, Expanding the genetic toolkit: ZFNs, TALENs, and CRISPR-Cas9, J Clin Invest 124(10):4154	
	38	Gustincich, 1991, A fast method for high-quality genomic DNA extraction from whole human blood, BioTechniques 11 (3):298-302	
	39	Gut, 1995, A procedure for selective DNA alkylation and detection by mass spectrometry, Nucl Acids Res 23 (8):1367-1373	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	40	Hallam, 2014, Validation for Clinical Use of, and Initial Clinical Experience with, a Novel Approach to Population-Based Carrier Screening using High-Throughput Next-Generation DNA Sequencing, J Mol Diagn 16:180-9	
	41	Hammond, 1996, Extraction of DNA from preserved animal specimens for use in randomly amplified polymorphic DNA analysis, Anal Biochem 240:298-300	
	42	Hardenbol, 2003, Multiplexed genotyping with sequence-tagged molecular inversion probes, Nat Biotech 21:673-8.	
	43	Hardenbol, 2005, Highly multiplexed molecular inversion probe genotyping: over 10,000 targeted SNPs genotyped in a single tube assay, Genome Res 15:269-75	
	44	Harris, 2006, Defects can increase the melting temperature of DNA-nanoparticle assemblies, J Phys Chem B 110 (33):16393-6	
	45	Harris, 2008, Helicos True Single Molecule Sequencing (tSMS) Science 320:106-109	
	46	Harris, 2008, Single-molecule DNA sequencing of a viral genome, Science 320(5872):106-9	
	47	Heger, 2006, Protonation of Cresol Red in Acidic Aqueous Solutions Caused by Freezing, J Phys Chem B 110 (3):1277-1287	
	48	Heid, 1996, Real time quantitative PCR, Genome Res 6:986-994	
	49	Hiatt, 2013, Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation, Genome Res 23:843-54	
	50	Hodges, 2007, Genome-wide in situ exon capture for selective resequencing, Nat Genet 39(12):1522-7	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (**Not for submission under 37 CFR 1.99)**		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

If you wish to add additional non-patent literature document citation information please click the Add button	Add		
EXAMINER SIGNATURE			
Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (**Not for submission under 37 CFR 1.99)**		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

See attached certification statement.

The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/Thomas C. Meyers/	Date (YYYY-MM-DD)	2021-01-15
Name/Print	Thomas C. Meyers	Registration Number	36,989

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

Doc code: IDS
Doc description: Information Disclosure Statement (IDS) Filed

PTO/SB/08a (02-18)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (**Not for submission under 37 CFR 1.99)**		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

U.S.PATENTS Remove

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Patent citation information please click the Add button. Add

U.S.PATENT APPLICATION PUBLICATIONS Remove

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Published Application citation information please click the Add button. Add

FOREIGN PATENT DOCUMENTS Remove

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1							

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	1	Holland, 2008, BioJava: an open-source framework for bioinformatics, Bioinformatics 24(18):2096-2097	
	2	Homer, 2008, Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays. PLoS One 4(8):e1000167	
	3	Homer, 2009, BFAST: An alignment tool for large scale genome resequencing, PLoS ONE 4(11):e7767	
	4	Housley, 2009, SNP discovery and haplotype analysis in the segmentally duplicated DRD5 coding region, Ann Hum Genet 73(3):274-282	
	5	Huang, 2008, Comparative analysis of common CFTR polymorphisms poly-T, TGrepeats and M470V in a healthy Chinese population, World J Gastroenterol 14(12):1925-30	
	6	Husemann, 2009, Phylogenetic Comparative Assembly, Algorithms in Bioinformatics: 9th International Workshop, pp. 145-156, Salzberg & Warnow, Eds. Springer-Verlag, Berlin, Heidelberg	
	7	Illumina, 2010, De Novo assembly using Illumina reads, Technical Note (8 pages)	
	8	International Human Genome Sequencing Consortium, 2004, Finishing the euchromatic sequence of the human genome, Nature 431:931-945	
	9	Iqbal, 2012, De novo assembly and genotyping of variants using colored de Bruijn graphs, Nature Genetics 44:226-232	
	10	Isosomppi, 2009, Disease-causing mutations in the CLRN1 gene alter normal CLRN1 protien trafficking to the plasma membrane, Mol Vis 15:1806-1818	
	11	Jaijo, 2010, Microarray-based mutation analysis of 183 Spanish families with Usher syndrome, Invest Ophthalmol Vis Sci 51(3):1311-7	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	12	Jensen, 2001, Orthologs and paralogs—we need to get it right, Genome Biol 2(8):1002-1002.3	
	13	Jones, 2008, Core signaling pathways in human pancreatic cancers revealed by global genomic analyses, Science 321(5897):1801-1806	
	14	Kambara, 1988, Optimization of Parameters in a DNA Sequenator Using Fluorescence Detection, Nature Biotechnology 6:816-821	
	15	Kennedy, 2013, Accessing more human genetic variation with short sequencing reads, American Society of Human Genetics 63rd Annual Meeting, Abstract, October 22, 2013	
	16	Kent, 2002, BLAT--The BLAST-like alignment tool, Genome Res 12(4): 656-664.	
	17	Kerem, 1989, Identification of the cystic fibrosis gene: genetic analysis, Science 245:1073-1080	
	18	Kinde, 2012, FAST-SeqS: a simple an effective method for detection of aneuploidy by massively parallel sequencing, PLoS One 7(7):e41162	
	19	Kircher, 2010, High-througput DNA sequencing - concepts and limitations, Bioassays 32:524-36	
	20	Kirpekar, 1994, Matrix assisted laser desorption/ionization mass spectrometry of enzymatically synthesized RNA up to 150 kDa, Nucl Acids Res 22:3866-3870	
	21	Klein, 2011, LOCAS-A low coverage sequence assembly tool for re-sequencing projects, PLoS One 6(8):e23455	
	22	Kneen, 1998, Green fluorescent protein as a noninvasive intracellular pH indicator, Biophys J 74(3):1591-99	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	23	Koboldt, 2009, VarScan: variant detection in massively parallel sequencing of individual and pooled samples, Bioinformatics 25:2283-85	
	24	Krawitz, 2010, Microindel detection in short-read sequence data, Bioinformatics 26(6):722-729	
	25	Kreindler, 2010, Cystic fibrosis: exploiting its genetic basis in the hunt for new therapies, Pharmacol Ther 125 (2):219-229	
	26	Krishnakumar, 2008, A comprehensive assay for targeted multiplex amplification of human DNA sequences, PNAS 105:9296-301	
	27	Kumar, 2010, Comparing de novo assemblers for 454 transcriptome data, Genomics 11:571	
	28	Kurtz, 2004, Versatile and open software for comparing large genomes, Genome Biol 5:R12	
	29	Lam, 2008, Compressed indexing and local alignment of DNA, Bioinformatics 24(6):791-97	
	30	Langmead, 2009, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, Genome Biol 10:R25	
	31	Larkin, 2007, Clustal W and Clustal X version 2.0, Bioinformatics, 23(21):2947-2948	
	32	Lecompte, 2001, Multiple alignment of complete sequences (MACS) in the post-genomic era, Gene 270(1-2):17-30	
	33	Li, 2003, DNA binding and cleavage by the periplasmic nuclease Vvn: a novel structure with a known active site, EMBO J 22(15):4014-4025	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	34	Li, 2008, SOAP: short oligonucleotide alignment program, Bioinformatics 24(5):713-14	
	35	Li, 2009, Fast and accurate short read alignment with Burrows-Wheeler transform, Bioinformatics, 25 (14):1754-60	
	36	Li, 2009, SOAP2: an improved ultrafast tool for short read alignment, Bioinformatics 25(15):1966-67	
	37	Li, 2009, The Sequence Alignment/Map format and SAMtools, Bioinformatics 25(16):2078-9	
	38	Li, 2010, Fast and accurate long-read alignment with Burrows-Wheeler transform, Bioinformatics 26(5):589-95.	
	39	Li, 2011, Improving SNP discovery by base alignment quality, Bioinformatics 27:1157	
	40	Li, 2011, Single nucleotide polymorphism genotyping and point mutation detection by ligation on microarrays, J Nanosci Nanotechnol 11(2):994-1003	
	41	Li, 2012, A new approach to detecting low-level mutations in next-generation sequence data, Genome Biol 13:1-15	
	42	Li, 2014, HUGO: Hierarchical mUlti-reference Genome cOmpression for aligned reads, JAMIA 21:363-373	
	43	Lin, 2008, ZOOM! Zillions of Oligos Mapped, Bioinformatics, 24:2431	
	44	Lin, 2010, A molecular inversion prove assay for detecting alternative splicing, BMC Genomics 11(712):1-14	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	45	Lin, 2012, Development and evaluation of a reverse dot blot assay for the simultaneous detection of common alpha and beta thalassemia in Chinese, Blood Cells Molecules, and Diseases 48(2):86-90.	
	46	Lipman, 1985, Rapid and sensitive protein similarity searches, Science 227(4693):1435-41	
	47	Liu, 2012, Comparison of next-generation sequencing systems, J Biomed Biotech 2012:251364	
	48	Llopis, 1998, Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins, PNAS 95(12):6803-08	
	49	Ma, 2006, Application of real-time polymerase chain reaction (RT-PCR), J Am Soc 1-15	
	50	MacArthur, 2014, Guidelines for investigating causality of sequence variants in human disease, Nature 508:469-76	

If you wish to add additional non-patent literature document citation information please click the Add button Add

EXAMINER SIGNATURE

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (**Not for submission under 37 CFR 1.99)**		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

See attached certification statement.

The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/Thomas C. Meyers/	Date (YYYY-MM-DD)	2021-01-15
Name/Print	Thomas C. Meyers	Registration Number	36,989

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

Doc code: IDS
Doc description: Information Disclosure Statement (IDS) Filed

PTO/SB/08a (02-18)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (**Not for submission under 37 CFR 1.99)**		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

U.S.PATENTS Remove

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Patent citation information please click the Add button. Add

U.S.PATENT APPLICATION PUBLICATIONS Remove

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Published Application citation information please click the Add button. Add

FOREIGN PATENT DOCUMENTS Remove

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1							

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	1	Maddalena, 2005, Technical standards and guidelines: molecular genetic testing for ultra-rare disorders, Genet Med 7:571-83	
	2	Malewicz, 2010, Pregel: a system for large-scale graph processing, Proc. ACM SIGMOD Int Conf Mgmt Data 135-46	
	3	Mamanova, 2010, Target-enrichment strategies for next-generation sequencing, Nat Meth 7(2):111-118	
	4	Margulies, 2005, Genome sequencing in micro-fabricated high-density picotiter reactors, Nature, 437:376-380	
	5	Marras, 1999, Multiplex detection of single-nucleotide variations using molecular beacons, Genetic Analysis: Biomolecular Engineering 14:151	
	6	Maxam, 1977, A new method for sequencing DNA, PNAS, 74:560-564	
	7	May, 1988, How Many Species Are There on Earth?, Science 241(4872):1441-9	
	8	McDonnell, 2007, Antisepsis, disinfection, and sterilization: types, action, and resistance, page 239	
	9	McKenna, 2010, The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data, Genome Research 20:1297-1303	
	10	Messiaen, 1999, Exon 10b of the NF1 gene represents a mutational hotspot and harbors a recurrent missense mutation Y489C associated with aberrant splicing, Genetics in Medicine, 1(6):248-253	
	11	Meyer, 2007, Targeted high-throughput sequencing of tagged nucleic acid samples, Nucleic Acids Research 35(15): e97 (5 pages).	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	12	Meyer, 2008, Parallel tagged sequencing on the 454 platform, Nat Protocol 3(2):267-278	
	13	Miesenbock, 1998, Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins, Nature 394(6689):192-95	
	14	Miller, 2010, Assembly algorithms for next-generation sequencing data, Genomics 95:315-327	
	15	Mills, 2010, Mapping copy number variation by population-scale genome sequencing, Nature 470(7332):59-65	
	16	Miner, 2004, Molecular barcodes detect redundancy and contamination in hairpin-bisulfite PCR, Nucl Acids Res 32 (17):e135	
	17	Minton, 2011, Mutation Surveyor: software for DNA sequence analysis, Meth Mol Biol 688:143-53	
	18	Miyake, 2009, PIK3CA gene mutations and umplification in uterine cancers, Canc Lett 261:120-126	
	19	Miyazaki, 2009, Characterization of deletion breakpoints in patients with dystrophinopathy carrying a deletion of exons 45-55 of the Duchenne muscular dystrophy (DMD) gene, J Hum Gen 54:127-30	
	20	Mockler, 2005, Applications of DNA tiling arrays for whole-genome analysis, Genomics 85(1):1-15	
	21	Mohammed, 2012, DELIMINATE-a fast and efficient methods for loss-less compression of genomice sequences, Bioinformatics 28(19):2527-2529	
	22	Moudrianakis, 1965, Base Sequence Determination in Nucleic Acids with the Electron Microscope, III. Chemistry and Microscopy of Guanine-Labeled DNA, PNAS, 53:564-71	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	23	Mullan, 2002, Multiple sequence alignment-the gateway to further analysis, Brief Bioinform 3(3):303-5	
	24	Munne, 2012, Preimplantation genetic diagnosis for aneuploidy and translocations using array comparative genomic hybridization, Curr Genomics 13(6):463-470	
	25	Nan, 2006, A novel CFTR mutation found in a Chinese patient with cystic fibrosis, Chinese Med J 119(2):103-9	
	26	Narang, 1979, Improved phosphotriester method for the synthesis of gene fragments, Meth Enz 68:90-98	
	27	Nelson, 1989, Bifunctional oligonucleotide probes synthesized using a novel CPG support are able to detect single base pair mutations, Nucl Acids Res 17(18):7187-7194	
	28	Ng, 2009, Targeted capture and massively parallel sequencing of 12 human exomes, Nature 461(7261):272-6	
	29	Nicholas, 2002, Strategies for multiple sequence alignment, Biotechniques 32:572-91	
	30	Nickerson, 1990, Automated DNA diagnostics using an ELISA-based oligonucleotide ligation assay, PNAS 87:8923-7	
	31	Nielsen, 1999, Peptide Nucleic Acids, Protocols and Applications (Norfolk: Horizon Scientific Press, 1-19)	
	32	Nilsson, 2006, Analyzing genes using closing and replicating circles, Trends in Biotechnology 24:83-8	
	33	Ning, 2001, SSAHA: a fast search method for large DNA databases, Genome Res 11(10):1725-9	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	34	Nordhoff, 1993, Ion stability of nucleic acids in infrared matrix-assisted laser desorption/ ionization mass spectrometry, Nucl Acid Res 21(15):3347-57	
	35	Nuttle, 2013, Rapid and accurate large-scale genotyping of duplicated genes and discovery of interlocus gene conversions, Nat Meth 10(9):903-909	
	36	Nuttle, 2014, Resolving genomic disorder-associated breakpoints within segmental DNA duplications using massively parallel sequencing, Nat Prot 9(6):1496-1513	
	37	O'Roak, 2012, Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders, Science 338(6114):1619-1622	
	38	Oefner, 1996, Efficient random sub-cloning of DNA sheared in a recirculating point-sink flow system, Nucleic Acids Res 24(20):3879-3886	
	39	Oka, 2006, Detection of loss of heterozygosity in the p53 gene in renal cell carcinoma and bladder cancer using the polymerase chain reaction, Mol Carcinogenesis 4(1):10-13	
	40	Okoniewski, 2013, Precise breakpoint localization of large genomic deletions using PacBio and Illumina next-generation sequencers, Biotechniques 54(2):98-100	
	41	Okou, 2007, Microarray-based genomic selection for high-throughput reseuqencing, Nat Meth 4(11):907-909	
	42	Oliphant, 2002, BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping, Biotechniques Suppl:56-8, 60-1	
	43	Ordahl, 1976, Sheared DNA fragment sizing: comparison of techniques, Nucleic Acids Res 3:2985-2999	
	44	Ostrer, 2001, A genetic profile of contemporary Jewish populations, Nat Rev Genet 2(11):891-8	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	45	Owens, 1998, Aspects of oligonucleotide and peptide sequencing with MALDI and electrospray mass spectrometry, Bioorg Med Chem 6:1547-1554	
	46	Parameswaran, 2007, A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing, Nucl Acids Rese 35:e130	
	47	Parkinson, 2012, Preparation of high-quality next-generation sequencing libraries from picogram quantities of target DNA, Genome Res 22:125-133	
	48	Pastor, 2010, Conceptual modeling of human genome mutations: a dichotomy between what we have and what we shoudl have, 2010 Proc BIOSTEC Bioinformatics, pp: 160-166	
	49	Paton, 2000, Conceptual modelling of genomic information, Bioinformatics 16(6):548-57	
	50	Pearson, 1988, Improved tools for biological sequence comparison, PNAS 85(8):2444-8	

If you wish to add additional non-patent literature document citation information please click the Add button Add

EXAMINER SIGNATURE

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (**Not for submission under 37 CFR 1.99)**		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

See attached certification statement.

The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/Thomas C. Meyers/	Date (YYYY-MM-DD)	2021-01-15
Name/Print	Thomas C. Meyers	Registration Number	36,989

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

Doc code: IDS
Doc description: Information Disclosure Statement (IDS) Filed

PTO/SB/08a (02-18)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (**Not for submission under 37 CFR 1.99)**		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

U.S.PATENTS Remove

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Patent citation information please click the Add button. Add

U.S.PATENT APPLICATION PUBLICATIONS Remove

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Published Application citation information please click the Add button. Add

FOREIGN PATENT DOCUMENTS Remove

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1							

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	1	Pertea, 2003, TIGR gene indices clustering tools (TGICL), Bioinformatics 19(5):651-52	
	2	Pieles, 1993, Matrix-assisted laser desorption ionization time-of-flight mass spectrometry: A powerful tool for the mass and sequence analysis of natural and modified oligonucleotides, Nucleic Acids Res 21:3191-3196	
	3	Pinho, 2013, MFCompress: a compression tool for FASTA and multi-FASTA data, Bioinformatics 30(1):117-8	
	4	Porreca, 2007, Multiplex amplificaiton of large sets of human exons, Nat Meth 4(11):931-936	
	5	Porreca, 2013, Analytical performance of a Next-Generation DNA sequencing-based clinical workflow for genetic carrier screening, American Society of Human Genetics 63rd Annual Meeting, Abstract, October 22, 2013	
	6	Pourmand, 2006, PathgoenMIPer: a tool for the design of molecular inversion probes, BMC informatics 7:500	
	7	Procter, 2006, Molecular diagnosis of Prader-Willi and Angelman syndromes by methylation-specific melting analysis and methylation-specific multiplex ligation-dependent probe amplification, Clin Chem 52(7):1276-83	
	8	Qiagen, 2011, Gentra Puregene handbook, 3d Ed. (72 pages)	
	9	Quail, 2010, DNA: Mechanical Breakage, In Encyclopedia of Life Sciences, John Wiley & Sons Ltd, Chicester (5 pages)	
	10	Rambaut, 1997, Seq-Gen:an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees, Bioinformatics 13:235-38	
	11	Richards, 2008 ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions, Genet Med 10(4):294-300	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	12	Richter, 2008, MetaSim-A Sequencing Simulator for Genomics and Metagenomics, PLoS ONE 3:e3373	
	13	Roberts, 1980, Restriction and modification enzymes and their recognition sequences, Nucleic Acids Res 8(1):r63–r80	
	14	Rodriguez, 2010, Constructions from Dots and Lines, Bull Am Soc Inf Sci Tech 36(6):35-41	
	15	Rosendahl, 2013, CFTR, SPINK1, CTRC and PRSS1 variants in chronic pancreatitis: is the role of mutated CFTR overestimated?, Gut 62:582-592	
	16	Rothberg, 2011, An integrated semiconductor device enabling non-optical genome sequencing, Nature 475:348-352	
	17	Rowntree, 2003, The phenotypic consequences of CFTR mutations, Ann Hum Gen 67:471-485	
	18	Saihan, 2009, Update on Usher syndrome, Curr Op Neurology 22(1):19-24	
	19	Sanger, 1977, DNA Sequencing with chain-terminating inhibitors, PNAS 74(12):5463-5467	
	20	Santa Lucia, 1998, A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics, PNAS 95(4):1460-5	
	21	Sargent, 1987, Isolation of differentially expressed genes, Meth Enzym 152:423-432	
	22	Sauro, 2004, How Do You Calculate a Z-Score/ Sigma Level?, https://www.measuringusability.com/zcalc.htm (online publication)	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	23	Sauro, 2004, What's a Z-score and Why Use It in Usability Testing?, https://www.measuringusability.com/z.htm (online publication)	
	24	Schadt, 2010, A window into third-generation sequencing, Human Mol Genet 19(R2):R227-40	
	25	Schatz, 2010, Assembly of large genomes using second-generation sequencing, Genome Res., 20:1165-1173	
	26	Schiffman, 2009, Molecular inversion probes reveal patterns of 9p21 deletion and copy number aberrations in childhood leukemia, Cancer Genetics and Cytogenetics 193:9-18.	
	27	Schneeberger, 2011, Reference-guided assembly of four diverse Arabidopsis thaliana genomes, PNAS 108 (25):10249-10254	
	28	Schouten, 2002, Relative Quantification of 40 Nucleic Acid Sequences by Multiplex Ligation-Dependent Probe Amplification, Nucle Acids Res 30 (12):257	
	29	Schrijver, 2005, Diagnostic testing by CFTR gene mutation analysis in a large group of Hispanics, J Mol Diag 7 (2):289-299	
	30	Schuette, 1995, Sequence analysis of phosphorothioate oligonucleotides via matrix-assisted laser desorption ionization time-of-flight mass spectrometry, J Pharm Biomed Anal 13:1195-1203	
	31	Schwartz, 2009, Identification of cystic fibrosis variants by polymerase chain reaction/oligonucleotide ligation assay, J Mol Diag 11(3):211-15	
	32	Schwartz, 2011, Clinical utility of single nucleotide polymorphism arrays, Clin Lab Med 31(4):581-94	
	33	Sequeira, 1997, Implementing generic, object-oriented models in biology, Ecological Modeling 94.1:17-31	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	34	Shagin, 2002, A novel method for SNP detection, Genome Res 12:1935	
	35	Shen, 2011, High quality DNA sequence capture of 524 disease candidate genes, PNAS 108(16):6549-6554	
	36	Shen, 2013, Multiplex capture with double-stranded DNA probes, Genome Medicine 5(50):1-8	
	37	Sievers, 2011, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, Mol Syst Biol 7:539	
	38	Simpson, 2009, ABySS: A parallel assembler for short read sequence data, Genome Res., 19(6):1117-23	
	39	Slater, 2005, Automated generation of heuristics for biological sequence comparison, BMC Bioinformatics 6:31	
	40	Smirnov, 1996, Sequencing oligonucleotides by exonuclease digestion and delayed extraction matrix-assisted laser desorption ionization time-of-flight mass spectrometry, Anal Biochem 238:19-25	
	41	Smith, 1985, The synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis, Nucl Acid Res 13:2399-2412	
	42	Smith, 2010, Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples, Nucleic Acids Research 38(13):e142 (8 pages).	
	43	Soni, 2007, Progress toward ultrafast DNA sequencing using solid-state nanopores, Clin Chem 53(11):1996-2001	
	44	Spanu, 2010, Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism, Science 330(6010):1543-46	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	45	Sproat, 1987, The synthesis of protected 5'-mercapto-2',5'-dideoxyribonucleoside-3'-O-phosphoramidites; uses of 5'-mercapto-oligodeoxyribonucleotides, Nucl Acid Res 15:4837-4848	
	46	Streit, 2003, CFTR gene: molecular analysis in patients from South Brazil, Molecular Genetics and Metabolism 78:259-264	
	47	Strom, 2005, Mutation detection, interpretation, and applications in the clinical laboratory setting, Mutat Res 573:160-67	
	48	Summerer, 2009, Enabling technologies of genomic-scale sequence enrichment for targeted high-throughput sequencing, Genomics 94(6):363-8	
	49	Summerer, 2010, Targeted High Throughput Sequencing of a Cancer-Related Exome Subset by Specific Sequence Capture With a Fully Automated Microarray Platform, Genomics 95(4):241-246	
	50	Sunnucks, 1996, Microsatellite and chromosome evolution of parthenogenetic sitobion aphids in Australia, Genetics 144:747-756	

If you wish to add additional non-patent literature document citation information please click the Add button Add

EXAMINER SIGNATURE

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (**Not for submission under 37 CFR 1.99)**		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

See attached certification statement.

The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/Thomas C. Meyers/	Date (YYYY-MM-DD)	2021-01-15
Name/Print	Thomas C. Meyers	Registration Number	36,989

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

Doc code: IDS
Doc description: Information Disclosure Statement (IDS) Filed

PTO/SB/08a (02-18)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (**Not for submission under 37 CFR 1.99)**		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

U.S.PATENTS Remove

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Patent citation information please click the Add button. Add

U.S.PATENT APPLICATION PUBLICATIONS Remove

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Published Application citation information please click the Add button. Add

FOREIGN PATENT DOCUMENTS Remove

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1							

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	1	Tan, 2014, Clinical outcome of preimplantation genetic diagnosis and screening using next generation sequencing, GigaScience 3(30):1-9.	
	2	Thauvin-Robinet, 2009, The very low penetrance of cystic fibrosis for the R117H mutation: a reappraisal for genetic counseling and newborn screening, J Med Genet 46:752-758	
	3	Thiyagarajan, 2006, PathogenMIPer: a tool for the design of molecular inversion probes to detect multiple pathogens, BMC Bioinformatics 7:500	
	4	Thompson, 1994, Clustal W: improving the sensitivity of progressive mulitple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nuc Acids Res 22:4673-80	
	5	Thompson, 2011, The properties and applications of single-molecule DNA sequencing, Genome Biol 12(2):217	
	6	Thorstenson, 1998, An Automated Hydrodynamic Process for Controlled, Unbiased DNA Shearing, Genome Res 8(8): 848-855	
	7	Thorvaldsdottir, 2012, Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration, Brief Bioinform 24(2):178-92	
	8	Tkachuk, 1990, Detection of bcr-abl Fusion in Chronic Myelogeneous Leukemia by in Situ Hybridization, Science 250:559	
	9	Tobler, 2005, The SNPlex Genotyping System: A Flexible and Scalable Platform for SNP Genotyping, J Biomol Tech 16(4):398	
	10	Tokino, 1996, Characterization of the human p57 KIP2 gene: alternative splicing, insertion/deletion polymorphisms in VNTR sequences in the coding region, and mutational analysis, Human Genetics 96:625-31.	
	11	Treangen, 2011, Repetitive DNA and next-generation sequencing: computational challenges and solutions, Nat Rev Gen 13(1):36-46	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	12	Turner, 2009, Massively parallel exon capture and library-free resequencing across 16 genomes, Nat Meth 6:315-316	
	13	Turner, 2009, Methods for genomic partitioning, Ann Rev Hum Gen 10:263-284	
	14	Umbarger, 2013, Detecting contamination in Next Generation DNA sequencing libraries, American Society of Human Genetics 63rd Annual Meeting, Abstract, October 22, 2013	
	15	Umbarger, 2014, Next-generation carrier screening, Gen Med 16(2):132-140	
	16	Veeneman, 2012, Oculus: faster sequence alignment by streaming read compression, BMC Bioinformatics 13:297	
	17	Wahl, 1979, Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridization by using dextran sulfate, PNAS 76:3683-3687	
	18	Wallace, 1979, Hybridization of synthetic oligodeoxyribonucteotides to dp x 174DNA:the effect of single base pair mismatch, Nucl Acids Res 6:3543-3557	
	19	Wallace, 1987, Oligonucleotide probes for the screening of recombinant DNA libraries, Meth Enz 152:432-442	
	20	Wang, 2005, Allele quantification using molecular inversion probes (MIP), Nucleic Acids Res 33(21):e183	
	21	Warner, 1996, A general method for the detection of large CAG repeat expansions by fluorescent PCR, J Med Genet 33(12):1022-6	
	22	Warren, 2007, Assembling millions of short DNA sequences using SSAKE, Bioinformatics, 23:500-501	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	23	Waszak, 2010, Systematic inference of copy-number genotypes from personal genome sequencing data reveals extensive olfactory gene content diversity, PLoS Comp Biol 6(11):e1000988	
	24	Watson, 2004, Cystic fibrosis population carrier screening: 2004 revision of American College of Medical Genetics mutation panel, Genetics in Medicine 6(5):387-391	
	25	Williams, 2003, Restriction endonucleases classification, properties, and applications, Mol Biotechnol 23(3):225-43	
	26	Wirth, 1999, Quantitative analysis of survival motor neuron copies, Am J Hum Genet 64:1340-1356	
	27	Wittung, 1997, Extended DNA-Recognition Repertoire of Peptide Nucleic Acid (PNA): PNA-dsDNA Triplex Formed with Cytosine-Rich Homopyrimidine PNA, Biochemistry 36:7973	
	28	Wu, 1998, Sequencing regular and labeled oligonucleotides using enzymatic digestion and ionspray mass spectrometry, Anal Biochem 263:129-138	
	29	Wu, 2001, Improved oligonucleotide sequencing by alkaline phosphatase and exonuclease digestions with mass spectrometry, Anal Biochem 290:347-352	
	30	Xu, 2012, FastUniq: A fast de novo duplicates removal tool for paired short reads, PLoS One 7(12):e52249	
	31	Yau, 1996, Accurate diagnosis of carriers of deletions and duplications in Duchenne/Becker muscular dystrophy by fluorescent dosage analysis, J Med Gen 33(7):550-8	
	32	Ye, 2009, Pindel: a pattern growth approach to detect break points of large deletions and medium size insertions from paired-end short reads, Bioinformatics 25(21):2865-2871	
	33	Yershov, 1996, DNA analysis and diagnostics on oligonucleotide microchips, PNAS 93:4913-4918	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	34	Yoo, 2009, Applications of DNA microarray in disease diagnostics, J Microbiol Biotech19(7):635-46	
	35	Yoon, 2014, MicroDuMIP: target-enrichment technique for microarray-based duplex molecular inversion probes, Nucl Ac Res 43(5):e28	
	36	Yoshida, 2004, Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage, Cancer Science 95(11)866-71	
	37	Yu, 2007, A novel set of DNA methylation markers in urine sediments for sensitive/specific detection of bladder cancer, Clin Cancer Res 13(24):7296-7304	
	38	Yuan, 1981, Structure and mechanism of multifunctional restriction endonucleases, Ann Rev Biochem 50:285–319	
	39	Zerbino, 2008, Velvet: Algorithms for de novo short read assembly using de Bruijn graphs, Genome Research 18 (5):821-829	
	40	Zhang, 2011, Is Mitochondrial tRNAphe Variant m.593T.Ca Synergistically Pathogenic Mutation in Chinese LHON Families with m.11778G.A? PLoS ONE 6(10):e26511	
	41	Zhao, 2009, PGA4genomics for comparative genome assembly based on genetic algorithm optimization, Genomics 94 (4):284-6	
	42	Zheng, 2011, iAssembler: a package for de novo assembly of Roche-454/Sanger transcriptome sequences, BMC Bioinformatics 12:453	
	43	Zhou, 2014, Bias from removing read duplication in ultra-deep sequencing experiments, Bioinformatics 30 (8):1073-1080	
	44	Zhulidov, 2004, Simple cDNA normalization using kamchatka crab duplex-specific nuclease, Nucl Acids Res 32(3):e37	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (**Not for submission under 37 CFR 1.99)**		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	45	Zimmerman, 2010, A novel custom resequencing array for dilated cardiomyopathy, Gen Med 12(5):268-78	
	46	Zimran, 1990, A glucocerebrosidase fusion gene in Gaucher disease, J Clin Invest 85:219-222	
	47	Zuckerman, 1987, Efficient methods for attachment of thiol specific probes to the 3'-ends of synthetic oligodeoxyribonucleotides, Nucl Acid Res 15(13):5305-5321	

If you wish to add additional non-patent literature document citation information please click the Add button Add

EXAMINER SIGNATURE

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/952,764	11/19/2020	Gregory Porreca	MLB-002/06US 34468/40	7812

21710 7590 02/01/2021

BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

EXAMINER	
POHNERT, STEVEN C	

ART UNIT	PAPER NUMBER
1634	

NOTIFICATION DATE	DELIVERY MODE
02/01/2021	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

ip@brownrudnick.com
usactions@brownrudnick.com

Office Action Summary	**Application No.** 16/952,764	**Applicant(s)** Porreca et al.	
	Examiner STEVEN C POHNERT PhD	**Art Unit** 1634	**AIA (FITF) Status** No

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 2 MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☑ Responsive to communication(s) filed on 11/19/20.

☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2a)☐ This action is **FINAL.** 2b) ☐ This action is non-final.

3)☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

4)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims*

5) ☑ Claim(s) 68-76 is/are pending in the application.

5a) Of the above claim(s) _____ is/are withdrawn from consideration.

6) ☐ Claim(s) _____ is/are allowed.

7) ☐ Claim(s) _____ is/are rejected.

8) ☐ Claim(s) _____ is/are objected to.

9) ☑ Claim(s) 68-76 are subject to restriction and/or election requirement

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to **PPHfeedback@uspto.gov.**

Application Papers

10)☐ The specification is objected to by the Examiner.

11)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a)☐ All b)☐ Some** c)☐ None of the:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)

2) ☐ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b) Paper No(s)/Mail Date _____.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date _____.

4) ☐ Other: _____.

Petitioner Illumina, Inc., EX1004, p. 838

DETAILED ACTION

Notice of Pre-AIA or AIA Status

1. The present application is being examined under the pre-AIA first to invent provisions.

Election/Restriction

2. Restriction to one of the following inventions is required under 35 U.S.C. 121:

I. Claims 68-75, drawn to a method for correcting for errors or bias introduced during nucleic acid analysis workflow correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count , classified in C12Q1/6827.

II. Claim 76, drawn to a method for correcting for errors or bias introduced during nucleic acid analysis workflow requiring unique differentiator tags and correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule, classified in C12Q1/6883.

The inventions are independent or distinct, each from the other because:

3. Inventions I and II are directed to related methods. The related inventions are distinct if: (1) the inventions as claimed are either not capable of use together or can have a materially different design, mode of operation, function, or effect; (2) the inventions do not overlap in scope, i.e., are mutually exclusive; and (3) the inventions as claimed are not obvious variants. See MPEP § 806.05(j). In the instant case, the inventions as claimed methods requiring correcting for error or bias introduced during

said workflow by collapsing target:differentiator tag combinations observed more than once into a single count with generic differentiator tags have a materially different design and effect, are not of the same scope and are not obvious variants of methods requiring unique differentiator tags and correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule . Furthermore, the inventions as claimed do not encompass overlapping subject matter and there is nothing of record to show them to be obvious variants. Thus searching both would place a serious search burden on the office.

4. Restriction for examination purposes as indicated is proper because all the inventions listed in this action are independent or distinct for the reasons given above and there would be a serious search and/or examination burden if restriction were not required because one or more of the following reasons apply:

(a) the inventions have acquired a separate status in the art in view of their different classification;

(b) the inventions have acquired a separate status in the art due to their recognized divergent subject matter;

(c) the inventions require a different field of search (for example, searching different classes/subclasses or electronic resources, or employing different search queries);

(d) the prior art applicable to one invention would not likely be applicable to another invention;

(e) the inventions are likely to raise different non-prior art issues under 35 U.S.C. 101 and/or 35 U.S.C. 112, first paragraph.

Applicant is advised that the reply to this requirement to be complete must include (i) an election of an invention to be examined even though the requirement may be traversed (37 CFR 1.143) **and (ii) identification of the claims encompassing the elected invention**.

The election of an invention may be made with or without traverse. To reserve a right to petition, the election must be made with traverse. If the reply does not distinctly and specifically point out supposed errors in the restriction requirement, the election shall be treated as an election without traverse. Traversal must be presented at the time of election in order to be considered timely. Failure to timely traverse the requirement will result in the loss of right to petition under 37 CFR 1.144. If claims are added after the election, applicant must indicate which of these claims are readable upon the elected invention.

Should applicant traverse on the ground that the inventions are not patentably distinct, applicant should submit evidence or identify such evidence now of record showing the inventions to be obvious variants or clearly admit on the record that this is the case. In either instance, if the examiner finds one of the inventions unpatentable over the prior art, the evidence or admission may be used in a rejection under 35 U.S.C. 103 or pre-AIA 35 U.S.C. 103(a) of the other invention.

5. This application contains claims directed to the following patentably distinct species

Applicant must elect bias or errors.

Applicant must elect if the biological sample is blood or a biopsy.

Applicant must elect if the plurality of target nucleic acid molecules comprises bacterial nucleic acids, viral nuclei acids, circulating tumor nucleic acids or maternal circulating fetal nucleic acids or placental nucleic acids. This election must be consistent with the election of a biological sample.

Applicant must further elect determining copy number of a genomic region or a transcript.

. The species are independent or distinct because methods of correcting errors will not inherently provide art on methods of correcting a bias and vice versa. Further methods requiring a blood sample will not inherently provide art on methods requiring a biopsy. Further methods for analyzing a bacterial sequence will not inherently provide art on methods of analyzing viral nuclei acids, circulating tumor nucleic acids or maternal circulating fetal nucleic acids or placental nucleic acids, or vice versa. Further methods of detecting copy number of a genomic region (DNA) will not inherently provide art on methods of copy number of a transcript (RNA).In addition, these species are not obvious variants of each other based on the current record.

Applicant is required under 35 U.S.C. 121 to elect a single disclosed species, or a single grouping of patentably indistinct species, for prosecution on the merits to which the claims shall be restricted if no generic claim is finally held to be allowable.

There is a search and/or examination burden for the patentably distinct species as set forth above because at least the following reason(s) apply:

(a) the inventions have acquired a separate status in the art in view of their different classification;

(b) the inventions have acquired a separate status in the art due to their recognized divergent subject matter;

(c) the inventions require a different field of search (for example, searching different classes/subclasses or electronic resources, or employing different search queries);

(d) the prior art applicable to one invention would not likely be applicable to another invention;

(e) the inventions are likely to raise different non-prior art issues under 35 U.S.C. 101 and/or 35 U.S.C. 112, first paragraph.

Applicant is advised that the reply to this requirement to be complete must include (i) an election of a species to be examined even though the requirement may be traversed (37 CFR 1.143) **and (ii) identification of the claims encompassing the elected species or grouping of patentably indistinct species**, including any claims subsequently added. An argument that a claim is allowable or that all claims are generic is considered nonresponsive unless accompanied by an election.

The election may be made with or without traverse. To preserve a right to petition, the election must be made with traverse. If the reply does not distinctly and specifically point out supposed errors in the election of species requirement, the election shall be treated as an election without traverse. Traversal must be presented at the time of election in order to be considered timely. Failure to timely traverse the requirement will result in the loss of right to petition under 37 CFR 1.144. If claims are added after the election, applicant must indicate which of these claims are readable on the elected species or grouping of patentably indistinct species.

Should applicant traverse on the ground that the species, or groupings of patentably indistinct species from which election is required, are not patentably distinct, applicant should submit evidence or identify such evidence now of record showing them to be obvious variants or clearly admit on the record that this is the case. In either instance, if the examiner finds one of the species unpatentable over the prior art, the evidence or admission may be used in a rejection under 35 U.S.C. 103 or pre-AIA 35 U.S.C. 103(a) of the other species.

Upon the allowance of a generic claim, applicant will be entitled to consideration of claims to additional species which depend from or otherwise require all the limitations of an allowable generic claim as provided by 37 CFR 1.141.

6. Applicant is reminded that upon the cancellation of claims to a non-elected invention, the inventorship must be corrected in compliance with 37 CFR 1.48(a) if one or more of the currently named inventors is no longer an inventor of at least one claim remaining in the application. A request to correct inventorship under 37 CFR 1.48(a) must be accompanied by an application data sheet in accordance with 37 CFR 1.76 that identifies each inventor by his or her legal name and by the processing fee required under 37 CFR 1.17(i).

Any inquiry concerning this communication or earlier communications from the examiner should be directed to STEVEN C POHNERT PhD whose telephone number is (571)272-3803. The examiner can normally be reached on Monday- Friday about 6:00 AM-5:30 PM, every second Friday off.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Dave Nguyen can be reached on 571-272-0731. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see https://ppair-my.uspto.gov/pair/PrivatePair. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Steven Pohnert/
Primary Examiner, Art Unit 1634

Electronic Acknowledgement Receipt

EFS ID:	41805360
Application Number:	16952764
International Application Number:	
Confirmation Number:	7812
Title of Invention:	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS
First Named Inventor/Applicant Name:	Gregory Porreca
Customer Number:	21710
Filer:	Thomas C. Meyers/Casey Suda
Filer Authorized By:	Thomas C. Meyers
Attorney Docket Number:	MLB-002/06US 34468/40
Receipt Date:	01-FEB-2021
Filing Date:	19-NOV-2020
Time Stamp:	19:23:02
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1		MLB-002-06US-Response-Feb012021.pdf	129828 93a26bf48c2835d6dc9f26a1bf0d10dff65a9241	yes	8

	Multipart Description/PDF files in .zip description		
	Document Description	**Start**	**End**
	Response to Election / Restriction Filed	1	1
	Claims	2	4
	Applicant Arguments/Remarks Made in an Amendment	5	8
Warnings:			
Information:			
	Total Files Size (in bytes):	129828	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.
National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.
New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Attorney Docket No.: MLB-002/06US 34468/40

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANT:	Molecular Loop Biosolutions, LLC	ART UNIT:	1634
SERIAL NUMBER:	16/952,764	CONF. NO.:	7812
FILING DATE:	November 19, 2020	EXAMINER:	Steven C. Pohnert
TITLE:	**METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS**		

FILED ELECTRONICALLY

Commissioner for Patents
P. O. Box 1450
Alexandria, VA 22313-1450

Response

This paper is in response to the non-final Office Action mailed from the United States Patent and Trademark Office on February 1, 2021.

A **Listing of the Claims** begin on page 2.

Remarks begin on page 5.

Applicant believes that no fees are due with respect to this Response.

LISTING OF THE CLAIMS

1-67. (Canceled)

68. (Previously presented) A method for correcting for errors or bias introduced during nucleic acid analysis workflow, the method comprising the steps of:

obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin;

introducing a set of differentiator tags, wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag;

amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons;

sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and

correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count.

69. (Previously presented) The method of claim 68, wherein said correcting step further comprises collapsing all sequence reads comprising a same differentiator tag into a single read before determining a consensus base call for the target nucleic acid molecule sequences.

Petitioner Illumina, Inc., EX1004, p. 849

70. (Previously presented) The method of claim 69, further comprising the step of determining the presence of a sequencing or amplification error based upon a number of said differentiator tag sequences.

71. (Previously presented) The method of claim 68, wherein said biological sample is selected from the group consisting of blood and biopsy tissue.

72. (Previously presented) The method of claim 68, wherein said plurality of target nucleic acid molecules comprises bacterial or viral nucleic acids isolated from said biological sample.

73. (Previously presented) The method of claim 68, wherein said plurality of target nucleic acid molecules are circulating tumor nucleic acid molecules.

74. (Previously presented) The method of claim 68, further comprising the step of determining the copy number of a genomic region or transcript in a patient from whom the sample is obtained.

75. (Previously presented) The method of claim 68, wherein said plurality of target nucleic acid molecules are maternally-circulating fetal or placental nucleic acid molecules.

76. (Previously presented) A method for correcting for errors or bias introduced during nucleic acid analysis workflow, the method comprising the steps of:

obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin;

introducing a set of unique differentiator tags, wherein members of said set of unique differentiator tags are randomly associated with members of said plurality of target nucleic acid molecules, such that any given tag is uniquely associated with said locus of origin;

amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons;

sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and

correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule.

Remarks

Claims 68-76 are pending in the application.

Restriction/Election

The Office Action, at pages 2-4, required that applicant elect from the follow groups of inventions to be examined:

I. Claims 68-75, drawn to a method for correcting errors or bias introduced during nucleic acid analysis workflow correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once in a single count; and

II. Claim 76, drawn to a method for correcting errors or bias introduced during nucleic acid analysis workflow requiring unique differentiator tags and correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule.

In response, applicant provisionally elects, with traverse, group I for examination, encompassing claims 68-75.

In traversing the restriction of the claims, applicant submits that group I and group II, identified in the Office Action, encompass overlapping related subject matter, are capable of use together, and searching all of the claims would not place a serious search burden on the office.

Claims of an application may properly be required to be restricted to one of two or more claimed inventions only if they are able to support separate patents and they are either independent or distinct. *See MPEP §803*. Two or more inventions are related, and therefore not independent, if they are disclosed as connected in at least one of design, operation, or effect. *See*

MPEP §802.01. Two or more inventions are distinct if the inventions as claimed are not connected in at least one of design, operation, or effect "and wherein at least one invention is PATENTABLE (novel and nonobvious) OVER THE OTHER (though they may each be unpatentable over the prior art)." *Id. (emphasis in original)*. Even if the application includes claims to independent or distinct inventions, if the search and examination of all the claims in an application can be made without serious burden, the examiner must examine them on the merits. *See MPEP §803.*

Claims 68-75 and 76 are all directed to associating target nucleic acid molecules with differentiator tags prior to amplification of the nucleic acid molecules. As a result, even originally identical nucleic acid molecules from the plurality of molecules, for example in the case of genetic copy number, are made differentiable from one another as target:differentiator combinations. After amplification of the target:differentiator molecules, combinations observed more than once are understood to be a product of the amplification step, and not indicative of multiple copies of a genomic loci. This understanding can be used to correct for bias or error by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule (as recited in the claim of group II), for example by collapsing target:differentiator tag combinations observed more than once in a single count (as recited in the claims of group I).

The claims of group I and group II recite significantly overlapping related subject matter regarding error correction during nucleic acid analysis workflow. The methods recited in the claims of group I and group II are connected in their operation and effect and are capable of use together. Moreover, a search required for group I would be required for group II and can be made without serious burden.

Applicant disagrees with the requirement for restriction and requests withdrawal of the requirement.

Species election

The Office Action, at pages 4-6, required that applicant elect a single species from each of the following groups of species to be examined:

(i) bias or errors, encompassing claims 68-75;

(ii) blood or a biopsy as the biological sample;

(iii) bacterial nucleic acids, viral nucleic acids, circulating tumor nucleic acids, maternal circulating fetal nucleic acids, or placental nucleic acids as the target nucleic acid molecules;

(iv) copy number of a genomic region or copy number of a transcript.

In response, applicant elects, without traverse:

(i) bias, encompassing claims 68, 69, and 71-75;

(ii) blood as the biological sample, encompassing claims 68-75;

(iii) circulating tumor nucleic acids as the target nucleic acid molecules, encompassing claims 68-71 and 73;

(iv) copy number of a genomic region, encompassing claims 68-75.

If there are any questions regarding the above remarks, the Examiner is invited and encouraged to contact Applicant's representative.

Dated: February 1, 2021

Respectfully submitted,

BROWN RUDNICK LLP

BROWN RUDNICK LLP
One Financial Center
Boston, MA 02111
Tel: (617) 856-8483
Fax: (617) 856-8201

/Thomas C. Meyers/
Thomas C. Meyers, Reg. No. 36,989
Attorney for Applicant
Email: tmeyers@brownrudnick.com

63911551 v1

Petitioner Illumina, Inc., EX1004, p. 855

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/952,764	11/19/2020	Gregory Porreca	MLB-002/06US 34468/40	7812

21710 7590 02/10/2021

BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

EXAMINER	
POHNERT, STEVEN C	

ART UNIT	PAPER NUMBER
1634	

NOTIFICATION DATE	DELIVERY MODE
02/10/2021	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

ip@brownrudnick.com
usactions@brownrudnick.com

Office Action Summary	**Application No.** 16/952,764	**Applicant(s)** Porreca et al.	
	Examiner STEVEN C POHNERT PhD	**Art Unit** 1634	**AIA (FITF) Status** No

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☑ Responsive to communication(s) filed on 2/1/2021.

☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2a)☐ This action is **FINAL.** 2b) ☑ This action is non-final.

3)☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

4)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims*

5) ☑ Claim(s) 68-76 is/are pending in the application.

5a) Of the above claim(s) 70,72 and 75 is/are withdrawn from consideration.

6) ☐ Claim(s) _____ is/are allowed.

7) ☑ Claim(s) 68-69,71,73-74 and 76 is/are rejected.

8) ☐ Claim(s) _____ is/are objected to.

9) ☐ Claim(s) _____ are subject to restriction and/or election requirement

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to **PPHfeedback@uspto.gov.**

Application Papers

10)☑ The specification is objected to by the Examiner.

11)☑ The drawing(s) filed on 11/19/20 is/are: a)☑ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a)☐ All b)☐ Some** c)☐ None of the:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☑ Notice of References Cited (PTO-892)

2) ☑ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b) Paper No(s)/Mail Date _____.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date _____.

4) ☐ Other: _____.

Petitioner Illumina, Inc., EX1004, p. 857

DETAILED ACTION

Notice of Pre-AIA or AIA Status

1. The present application is being examined under the pre-AIA first to invent provisions.

Election/Restrictions

In view of arguments the restriction between groups I and II are withdrawn.

1. Applicant's election without traverse of bias, blood as the biological sample, circulating tumor cells and copy number of a genomic region in the reply filed on 2/1/2021 is acknowledged.

Claims 70, 72, 75 are withdrawn from further consideration pursuant to 37 CFR 1.142(b) as being drawn to a nonelected species, there being no allowable generic or linking claim. Election was made **without** traverse in the reply filed on 2/1/2021.

Claims 68-69,71, 73-74, 76 are being examined.

Priority

The instant application was filed 11/19/2020 is a continuation of 15231687, filed 08/08/2016 which is a continuation of 13266862, filed 03/13/2012 which is a national stage entry of PCT/US10/01293 having an international filing date: 04/30/2010 which claims priority from provisional application 61182089, filed 05/28/2009, claims priority from provisional application 61179358, filed 05/18/2009 claims priority from provisional application 61178923, filed 05/15/2009 and claims priority from provisional application 61174470, filed 04/30/2009.

Applicant's claim for the benefit of a prior-filed application under 35 U.S.C. 119(e) or under 35 U.S.C. 120, 121, 365(c), or 386(c) is acknowledged. Applicant has not

complied with one or more conditions for receiving the benefit of an earlier filing date under 35 U.S.C. 35 U.S.C. 119 (e)120, 121, 365(c), or 386(c) as follows: The later-filed application must be an application for a patent for an invention which is also disclosed in the prior application (the parent or original nonprovisional application or provisional application). The disclosure of the invention in the parent application and in the later-filed application must be sufficient to comply with the requirements of 35 U.S.C. 112(a) or the first paragraph of pre-AIA 35 U.S.C. 112, except for the best mode requirement. See *Transco Products, Inc. v. Performance Contracting, Inc.*, 38 F.3d 551, 32 USPQ2d 1077 (Fed. Cir. 1994)

The disclosure of the prior-filed application, Application No. 61182089, 61179358, 61178923, 61174470, fails to provide adequate support or enablement in the manner provided by 35 U.S.C. 112(a) or pre-AIA 35 U.S.C. 112, first paragraph for one or more claims of this application.

The instant claims were submitted as a preliminary amendment on 11/19/20. The instant claims require, "introducing a set of differentiator tags, wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag; amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons; sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count."

Provisional application 61/179358 is the only application that discloses collapsing sequencing reads. However it requires each target loci has a different differentiator tags. Review of provisional applications, 61182089, 61178923, 61174470 did not reveal support for the collapsing or introducing step. Thus the claims are being afforded priority to the instant filing date.

Information Disclosure Statement

2. The information disclosure statement (IDS) submitted on 1/15/21 is being considered by the examiner.

Specification

2. The specification is objected to as failing to provide proper antecedent basis for the claimed subject matter. See 37 CFR 1.75(d)(1) and MPEP § 608.01(o). Correction of the following is required: Claim 73 has been added by amendment and recites, "wherein said plurality of target nucleic acid molecules are circulating tumor nucleic acid molecules." Review and searching of the specification did not reveal antecedent basis for this amendment.

Claim Rejections - 35 USC § 112

3. The following is a quotation of the first paragraph of 35 U.S.C. 112(a):

> (a) IN GENERAL.—The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor or joint inventor of carrying out the invention.

The following is a quotation of the first paragraph of pre-AIA 35 U.S.C. 112:

> The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor of carrying out his invention.

4. Claims 68-69,71, 73-74, 76 rejected under 35 U.S.C. 112(a) or 35 U.S.C. 112 (pre-AIA), first paragraph, as failing to comply with the written description requirement. The claim(s) contains subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor or a joint inventor, or for applications subject to pre-AIA 35 U.S.C. 112, the inventor(s), at the time the application was filed, had possession of the claimed invention.

Claim 68 recites, "correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count." The specification recites collapsing three times. The specification in figure 5 demonstrates 4 bases with one base difference in sequences. It demonstrates in a very simple example that sequences can be preferentially amplified. The specification does not any guidance on what is required of collapsed, how collapsing is done or how collapsing corrects errors. How to determine to collapse sequences of greater complexity, or what is require of a target sequence to identify it to be collapsed. Thus the claims lack adequate written description.

Further claim 74 recites, "further comprising the step of determining the copy number of a genomic region or transcript in a patient from whom the sample is obtained.." The specification and claims provide no guidance on how to determine a copy number of genomic region or transcript based on collapsing reads of target: differentiator tag combinations. Thus the claims lack adequate written description.

Claim 76 recites, "correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from

the same input molecule.." The specification recites recognize six times. The specification does not recite, in the context of the limitations of claim 76.The specification in figure 5 demonstrates 4 bases with one base difference in sequences. It demonstrates in a very simple example that sequences can be preferentially amplified. The specification does not any guidance on what is required of recognizing, how recognizing is done or how recognizing corrects errors. How to determine to recognizing sequences of greater complexity, or what is require of a target sequence to identify it to be collapsed. Thus the claims lack adequate written description.

5. The following is a quotation of 35 U.S.C. 112(b):

> (b) CONCLUSION.—The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the inventor or a joint inventor regards as the invention.

The following is a quotation of 35 U.S.C. 112 (pre-AIA), second paragraph:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

6. Claims 68-69,71, 73-74, 76 rejected under 35 U.S.C. 112(b) or 35 U.S.C. 112 (pre-AIA), second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which the inventor or a joint inventor (or for applications subject to pre-AIA 35 U.S.C. 112, the applicant), regards as the invention.

Claim 68 recites, "correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count." The specification recites collapsing three times. The specification provides no guidance on how to perform collapsing, why collapsing is done or how collapsing corrects errors of bias. Thus the recitation is incomplete and unclear as it provides no guidance on how to perform collapsing, why collapsing is done or how collapsing corrects errors of bias.

Further claim 74 recites, "further comprising the step of determining the copy number of a genomic region or transcript in a patient from whom the sample is obtained.." The recitation is incomplete and unclear as the specification and claims provide no guidance on how to determine a copy number of genomic region or transcript based on collapsing reads of target: differentiator tag combinations.

Claim 76 recites, "correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule.." The specification recites recognize six times. The specification provides no guidance on how to perform recognizing, why recognizing is done or how recognizing corrects errors of bias. Thus the recitation is incomplete and unclear as it provides no guidance on how to perform recognizing, why recognizing is done or how recognizing corrects errors of bias.

Claim Rejections - 35 USC § 101

7. 35 U.S.C. 101 reads as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

8. Claims 68-69, 71, 73-74, 76 are rejected under 35 U.S.C. 101 because the claimed invention is directed to a mental step or abstract idea without significantly more. The claim(s) recite(s) the abstract idea or mental step of correcting. This judicial exception is not integrated into a practical application because if there is no steps which depend from or otherwise flow from the correcting step. The claim(s) does/do not

include additional elements that are sufficient to amount to significantly more than the judicial exception because if there are no additional steps integrating the correcting.

Claim analysis

The instant claim 68 is directed towards a method for correcting for errors or bias introduced during nucleic acid analysis workflow, the method comprising the steps of: obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin; introducing a set of differentiator tags, wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag; amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons; sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and correcting for error or bias introduced during said workflow. The correcting step is a mental step or abstract idea.

The instant claim 76 is directed towards method for correcting for errors or bias introduced during nucleic acid analysis workflow, the method comprising the steps of: obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin; introducing a set of unique differentiator tags, wherein members of said set of unique differentiator tags are randomly associated with members of said plurality of target nucleic acid molecules, such that any given tag is uniquely associated with said locus of origin; amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons; sequencing the amplicons

obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule.. The correcting step is a mental step or abstract idea.

The steps of obtaining, introducing, amplifying and sequencing are active steps.

Dependent claims set forth further limitations to about the reference level, method of detecting expression, subject and source of cancer.

According to the 2019 Patent Eligibility Guidance an initial two step analysis is required for determining statutory eligibility.

Step 1. Is the claim directed to a process, machine, manufacture, or composition of matter? In the instant case the Step 1 requirement is satisfied as the claims are directed towards a process.

Step 2A Prong one. Does the claim recite a law of nature, a natural phenomenon or an abstract idea? Yes, abstract idea and law.

With regards to claim 68, the claim recites, "and correcting for error or bias introduced during said workflow." This is an abstract idea or mental step.

With regards to claim 76, the claim recites, "correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule." This is an abstract idea or mental step.

Step 2A prong two. Does the claim recite additional elements that integrate the judicial exception into a practical application? The answer is no as the claims provide no active step in which the correcting is relied upon or otherwise integrated.

Step 2B. Does the claim recite additional elements that are significantly more than the judicial exceptions? No

Claims 68 and 76 require the steps of obtaining, introducing, amplifying and sequencing are active steps.. The specification teaches:

Sample preparative methods may fall into three classes: 1) single- or several-target amplification (e.g., uniplex PCR, 'multiplex' PCR), 2) multi-target hybridization enrichment (e.g., Agilent SureSelect 'hybrid capture' [Gnirke et al 2009, Nature methods 27:182-9], Roche/Nimblegen 'sequence capture' [Hodges et al 2007, Nature genetics 39:1522-7], and 3) multi-target circularization selection (e.g. molecular inversion probes or padlock probes, [Porreca et al 2007, Nature methods 4:931-6, Turner et al 2009, Nature methods 6:315-6], 'selectors' [Dahl et al 2005, Nucleic acids research 33:e71]). arrays, single nucleotide polymorphism (SNP) arrays (Affymetrix, Santa Clara, Calif.).

Claim Rejections - 35 USC § 103

3. In the event the determination of the status of the application as subject to AIA 35 U.S.C. 102 and 103 (or as subject to pre-AIA 35 U.S.C. 102 and 103) is incorrect, any correction of the statutory basis for the rejection will not be considered a new ground of rejection if the prior art relied upon, and the rationale supporting the rejection, would be the same under either status.

4. The following is a quotation of pre-AIA 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

5. Claim 68-69, 76 is/are rejected under pre-AIA 35 U.S.C. 103(a) as being unpatentable over Porecca(Nature methods (2007) volume 4, pages 931-936).

The specification provides no guidance on how, "correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count" is done thus this is considered broad.

Further the specification teaches, "As used herein, a "differentiator tag sequence" is a sequence of a nucleic acid (a preparative nucleic acid), which in the context of a plurality of different isolated nucleic acids, identifies a unique, independently isolated nucleic acid." Thus this broadly encompasses the targeting arms of a molecular inversion probe (MIP).

Porecca teaches figure 1:

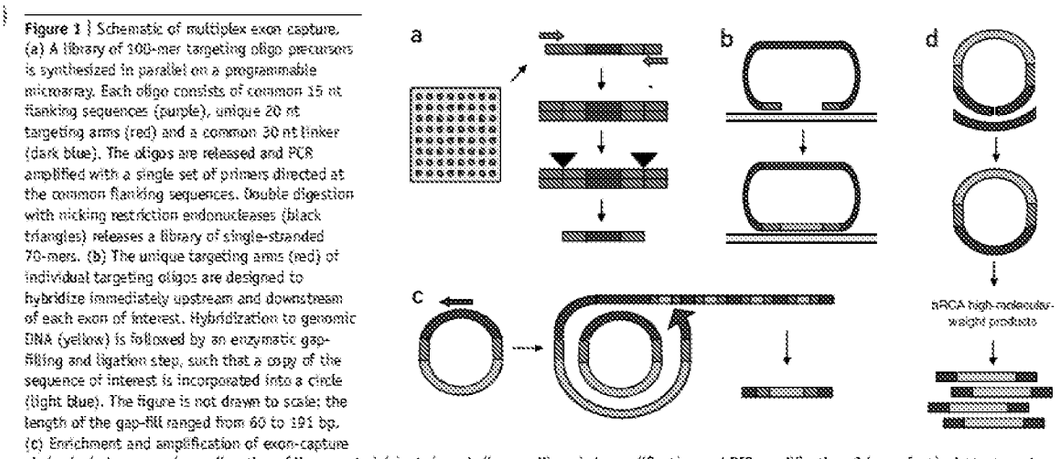


Figure 1 | Schematic of multiplex exon capture. (a) A library of 100-mer targeting oligo precursors is synthesized in parallel on a programmable microarray. Each oligo consists of common 15 nt flanking sequences (purple), unique 20 nt targeting arms (red) and a common 30 nt linker (dark blue). The oligos are released and PCR amplified with a single set of primers directed at the common flanking sequences. Double digestion with nicking restriction endonucleases (black triangles) releases a library of single-stranded 70-mers. (b) The unique targeting arms (red) of individual targeting oligos are designed to hybridize immediately upstream and downstream of each exon of interest. Hybridization to genomic DNA (yellow) is followed by an enzymatic gap-filling and ligation step, such that a copy of the sequence of interest is incorporated into a circle (light blue). The figure is not drawn to scale; the length of the gap-fill ranged from 60 to 191 bp. (c) Enrichment and amplification of exon-capture circles includes exonuclease digestion of linear material (not shown), linear rolling circle amplification and PCR amplification. Primers for the latter two steps are directed at the common linker sequence (dark blue). (d) To generate a shotgun sequencing library, amplicons are recircularized via their common linker (dark blue) and subjected to hyperbranched RCA with phi29 polymerase and random hexamers. High-molecular-weight products of the hRCA reaction are randomly sheared and appended with universal adaptors to generate the shotgun sequencing library.

On page 935, 2nd column Porrecca teaches genomic DNA from lyphoblastoid cell lines were hybridized to targeting oligos (MIPS via unique 20 nucleotide targeting arms), ligated and amplified and then sequenced.

Porecca does not specifically teach, "correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count"

However, Porecca teaches, "All evaluated positions at which the genotype is homozygous were dominated by the correct base-call (n ¼ 388). The data on heterozygous positions (n ¼ 90) are more ambiguous. For example, if we were to call positions with allele frequencies between 0.2 and 0.8 as heterozygous, then only 25 of 90 heterozygous positions (31%) would be called correctly. An alternative perspective is that because the goal of variation discovery is often to interrogate for rare variants, the key task is to identify positions that deviate from the reference genome with high sensitivity and specificity. If we call positions with observed allele frequencies of o0.8 as likely to be mutated relative to the reference genome, we can estimate a sensitivity of 63% for detecting rare mutations while maintaining high specificity (sensitivity, 57 of 90 heterozygous positions had a reference allele frequency of less than 0.8; specificity, 0 of 313 homozygous positions had a reference allele frequency of less than 0.8)." (page 934, 1st column)

Therefore it would have been prima facie obvious to one of ordinary skill in the art at the time the invention was made to sequence the arms of the MIPS and collapse (or identify) sequences including arms with greater than 0.8 allele frequency and identify they them as homozygous. The artisan would have a reasonable expectation of success as the artisan is merely using the guidance of Porreca to identify and collapse sequences that are homozygous. The artisan would have a reasonable expectation of success as the artisan is using techniques demonstrated by the art to work.

With regards to claim 69, Porecca teaches homozygous positions were dominated by the correct base call. Thus it is obvious to collapse all sequence reads before making a base call.

The specification provides no guidance on how, "correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule" is done thus this is considered broad.

Further the specification teaches, "As used herein, a "differentiator tag sequence" is a sequence of a nucleic acid (a preparative nucleic acid), which in the context of a plurality of different isolated nucleic acids, identifies a unique, independently isolated nucleic acid." Thus this broadly encompasses the targeting arms of a molecular inversion probe (MIP).

However, Porecca teaches, "All evaluated positions at which the genotype is homozygous were dominated by the correct base-call (n ¼ 388). The data on heterozygous positions (n ¼ 90) are more ambiguous. For example, if we were to call positions with allele frequencies between 0.2 and 0.8 as heterozygous, then only 25 of 90 heterozygous positions (31%) would be called correctly. An alternative perspective is that because the goal of variation discovery is often to interrogate for rare variants, the key task is to identify positions that deviate from the reference genome with high sensitivity and specificity. If we call positions with observed allele frequencies of o0.8 as likely to be mutated relative to the reference genome, we can estimate a sensitivity of 63% for detecting rare mutations while maintaining high specificity (sensitivity, 57 of 90 heterozygous positions had a reference allele frequency of less than 0.8; specificity, 0 of

313 homozygous positions had a reference allele frequency of less than 0.8)." (page 934, 1st column)

Therefore it would have been prima facie obvious to one of ordinary skill in the art at the time the invention was made to sequence the arms of the MIPS and recognize (or identify) sequences including arms with greater than 0.8 allele frequency and identify they them as homozygous. The artisan would have a reasonable expectation of success as the artisan is merely using the guidance of Porreca to recognize sequences with the same target sequences and MIP targeting arms as homozygous. The artisan would have a reasonable expectation of success as the artisan is using techniques demonstrated by the art to work.

6. Claim74 is rejected under pre-AIA 35 U.S.C. 103(a) as being unpatentable over Porecca(Nature methods (2007) volume 4, pages 931-936).as applied to claim 68-69, 76 above, and further in view of Wang (Genome Biology (2007) volume 8, R246.1-246.14).

The teachings of Porecca are set forth above and teach the detection of homozygous alleles by use of unique targeting arms of MIPS and recognizing or collapsing the combination of unique target arms and alleles (or target sequences).

Porecca does not specifically teach determination of copy number.

However, Wang teaches, "One of the techniques that have previously been described for the measurement of ACN is molecular inversion probes (MIPs) [10-12]. Briefly, MIP probes are circularizable oligonucleotides, where the two ends carry two sequences that are complementary to two sequences on the genome separated by one nucleotide (exactly where the variant to be genotyped is). After hybridization to the

genomic DNA, the reaction is split into four tubes where a single nucleotide is added to each tube. Upon the addition of the nucleotide, the MIP probe is ligated closed (but this only occurs in the tube with the nucleotide that is complementary to the allele on the genome), turning the probe into a circle. This structure can be selected for by the use of exonucleases, allowing for minimal cross talk' between probes and making it possible to obtain high quality data from highly multiplexed assays (>50,000- plex)" (page 1-2).

Wang teaches, "We describe in this manuscript significant improvements we have made to the MIP-based measurements of ACN. By increasing the proportion of genomic targets that are hybridized to the MIP probes, we have improved the performance while requiring a smaller amount of DNA. Additionally, for copy number measurements there are substantial advantages in uniformity and robustness when utilizing one-color readouts, especially at high levels of multiplexing" (page 7). Wang teaches, "These observations led us to suspect that it should be possible to genotype mixed DNA populations, such as occurs in tumor samples contaminated with normal tissue."

Therefore it would have been prima facie obvious to one of skill in the art at the time the invention was made to use MIPs to determine copy number. The artisan would be motivated as it would allow detection using less DNA and would allow genotype of mixed populations of cells to be determined. The artisan would have a reasonable expectation of success as the artisan is merely using MIPs as the art suggests.

7. Claims 71, 73 are rejected under pre-AIA 35 U.S.C. 103(a) as being unpatentable over Porecca(Nature methods (2007) volume 4, pages 931-936).as

applied to claim 68-69, 76 above, and further in view of Schiffman (Journal of Clinical Oncology (2007) volume 25, p530).

The teachings of Porecca are set forth above and teach the detection of homozygous alleles by use of unique targeting arms of MIPS and recognizing or collapsing the combination of unique target arms and alleles (or target sequences).

Porecca does not specifically teach determination of blood samples or circulating tumor nucleic acids.

However, Schiffman teaches the use of MIPs for analysis of blood samples subjects with leukemia (leukemia is a circulating cancer or tumor).

Therefore it would have been prima facie obvious to one of ordinary skill in the art at the time the invention was made to use blood samples from subjects with leukemia for examination by MIPs. The artisan would be motivated to examine gene copy number changes and LOH in blood samples to characterize the leukemia. The artisan would have a reasonable expectation of success as the artisan is using known samples in know methods.

Summary

No claims are allowed.

Conclusion

8. Any inquiry concerning this communication or earlier communications from the examiner should be directed to STEVEN C POHNERT PhD whose telephone number is (571)272-3803. The examiner can normally be reached on Monday- Friday about 6:00 AM-5:30 PM, every second Friday off.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Dave Nguyen can be reached on 571-272-0731. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see https://ppair-my.uspto.gov/pair/PrivatePair. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Steven Pohnert/
Primary Examiner, Art Unit 1634

Notice of References Cited	Application/Control No. 16/952,764	Applicant(s)/Patent Under Reexamination Porreca et al.	
	Examiner STEVEN C POHNERT PhD	Art Unit 1634	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	CPC Classification	US Classification
	A					
	B					
	C					
	D					
	E					
	F					
	G					
	H					
	I					
	J					
	K					
	L					
	M					

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	CPC Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	Porecca(Nature methods (2007) volume 4, pages 931-936).
	V	Wang (Genome Biology (2007) volume 8, R246.1-246.14)
	W	Schiffman (Journal of Clinical Oncology (2007) volume 25, p530).
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

Petitioner Illumina, Inc., EX1004, p. 874

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

See attached certification statement.

The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/Thomas C. Meyers/	Date (YYYY-MM-DD)	2021-01-15
Name/Print	Thomas C. Meyers	Registration Number	36,989

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

Doc code: IDS
Doc description: Information Disclosure Statement (IDS) Filed

PTO/SB/08a (02-18)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (**Not for submission under 37 CFR 1.99)**		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

U.S.PATENTS							Remove
Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	
	1						
If you wish to add additional U.S. Patent citation information please click the Add button.							Add

U.S.PATENT APPLICATION PUBLICATIONS							Remove
Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	
	1						
If you wish to add additional U.S. Published Application citation information please click the Add button.							Add

FOREIGN PATENT DOCUMENTS								Remove
Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1							
If you wish to add additional Foreign Patent Document citation information please click the Add button								Add

NON-PATENT LITERATURE DOCUMENTS			Remove
Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

/SP/	1	Tan, 2014, Clinical outcome of preimplantation genetic diagnosis and screening using next generation sequencing, GigaScience 3(30):1-9.	
/SP/	2	Thauvin-Robinet, 2009, The very low penetrance of cystic fibrosis for the R117H mutation: a reappraisal for genetic counseling and newborn screening, J Med Genet 46:752-758	
/SP/	3	Thiyagarajan, 2006, PathogenMIPer: a tool for the design of molecular inversion probes to detect multiple pathogens, BMC Bioinformatics 7:500	
/SP/	4	Thompson, 1994, Clustal W: improving the sensitivity of progressive mulitple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nuc Acids Res 22:4673-80	
/SP/	5	Thompson, 2011, The properties and applications of single-molecule DNA sequencing, Genome Biol 12(2):217	
/SP/	6	Thorstenson, 1998, An Automated Hydrodynamic Process for Controlled, Unbiased DNA Shearing, Genome Res 8(8): 848-855	
/SP/	7	Thorvaldsdottir, 2012, Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration, Brief Bioinform 24(2):178-92	
/SP/	8	Tkachuk, 1990, Detection of bcr-abl Fusion in Chronic Myelogeneous Leukemia by in Situ Hybridization, Science 250:559	
/SP/	9	Tobler, 2005, The SNPlex Genotyping System: A Flexible and Scalable Platform for SNP Genotyping, J Biomol Tech 16(4):398	
/SP/	10	Tokino, 1996, Characterization of the human p57 KIP2 gene: alternative splicing, insertion/deletion polymorphisms in VNTR sequences in the coding region, and mutational analysis, Human Genetics 96:625-31.	
/SP/	11	Treangen, 2011, Repetitive DNA and next-generation sequencing: computational challenges and solutions, Nat Rev Gen 13(1):36-46	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

/SP/	12	Turner, 2009, Massively parallel exon capture and library-free resequencing across 16 genomes, Nat Meth 6:315-316	
/SP/	13	Turner, 2009, Methods for genomic partitioning, Ann Rev Hum Gen 10:263-284	
/SP/	14	Umbarger, 2013, Detecting contamination in Next Generation DNA sequencing libraries, American Society of Human Genetics 63rd Annual Meeting, Abstract, October 22, 2013	
/SP/	15	Umbarger, 2014, Next-generation carrier screening, Gen Med 16(2):132-140	
/SP/	16	Veeneman, 2012, Oculus: faster sequence alignment by streaming read compression, BMC Bioinformatics 13:297	
/SP/	17	Wahl, 1979, Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridization by using dextran sulfate, PNAS 76:3683-3687	
/SP/	18	Wallace, 1979, Hybridization of synthetic oligodeoxyribonucteotides to dp x 174DNA:the effect of single base pair mismatch, Nucl Acids Res 6:3543-3557	
/SP/	19	Wallace, 1987, Oligonucleotide probes for the screening of recombinant DNA libraries, Meth Enz 152:432-442	
/SP/	20	Wang, 2005, Allele quantification using molecular inversion probes (MIP), Nucleic Acids Res 33(21):e183	
/SP/	21	Warner, 1996, A general method for the detection of large CAG repeat expansions by fluorescent PCR, J Med Genet 33(12):1022-6	
/SP/	22	Warren, 2007, Assembling millions of short DNA sequences using SSAKE, Bioinformatics, 23:500-501	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

/SP/	23	Waszak, 2010, Systematic inference of copy-number genotypes from personal genome sequencing data reveals extensive olfactory gene content diversity, PLoS Comp Biol 6(11):e1000988	
/SP/	24	Watson, 2004, Cystic fibrosis population carrier screening: 2004 revision of American College of Medical Genetics mutation panel, Genetics in Medicine 6(5):387-391	
/SP/	25	Williams, 2003, Restriction endonucleases classification, properties, and applications, Mol Biotechnol 23(3):225-43	
/SP/	26	Wirth, 1999, Quantitative analysis of survival motor neuron copies, Am J Hum Genet 64:1340-1356	
/SP/	27	Wittung, 1997, Extended DNA-Recognition Repertoire of Peptide Nucleic Acid (PNA): PNA-dsDNA Triplex Formed with Cytosine-Rich Homopyrimidine PNA, Biochemistry 36:7973	
/SP/	28	Wu, 1998, Sequencing regular and labeled oligonucleotides using enzymatic digestion and ionspray mass spectrometry, Anal Biochem 263:129-138	
/SP/	29	Wu, 2001, Improved oligonucleotide sequencing by alkaline phosphatase and exonuclease digestions with mass spectrometry, Anal Biochem 290:347-352	
/SP/	30	Xu, 2012, FastUniq: A fast de novo duplicates removal tool for paired short reads, PLoS One 7(12):e52249	
/SP/	31	Yau, 1996, Accurate diagnosis of carriers of deletions and duplications in Duchenne/Becker muscular dystrophy by fluorescent dosage analysis, J Med Gen 33(7):550-8	
/SP/	32	Ye, 2009, Pindel: a pattern growth approach to detect break points of large deletions and medium size insertions from paired-end short reads, Bioinformatics 25(21):2865-2871	
/SP/	33	Yershov, 1996, DNA analysis and diagnostics on oligonucleotide microchips, PNAS 93:4913-4918	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

/SP/	34	Yoo, 2009, Applications of DNA microarray in disease diagnostics, J Microbiol Biotech19(7):635-46	
/SP/	35	Yoon, 2014, MicroDuMIP: target-enrichment technique for microarray-based duplex molecular inversion probes, Nucl Ac Res 43(5):e28	
/SP/	36	Yoshida, 2004, Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage, Cancer Science 95(11)866-71	
/SP/	37	Yu, 2007, A novel set of DNA methylation markers in urine sediments for sensitive/specific detection of bladder cancer, Clin Cancer Res 13(24):7296-7304	
/SP/	38	Yuan, 1981, Structure and mechanism of multifunctional restriction endonucleases, Ann Rev Biochem 50:285–319	
/SP/	39	Zerbino, 2008, Velvet: Algorithms for de novo short read assembly using de Bruijn graphs, Genome Research 18 (5):821-829	
/SP/	40	Zhang, 2011, Is Mitochondrial tRNAphe Variant m.593T.Ca Synergistically Pathogenic Mutation in Chinese LHON Families with m.11778G.A? PLoS ONE 6(10):e26511	
/SP/	41	Zhao, 2009, PGA4genomics for comparative genome assembly based on genetic algorithm optimization, Genomics 94 (4):284-6	
/SP/	42	Zheng, 2011, iAssembler: a package for de novo assembly of Roche-454/Sanger transcriptome sequences, BMC Bioinformatics 12:453	
/SP/	43	Zhou, 2014, Bias from removing read duplication in ultra-deep sequencing experiments, Bioinformatics 30 (8):1073-1080	
/SP/	44	Zhulidov, 2004, Simple cDNA normalization using kamchatka crab duplex-specific nuclease, Nucl Acids Res 32(3):e37	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

/SP/	45	Zimmerman, 2010, A novel custom resequencing array for dilated cardiomyopathy, Gen Med 12(5):268-78	
/SP/	46	Zimran, 1990, A glucocerebrosidase fusion gene in Gaucher disease, J Clin Invest 85:219-222	
/SP/	47	Zuckerman, 1987, Efficient methods for attachment of thiol specific probes to the 3'-ends of synthetic oligodeoxyribonucleotides, Nucl Acid Res 15(13):5305-5321	

If you wish to add additional non-patent literature document citation information please click the Add button Add

EXAMINER SIGNATURE

Examiner Signature	/Steven Pohnert/ (02/04/2021)	Date Considered	02/04/2021

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

See attached certification statement.

The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/Thomas C. Meyers/	Date (YYYY-MM-DD)	2021-01-15
Name/Print	Thomas C. Meyers	Registration Number	36,989

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

Doc code: IDS
Doc description: Information Disclosure Statement (IDS) Filed

PTO/SB/08a (02-18)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (**Not for submission under 37 CFR 1.99)**		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

U.S.PATENTS Remove

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Patent citation information please click the Add button. Add

U.S.PATENT APPLICATION PUBLICATIONS Remove

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	20120059594	A1	2012-03-08	Hatchwell et al.	
	2	20120074925	A1	2012-03-29	Oliver	
	3	20120079980	A1	2012-04-05	TAYLOR et al.	
	4	20120115736	A1	2012-05-10	Bjornson et al.	
	5	20120164630	A1	2012-06-28	Porreca et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	6	20120165202	A1	2012-06-28	Porreca et al.	
	7	20120179384	A1	2012-07-12	Kuramitsu et al.	
	8	20120214678	A1	2012-08-23	RAVA et al.	
	9	20120216151	A1	2012-08-23	Sarkar et al.	
	10	20120236861	A1	2012-09-20	Ganeshalingam et al.	
	11	20120245041	A1	2012-09-27	Brenner et al.	
	12	20120252020	A1	2012-10-04	Shuber	
	13	20120252684	A1	2012-10-04	Selifonov et al.	
	14	20120258461	A1	2012-10-11	Weisbart	
	15	20120270212	A1	2012-10-25	RABINOWITZ et al.	
	16	20120270739	A1	2012-10-25	RAVA et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	17	20130129755	A1	2013-05-23	Song	
	18	20130130921	A1	2013-05-23	Gao et al.	
	19	20130178378	A1	2013-07-11	HATCH et al.	
	20	20130183672	A1	2013-07-18	de Laat et al.	
	21	20130222388	A1	2013-08-29	McDonald	
	22	20130268206	A1	2013-10-10	Porreca et al.	
	23	20130268474	A1	2013-10-10	Nizzari et al.	
	24	20130274146	A1	2013-10-17	Umbarger et al.	
	25	20130275103	A1	2013-10-17	Struble et al.	
	26	20130288242	A1	2013-10-31	Stoughton et al.	
	27	20130323730	A1	2013-12-05	CURRY et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	28	20130332081	A1	2013-12-12	Reese et al.	
	29	20130337447	A1	2013-12-19	Porreca et al.	
	30	20130344096	A1	2013-12-26	Chiang et al.	
	31	20140129201	A1	2014-05-08	Kennedy et al.	
	32	20140136120	A1	2014-05-15	COLWELL et al.	
	33	20140206552	A1	2014-07-24	Rabinowitz et al.	
	34	20140222349	A1	2014-08-07	Higgins et al.	
	35	20140228226	A1	2014-08-14	Yin et al.	
	36	20140255931	A1	2014-09-11	Porreca et al.	
	37	20140308667	A1	2014-10-16	Umbarger	
	38	20140318274	A1	2014-10-30	Zimmerman et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	39	20140342354	A1	2014-11-20	Evans et al.	
	40	20140361022	A1	2014-12-11	Finneran	
	41	20150051085	A1	2015-02-19	Vogelstein et al.	
	42	20150056613	A1	2015-02-26	Kural	
	43	20150111208	A1	2015-04-23	Umbarger et al.	
	44	20150178445	A1	2015-06-25	CIBULSKIS et al.	
	45	20150258170	A1	2015-09-17	McCabe et al.	
	46	20150299767	A1	2015-10-22	Armour et al.	
	47	20150310163	A1	2015-10-29	Kingsmore et al.	
	48	20150354003	A1	2015-12-10	Umbarger	
	49	20160003812	A1	2016-01-07	Porreca et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	50	20160034638	A1	2016-02-04	Spence et al.	
	51	20160068889	A1	2016-03-10	Gole et al.	
	52	20160210486	A1	2016-07-21	Porreca et al.	
	53	20160251719	A1	2016-09-01	Umbarger	
	54	20170044610	A1	2017-02-16	Johnson	
	55	20170129964	A1	2017-05-11	CHEUNG	
	56	20170183731	A1	2017-06-29	Mann et al.	
	57	20170275676	A1	2017-09-28	Umbarger	
	58	20180371533	A1	2018-12-27	Gore et al.	
	59	20190233881	A1	2019-08-01	Umbarger et al.	
	60	20200181696	A1	2020-06-11	Porreca et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

If you wish to add additional U.S. Published Application citation information please click the Add button. Add

FOREIGN PATENT DOCUMENTS Remove

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1							

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]
	1	Craig, 1997, Removal of repetitive sequences from FISH probes, Hum Genet 100:472	
	2	Cremers, 1998, Autosomal Recessive Retinitis Pigmentosa and Cone-Rod Dystrophy Caused by Splice Site Mutations in the Stargardt's Disease Gene ABCR, Hum Mol Gen 7(3):355	
	3	Cronin, 1996, Cystic Fibrosis Mutation Detection by Hybridization to Light-Generated DNA Probe Arrays" Human Mutation 7:244	
	4	Dahl, 2005, Multiplex amplification enabled by selective circularization of large sets of genomic DNA fragments, Nucleic Acids Res 33(8):e71	
	5	Danecek, 2011, The variant call format and VCFtools, Bioinformatics 27(15):2156-2158	
	6	de la Bastide, 2007, Assembling genome DNA sequences with PHRAP, Current Protocols in Bioinformatics 17:11.4.1-11.4.15	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	7	Delcher, 1999, Alignment of whole genomes, Nuc Acids Res 27(11):2369-2376	
	8	den Dunnen, 2003, Mutation Nomenclature, Curr Prot Hum Genet 7.13.1-7.13.8	
	9	Deng et. al., 2012, Supplementary Material, Nature Biotechnology, S1-1 - S1-1 1, Retrieved from the Internet on 2012-10-24	
	10	Deng, 2009, targeted bisulfite sequencing reveals changes in DNA methylation, Nat Biotech 27(4):353-360	
	11	Deorowicz, 2013, Data compression for sequencing data, Alg for Mole Bio 8:25	
	12	Diep, 2012, Library-free methylation sequencing with bisulfite padlock probes, Nature Methods 9:270-272 (and supplemental information)	
	13	DiGuistini, 2009, De novo sequence assembly of a filamentous fungus using Sanger, 454 and Illumina sequence data, Genome Biology, 10:R94	
	14	Dolinsek, 2013, Depletion of unwanted nucleic acid templates by selection cleavage: LNAzymes, catalytically active oligonucleotides containing locked nucleic acids, open a new window for detecting rare microbial community members, App Env Microbiol 79(5):1534-1544	
	15	Dong, 2011, Mutation surveyor: An in silico tool for sequencing analysis, Methods Mol Biol 760:223-37	
	16	Drmanac, 1992, Sequencing by hybridization: towards an automated sequencing of one million M13 clones arrayed on membranes, Elctrophoresis 13:566-573	
	17	Dudley, 2009, A quick guide for developing effective bioinformatics programming skills, PLoS Comp Biol 5(12): e1000589	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	18	Ericsson, 2008, A dual-tag microarray platform for high-performance nucleic acid and protein analyses, Nucl Acids Res 36:e45	
	19	Fares, 2008, Carrier frequency of autosomal-recessive disorders in the Ashkenazi Jewish population: should the rationale for mutation choice for screening be reevaluated?, Prenatal Diagnosis 28:236-41.	
	20	Faulstich, 1997, A sequencing method for RNA oligonucleotides based on mass spectrometry, Anal Chem 69:4349-4353	
	21	Faust, 2014, SAMBLASTER: fast duplicate marking and structural variant read extraction, Bioinformatics published online May 7, 2014	
	22	Fitch, 1970, Distinguishing homologs from analogous proteins, Syst Biol 19(2):99-113	
	23	Flaschker, 2007, Description of the mutations in 15 subjects with variant forms of maple syrup urine disease, J Inherit Metab Dis 30:903-909	
	24	Frey, 2006, Statistics Hacks 108-115.	
	25	Friedenson, 2005, BRCA1 and BRCA2 Pathways and the Risk of Cancers Other Than Breast or Ovarian, Medscape General Medicine 7(2):60.	
	26	Furtado, 2011, Characterization of large genomic deletions in the FBN1 gene using multiplex ligation-dependent probe amplification, BMC Med Gen 12:119-125	
	27	Garber, 2008, Fixing the front end, Nat Biotech 26(10):1101-1104	
	28	Gemayel, 2010, Variable tandem repeats accelerate evolution of coding and regulatory sequences, Ann Rev Genet 44:445-77	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	29	Giusti, 1993, Synthesis and Characterization of f'-Fluorescent-dye-labeled Oligonucleotides, PCR Meth Appl 2:223-227	
	30	Glover, 1995, Sequencing of oligonucleotides using high performance liquid chromatography and electrospray mass spectrometry, Rapid Com Mass Spec 9:897-901	
	31	Gnirke, 2009, Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing, nature biotechnology 27:182-9.	
	32	Goto, 1994, A Study on Development of a Deductive Object-Oriented Database and Its Application to Genome Analysis, PhD Thesis, Kyushu University, Kyushu, Japan (106 pages)	
	33	Goto, 2010, BioRuby: bioinformatics software for the Ruby programming language, Bioinformatics 26(20):2617-2619	
	34	Green, 2005, Suicide polymerase endonuclease restriction, a novel technique for enhancing PCR amplification of minor DNA template, Appl Env Microbiol 71(8):4721-4727	
	35	Guerrero-Fernandez, 2013, FQbin: a compatible and optimize dformat for storing and managing sequence data, IWBBIO Proceedings, Granada 337-344	
	36	Gupta, 1991, A general method for the synthesis of 3'-sulfhydryl and phosphate group containing oligonucleotides, Nucl Acids Res 19(11):3019-3025	
	37	Gupta, 2014, Expanding the genetic toolkit: ZFNs, TALENs, and CRISPR-Cas9, J Clin Invest 124(10):4154	
	38	Gustincich, 1991, A fast method for high-quality genomic DNA extraction from whole human blood, BioTechniques 11 (3):298-302	
	39	Gut, 1995, A procedure for selective DNA alkylation and detection by mass spectrometry, Nucl Acids Res 23 (8):1367-1373	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	40	Hallam, 2014, Validation for Clinical Use of, and Initial Clinical Experience with, a Novel Approach to Population-Based Carrier Screening using High-Throughput Next-Generation DNA Sequencing, J Mol Diagn 16:180-9	
	41	Hammond, 1996, Extraction of DNA from preserved animal specimens for use in randomly amplified polymorphic DNA analysis, Anal Biochem 240:298-300	
	42	Hardenbol, 2003, Multiplexed genotyping with sequence-tagged molecular inversion probes, Nat Biotech 21:673-8.	
	43	Hardenbol, 2005, Highly multiplexed molecular inversion probe genotyping: over 10,000 targeted SNPs genotyped in a single tube assay, Genome Res 15:269-75	
	44	Harris, 2006, Defects can increase the melting temperature of DNA-nanoparticle assemblies, J Phys Chem B 110 (33):16393-6	
	45	Harris, 2008, Helicos True Single Molecule Sequencing (tSMS) Science 320:106-109	
	46	Harris, 2008, Single-molecule DNA sequencing of a viral genome, Science 320(5872):106-9	
	47	Heger, 2006, Protonation of Cresol Red in Acidic Aqueous Solutions Caused by Freezing, J Phys Chem B 110 (3):1277-1287	
	48	Heid, 1996, Real time quantitative PCR, Genome Res 6:986-994	
	49	Hiatt, 2013, Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation, Genome Res 23:843-54	
	50	Hodges, 2007, Genome-wide in situ exon capture for selective resequencing, Nat Genet 39(12):1522-7	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

If you wish to add additional non-patent literature document citation information please click the Add button Add

EXAMINER SIGNATURE

Examiner Signature	/Steven Pohnert/ (02/04/2021)	Date Considered	02/04/2021

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

See attached certification statement.

The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/Thomas C. Meyers/	Date (YYYY-MM-DD)	2021-01-15
Name/Print	Thomas C. Meyers	Registration Number	36,989

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

Doc code: IDS
Doc description: Information Disclosure Statement (IDS) Filed

PTO/SB/08a (02-18)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (**Not for submission under 37 CFR 1.99)**		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

U.S.PATENTS						Remove
Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	4683195		1987-07-28	Mullis et al.	
	2	4683202		1987-07-28	Mullis	
	3	4988617		1991-01-29	Landegren et al.	
	4	5060980		1991-10-29	Johnson et al.	
	5	5210015		1993-05-11	Gelfand et al.	
	6	5234809		1993-08-10	Boom et al.	
	7	5242794		1993-09-07	Whiteley et al.	
	8	5348853		1994-09-20	Wang et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	9	5434049	A	1995-07-18	Okano et al.	
	10	5459307		1995-10-17	Klotz, Jr.	
	11	5486686		1996-01-23	Zdybel, Jr. et al.	
	12	5491224	A	1996-02-13	Bittner et al.	
	13	5494810		1996-02-27	Barany et al.	
	14	5567583		1996-10-22	Wang et al.	
	15	5583024		1996-12-10	McElroy, deceased et al.	
	16	5604097		1997-02-18	Brenner	
	17	5636400		1997-06-10	Young	
	18	5674713		1997-10-07	McElroy, deceased et al.	
	19	5695934		1997-12-09	Brenner	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	20	5700673		1997-12-23	McElroy, deceased et al.	
	21	5701256		1997-12-23	Marr et al.	
	22	5830064		1998-11-03	Bradish et al.	
	23	5846719		1998-12-08	Brenner et al.	
	24	5863722		1999-01-26	Brenner	
	25	5866337		1999-02-02	Schon	
	26	5869252		1999-02-09	Bouma et al.	
	27	5869717		1999-02-09	Frame et al.	
	28	5871921		1999-02-16	Landegren et al.	
	29	5888788		1999-03-30	De Miniac	
	30	5942391		1999-08-24	Zhang et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	31	5971921		1999-10-26	Timbel	
	32	5993611		1999-11-30	Moroney, III et al.	
	33	5994056		1999-11-30	Higuchi	
	34	6020127	A	2000-02-01	MacKenzie et al.	
	35	6033854		2000-03-07	Kumit et al.	
	36	6033872		2000-03-07	Bergsma et al.	
	37	6100099		2000-08-08	Gordon et al.	
	38	6138077		2000-10-24	Brenner	
	39	6150516		2000-11-21	Brenner et al.	
	40	6171785		2001-01-09	Higuchi	
	41	6172214		2001-01-09	Brenner	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	42	6172218		2001-01-09	Brenner	
	43	6197508		2001-03-06	Stanley	
	44	6197574		2001-03-06	Miyamoto et al.	
	45	6210891		2001-04-03	Nyren et al.	
	46	6223128	B1	2001-04-24	Allex et al.	
	47	6235472		2001-05-22	Landegren et al.	
	48	6235475		2001-05-22	Brenner et al.	
	49	6235501		2001-05-22	Gautsch et al.	
	50	6235502		2001-05-22	Weissman et al.	
	51	6258568		2001-07-10	Nyren	
	52	6274320		2001-08-14	Rothberg et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	53	6306597		2001-10-23	Macevicz	
	54	6352828		2002-03-05	Brenner	
	55	6360235		2002-03-19	Tilt et al.	
	56	6361940		2002-03-26	Van Ness et al.	
	57	6403320		2002-06-11	Read et al.	
	58	6462254		2002-10-08	Vernachio et al.	
	59	6489105		2002-12-03	Matlashewski et al.	
	60	6558928		2003-05-06	Landegren	
	61	6569920		2003-05-27	Wen et al.	
	62	6582938		2003-06-24	Su et al.	
	63	6585938		2003-07-01	Machida et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	64	6613516		2003-09-02	Christians et al.	
	65	6714874		2004-03-30	Myers et al.	
	66	6716580		2004-04-06	Gold et al.	
	67	6719449		2004-04-13	Laugharn, Jr. et al.	
	68	6818395	B1	2004-11-16	Quake et al.	
	69	6828100		2004-12-07	Ronaghi	
	70	6833246		2004-12-21	Balasubramanian	
	71	6858412		2005-02-22	Willis et al.	
	72	6911345		2005-06-28	Quake et al.	
	73	6913879		2005-07-05	TeleChem International Inc.	
	74	6927024		2005-08-09	Dodge et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	64	6613516		2003-09-02	Christians et al.	
	65	6714874		2004-03-30	Myers et al.	
	66	6716580		2004-04-06	Gold et al.	
	67	6719449		2004-04-13	Laugharn, Jr. et al.	
	68	6818395	B1	2004-11-16	Quake et al.	
	69	6828100		2004-12-07	Ronaghi	
	70	6833246		2004-12-21	Balasubramanian	
	71	6858412		2005-02-22	Willis et al.	
	72	6911345		2005-06-28	Quake et al.	
Change(s) applied to document, /A.C./ 10/5/2023	73	6913879		2005-07-05	~~TeleChem International Inc~~. Schena	
	74	6927024		2005-08-09	Dodge et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	75	6941317		2005-09-06	Chamberlin et al.	
	76	6948843		2005-09-27	Laugharn, Jr. et al.	
	77	7034143		2006-04-25	Preparata et al.	
	78	7041481		2006-05-09	Anderson et al.	
	79	7049077		2006-05-23	Yang	
	80	7057026		2006-06-06	Barnes et al.	
	81	7071324		2006-07-04	Preparata et al.	
	82	7074564		2006-07-11	Landegren	
	83	7074586		2006-07-11	Cheronis et al.	
	84	7115400		2006-10-03	Adessi et al.	
	85	7169560		2007-01-30	Lapidus et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	86	7211390		2007-05-01	Rothberg et al.	
	87	7232656		2007-06-19	Balasubramanian et al.	
	88	7244559		2007-07-17	Rothberg et al.	
	89	7264929		2007-09-04	Rothberg et al.	
	90	7282337		2007-10-16	Harris	
	91	7297518		2007-11-20	Quake et al.	
	92	7320860		2008-01-22	Landegren et al.	
	93	7323305		2008-01-29	Leamon et al.	
	94	7335762		2008-02-26	Rothberg et al.	
	95	7351528		2008-04-01	Landegren	
	96	7393665		2008-07-01	Brenner	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	97	7510829		2009-03-31	Faham et al.	
	98	7523117		2009-04-21	Zhang et al.	
	99	7537889		2009-05-26	Sinha et al.	
	100	7537897		2009-05-26	Brenner et al.	
	101	7544473		2009-06-09	Brenner	
	102	7582431		2009-09-01	Drmanac et al.	
	103	7598035		2009-10-06	Macevicz	
	104	7629151		2009-12-08	Gold et al.	
	105	7642056		2010-01-05	Ahn et al.	
	106	7666593		2010-02-23	Lapidus	
	107	7700323		2010-04-20	Willis et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	108	7774962		2010-08-17	Ladd	
	109	7776616		2010-08-17	Heath et al.	
	110	7790388		2010-09-07	Landegren et al.	
	111	7809509		2010-10-05	Milosavljevic	
	112	7835871		2010-11-16	Kain et al.	
	113	7862999		2011-01-04	Zheng et al.	
	114	7865534		2011-01-04	Genstruct	
	115	7883849		2011-02-08	Dahl	
	116	7957913		2011-06-07	Chinitz et al.	
	117	7960120		2011-06-14	Rigatti et al.	
	118	7985716		2011-07-26	Yershov et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	119	7993880		2011-08-09	Willis et al.	
	120	8024128		2011-09-20	Rabinowitz et al.	
	121	8114027		2012-02-14	Triva	
	122	8165821		2012-04-24	Zhang	
	123	8209130		2012-06-26	Kennedy et al.	
	124	8283116		2012-10-09	Bhattacharyya et al.	
	125	8462161		2013-06-11	Barber	
	126	8463895		2013-06-11	Arora et al.	
	127	8474228		2013-07-02	Adair et al.	
	128	8496166		2013-07-30	Burns et al.	
	129	8529744		2013-09-10	Marziali et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	130	8738300		2014-05-27	Porreca et al.	
	131	8778609		2014-07-15	Umbarger	
	132	8812422		2014-08-19	Nizzari et al.	
	133	8847799		2014-09-30	Kennedy et al.	
	134	8976049		2015-03-10	Kennedy et al.	
	135	9074244		2015-07-07	Sparks et al.	
	136	9115387		2015-08-25	Umbarger	
	137	9228233		2016-01-05	Kennedy et al.	
	138	9292527		2016-03-22	Kennedy et al.	
	139	9535920		2017-01-03	Kennedy et al.	
	140	9567639		2017-02-14	Oliphant et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	141	9677124		2017-06-13	Umbarger	
	142	10066259	B2	2018-09-04	Gore et al.	
	143	10202637		2019-02-12	Umbarger	
	144	10227635		2019-03-12	Umbarger et al.	
	145	10604799		2020-03-31	Porreca et al.	
	146	10683533		2020-06-16	Umbarger et al.	
	147	RE39793		2007-08-21	Brenner	
	148	RE41780		2010-09-28	Anderson et al.	

If you wish to add additional U.S. Patent citation information please click the Add button. Add

U.S.PATENT APPLICATION PUBLICATIONS Remove

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	20010007742	A1	2001-07-12	LANDERGREN	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	2	20010046673	A1	2001-11-29	French et al.	
	3	20020001800	A1	2002-01-03	LAPIDUS	
	4	20020040216	A1	2002-04-04	Dumont et al.	
	5	20020042052	A1	2002-04-11	NILSEN et al.	
	6	20020091666	A1	2002-07-11	Rice et al.	
	7	20020164629	A1	2002-11-07	Quake et al.	
	8	20020172954	A1	2002-11-21	Mao et al.	
	9	20020182609	A1	2002-12-05	Arcot	
	10	20020187496	A1	2002-12-12	Andersson et al.	
	11	20020190663	A1	2002-12-19	Rasmussen	
	12	20030166057	A1	2003-09-04	Hildebrand et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	13	20030175709	A1	2003-09-18	Murphy et al.	
	14	20030177105	A1	2003-09-18	Xiao et al.	
	15	20030203370	A1	2003-10-30	Yakhini et al.	
	16	20030208454	A1	2003-11-06	Rienhoff et al.	
	17	20030224384	A1	2003-12-04	Sayood et al.	
	18	20040029264	A1	2004-02-12	Robbins	
	19	20040053275	A1	2004-03-18	Shafer	
	20	20040106112	A1	2004-06-03	Nilsson et al.	
	21	20040121373	A1	2004-06-24	Friedlander et al.	
	22	20040142325	A1	2004-07-22	Mintz et al.	
	23	20040152108	A1	2004-08-05	Keith et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	24	20040161773	A1	2004-08-19	Rogan et al.	
	25	20040170965	A1	2004-09-02	Scholl et al.	
	26	20040171051	A1	2004-09-02	Holloway	
	27	20040197813	A1	2004-10-07	Hoffman et al.	
	28	20040209299	A1	2004-10-21	Pinter et al.	
	29	20050003369	A1	2005-01-06	Christians et al.	
	30	20050026204	A1	2005-02-03	Landegren	
	31	20050032095	A1	2005-02-10	Wigler et al.	
	32	20050048505	A1	2005-03-03	Fredrick et al.	
	33	20050059048	A1	2005-03-17	Gunderson et al.	
	34	20050100900	A1	2005-05-12	Kawashima et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	35	20050112590	A1	2005-05-26	Boom et al.	
	36	20050186589	A1	2005-08-25	Kowalik et al.	
	37	20050214811	A1	2005-09-29	Margulies et al.	
	38	20050244879	A1	2005-11-03	Schumm et al.	
	39	20050250147	A1	2005-11-10	Macevicz	
	40	20050272065	A1	2005-12-08	Lakey et al.	
	41	20060008824	A1	2006-01-12	Ronaghi et al.	
	42	20060019304	A1	2006-01-26	Hardenbol et al.	
	43	20060024681	A1	2006-02-02	Smith et al.	
	44	20060030536	A1	2006-02-09	Yu et al.	
	45	20060078894	A1	2006-04-13	Winkler et al.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	46	20060149047	A1	2006-07-06	Nanduri et al.	
	47	20060177837	A1	2006-08-10	Borozan et al.	
	48	20060183132	A1	2006-08-17	Fu et al.	
	49	20060192047	A1	2006-08-31	Goossen	
	50	20060195269	A1	2006-08-31	Yeatman et al.	
	51	20060246500	A1	2006-11-02	Browne	
	52	20060263789	A1	2006-11-23	Kincaid	
	53	20060281098	A1	2006-12-14	Miao et al.	
	54	20060286577	A1	2006-12-21	Jia	
	55	20060292585	A1	2006-12-28	Nautiyal et al.	
	56	20060292611	A1	2006-12-28	Berka et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	57	20070009925	A1	2007-01-11	Fang et al.	
	58	20070020640	A1	2007-01-25	McCloskey et al.	
	59	20070042369	A1	2007-02-22	Reese et al.	
	60	20070092883	A1	2007-04-26	Schouten et al.	
	61	20070114362	A1	2007-05-24	Feng et al.	
	62	20070128624	A1	2007-06-07	Gormley et al.	
	63	20070161013	A1	2007-07-12	Hantash	
	64	20070162983	A1	2007-07-12	Hesterkamp et al.	
	65	20070166705	A1	2007-07-19	Milton et al.	
	66	20070212704	A1	2007-09-13	Dong et al.	
	67	20070225487	A1	2007-09-27	Nilsson et al.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	68	20070238122	A1	2007-10-11	Allbritton et al.	
	69	20070244675	A1	2007-10-18	Shai et al.	
	70	20070264653	A1	2007-11-15	Berlin et al.	
	71	20080003142	A1	2008-01-03	Link et al.	
	72	20080014589	A1	2008-01-17	Link et al.	
	73	20080076118	A1	2008-03-27	Tooke et al.	
	74	20080081330	A1	2008-04-03	Kahvejian	
	75	20080085836	A1	2008-04-10	Kearns et al.	
	76	20080090239	A1	2008-04-17	SHOEMAKER et al.	
	77	20080125324	A1	2008-05-29	Petersdorf et al.	
	78	20080176209	A1	2008-07-24	Muller et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	79	20080269068	A1	2008-10-30	CHURCH et al.	
	80	20080280955	A1	2008-11-13	McCamish	
	81	20080293589	A1	2008-11-27	Shapero	
	82	20090009904	A1	2009-01-08	Yasuna et al.	
	83	20090019156	A1	2009-01-15	Mo et al.	
	84	20090026082	A1	2009-01-29	Rothberg et al.	
	85	20090029385	A1	2009-01-29	Christians et al.	
	86	20090035777	A1	2009-02-05	Kokoris et al.	
	87	20090042206	A1	2009-02-12	Schneider et al.	
	88	20090098551	A1	2009-04-16	Landers et al.	
	89	20090099041	A1	2009-04-16	CHURCH et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	90	20090105081	A1	2009-04-23	Rodesch et al.	
	91	20090119313	A1	2009-05-07	Pearce	
	92	20090127589	A1	2009-05-21	Rothberg et al.	
	93	20090129647	A1	2009-05-21	Dimitrova et al.	
	94	20090156412	A1	2009-06-18	Boyce, IV et al.	
	95	20090163366	A1	2009-06-25	Nickerson et al.	
	96	20090181389	A1	2009-07-16	Li et al.	
	97	20090191565	A1	2009-07-30	Lapidus et al.	
	98	20090192047	A1	2009-07-30	Parr et al.	
	99	20090202984	A1	2009-08-13	Cantor	
	100	20090203014	A1	2009-08-13	WU et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	101	20090220955	A1	2009-09-03	Verrant	
	102	20090226975	A1	2009-09-10	Sabot et al.	
	103	20090233814	A1	2009-09-17	Bashkirov et al.	
	104	20090298064	A1	2009-12-03	Batzoglou et al.	
	105	20090301382	A1	2009-12-10	Patel	
	106	20090318310	A1	2009-12-24	Liu et al.	
	107	20100035243	A1	2010-02-11	Muller et al.	
	108	20100035252	A1	2010-02-11	Rothberg et al.	
	109	20100063742	A1	2010-03-11	Hart et al.	
	110	20100069263	A1	2010-03-18	Shendure et al.	
	111	20100076185	A1	2010-03-25	Adey et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	112	20100086914	A1	2010-04-08	BENTLEY et al.	
	113	20100086926	A1	2010-04-08	Craig et al.	
	114	20100105107	A1	2010-04-29	Hildebrand et al.	
	115	20100137143	A1	2010-06-03	Rothberg et al.	
	116	20100137163	A1	2010-06-03	Link et al.	
	117	20100143908	A1	2010-06-10	Gillevet	
	118	20100159440	A1	2010-06-24	Messier et al.	
	119	20100188073	A1	2010-07-29	Rothberg et al.	
	120	20100196911	A1	2010-08-05	Hoffman et al.	
	121	20100197507	A1	2010-08-05	Rothberg et al.	
	122	20100216151	A1	2010-08-26	LAPIDUS et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	123	20100216153	A1	2010-08-26	LAPIDUS et al.	
	124	20100248984	A1	2010-09-30	Shaffer et al.	
	125	20100282617	A1	2010-11-11	Rothberg et al.	
	126	20100285578	A1	2010-11-11	Selden et al.	
	127	20100297626	A1	2010-11-25	MCKERNAN et al.	
	128	20100300559	A1	2010-12-02	Schultz et al.	
	129	20100300895	A1	2010-12-02	Nobile et al.	
	130	20100301042	A1	2010-12-02	Kahlert	
	131	20100301398	A1	2010-12-02	Rothberg et al.	
	132	20100304982	A1	2010-12-02	Hinz et al.	
	133	20100311061	A1	2010-12-09	KORLACH et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	134	20100330619	A1	2010-12-30	Willis et al.	
	135	20110004413	A1	2011-01-06	Carnevali et al.	
	136	20110009278	A1	2011-01-13	Kain et al.	
	137	20110015863	A1	2011-01-20	Pevzner et al.	
	138	20110021366	A1	2011-01-27	Chinitz et al.	
	139	20110034342	A1	2011-02-10	Fox	
	140	20110092375	A1	2011-04-21	Zamore et al.	
	141	20110098193	A1	2011-04-28	Kingsmore et al.	
	142	20110117544	A1	2011-05-19	Lexow	
	143	20110159499	A1	2011-06-30	HINDSON et al.	
	144	20110166029	A1	2011-07-07	Margulies et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	145	20110224105	A1	2011-09-15	Kurn et al.	
	146	20110230365	A1	2011-09-22	Rohlfs et al.	
	147	20110257889	A1	2011-10-20	Klammer et al.	
	148	20110288780	A1	2011-11-24	Rabinowitz et al.	
	149	20110301042	A1	2011-12-08	Steinmann et al.	
	150	20120015050	A1	2012-01-19	Abkevich et al.	
	151	20120021930	A1	2012-01-26	Schoen et al.	
	152	20120046877	A1	2012-02-23	HYLAND et al.	

If you wish to add additional U.S. Published Application citation information please click the Add button. Add

FOREIGN PATENT DOCUMENTS Remove

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2]	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1	1 321 477	EP	A1	2003-06-25	Grawunder, Ulf		

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	2	1 564 306	EP	A2	2005-08-17	Affymetrix, Inc		
	3	10770071.8	EP		2010-11-04	Good Start Genetics, Inc.		
	4	2 437 191	EP	A2	2012-04-04	Gene Security Network Llc		
	5	2425240	EP	A2	2012-03-07	Good Start Genetics, Inc		
	6	2716766	EP	A1	2014-04-09	Berry Genomics Co., Ltd		
	7	00/018957	WO	A1	2000-04-06	Applied Research Systems		
	8	02/093453	WO	A2	2002-11-21	X-mine, Inc		
	9	1996/019586	WO	A1	1996-06-27	Visible Genetics Inc		
	10	2004/018497	WO	A2	2004-03-04	Solexa Ltd		
	11	2004/083819	WO	A2	2004-09-30	Trace Genetics, Inc		
	12	2005/003304	WO	A2	2005-01-13	Illumina, Inc		

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	13	2006/084132	WO	A2	2006-08-10	Agencourt Bioscience Corp		
	14	2007/010251	WO	A2	2007-01-25	Solexa Ltd		
	15	2007/061284	WO	A1	2007-05-31	Plant Res Int BV		
	16	2007/107717	WO	A1	2007-09-27	Ucl Business Plc		
	17	2007/123744	WO	A2	2007-11-01	Solexa, Inc		
	18	2007/135368	WO	A2	2007-11-29	Solexa Ltd		
	19	2008067551	WO	A2	2008-06-05	Navigenics Inc		
	20	2009/036525	WO	A2	2009-03-26	Univ Leuven Kath		
	21	2009/076238	WO	A2	2009-06-18	Complete Genomics, Inc		
	22	2010/024894	WO	A1	2010-03-04	23andme, Inc		
	23	2010/115154	WO	A1	2010-10-07	Fluidigm Corp		

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	24	2010/126614	WO	A2	2010-11-04	Good Start Genetics, Inc		
	25	2011/006020	WO	A1	2011-01-13	Qualcomm Inc		
	26	2011/102998	WO	A2	2011-08-25	Helicos Biosciences Corp		
	27	2011/155833	WO	A2	2011-12-15	Keygene Nv		
	28	2011066476	WO	A1	2011-06-03	Quantalife, Inc		
	29	2011067378	WO	A1	2011-06-09	Olink Genomics Ab		
	30	2012/006291	WO	A2	2012-01-12	Life Technologies Corp		
	31	2012/051208	WO	A2	2012-04-19	Complete Genomics, Inc		
	32	2012/087736	WO	A1	2012-06-28	Good Start Genetics, Inc		
	33	2012/109500	WO	A2	2012-08-16	Bio-rad Laboratories, Inc		
	34	2012/134884	WO	A1	2012-10-04	Good Start Genetics, Inc		

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	35	2012/149171	WO	A1	2012-11-01	Univ California		
	36	2012/170725	WO	A2	2012-12-13	Sinai School Medicine		
	37	2012040387	WO	A1	2012-03-29	Univ Leland Stanford Junior		
	38	2013/058907	WO	A1	2013-04-25	Good Start Genetics, Inc		
	39	2013/148496	WO	A1	2013-10-03	Univ Johns Hopkins		
	40	2013/177086	WO	A1	2013-11-28	Sequenom, Inc		
	41	2013/191775	WO	A2	2013-12-27	Nugen Technologies, Inc		
	42	2014/052909	WO	A2	2014-04-03	Childrens Mercy Hospital		
	43	2014/074246	WO	A1	2014-05-15	Good Start Genetics, Inc		
	44	2015/119941	WO	A2	2015-08-13	Igenomx Internat Genomics Corp		
	45	95/011995	WO	A1	1995-05-04	Affymax Tech Nv		

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	46	98/014275	WO	A1	1998-04-09	Intronn Llc		
	47	98/044151	WO	A1	1998-10-08	Glaxo Group Ltd		

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T5
	1	Abravaya, 1995, Detection of point mutations with a modified ligase chain reaction (Gap-LCR), Nucleic Acids Research, 23(4): 675-682	
	2	Adey, 2010, Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition, Genome Biol 11:R119	
	3	Ageno, 1969, The alkaline denaturation of DNA, Biophys J 9(11):1281-1311	
	4	Agrawal, 1990, Site-specific functionalization of oligodeoxynucleotides for non-radioactive labelling, Tetrahedron Let 31:1543-1546	
	5	Akhras, 2007, Connector inversion probe technology: A powerful one-primer multiplex DNA amplification system for numerous scientific applications, PLoSOne 9:e915	
	6	Akhras, 2007, PathogenMip Assay: a multiplex pathogen detection assay, PLOS One 2:e2230	
	7	Alazard, 2002, Sequencing of production-scale synthetic oligonucleotides by enriching for coupling failures using matrix-assisted laser desorption/ ionization time-of-flight mass spectrometry, Anal Biochem 301:57-64	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	8	Alazard, 2006, Sequencing oligonucleotides by enrichment of coupling failures using matrix-assisted laser desorption/ ionization time-of-flight mass spectrometry, Curr Protoc Nucleic Acid Chem, Chapter 10, Unit 10:1-7	
	9	Albert, 2007, Direct selection of human genomic loci by microarray hybridization, Nature Methods 4(11):903-5	
	10	Aljanabi, 1997, Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques, Nucl. Acids Res 25:4692-4693	
	11	Antonarakis and the Nomenclature Working Group, 1998, Recommendations for a nomenclature system for human gene mutations, Human Mutation 11:1-3	
	12	Archer, 2014, Selective and flexible depletion of problematic sequences from RNA-seq libraries at the cDNA stage, BMC Genomics 15(1):401	
	13	Ball, 2009, Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells, Nat Biotech 27:361-8	
	14	Balzer, 2013, Filtering duplicate reads from 454 pyrosequencing data, Bioinformatics 29(7):830-836	
	15	Barany, 1991, Genetic disease detection and DNA amplification using cloned thermostable ligase, PNAS 88:189-193	
	16	Barany, 1991, The Ligase Chain Reaction in a PCR World, Genome Research 1:5-16	
	17	Bau, 2008, Targeted next-generation sequencing by specific capture of multiple genomic loci using low-volume microfluidic DNA arrays, Analytical and Bioanal Chem 393(1):171-5	
	18	Beer, 1962, Determination of base sequence in nucleic acids with the electron microscope: visibility of a marker, PNAS 48(3):409-416	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	19	Bell, 2011, Carrier testing for severe childhood recessive diseases by next-generation sequencing, Sci Trans Med 3 (65ra4)	
	20	Benner, 2001, Evolution, language and analogy in functional genomics, Trends Genet 17:414-8	
	21	Bentzley, 1996, Oligonucleotide sequence and composition determined by matrix-assisted laser desorption/ionization, Anal Chem 68:2141-2146	
	22	Bentzley, 1998, Base specificity of oligonucleotide digestion by calf spleen phosphodiesterase with matrix-assisted laser desorption ionization analysis, Anal Biochem 258:31-37	
	23	Bhangale, 2006, Automating resequencing-based detection of insertion-deletion polymorphisms, Nature Genetics 38:1457-1462	
	24	Bickle, 1993, Biology of DNA Restriction, Microbiol Rev 57(2):434-50	
	25	Bonfield, 2013, Compression of FASTQ and SAM format sequencing data, PLoS One 8(3):e59190	
	26	Bose, 2012, BIND-An algorithm for loss-less compression of nucleotide sequence data, J Biosci 37(4):785-789	
	27	Boyden, 2013, High-throughput screening for SMN1 copy number loss by next-generation sequencing, American Society of Human Genetics 63rd Annual Meeting, Abstract, October 22, 2013	
	28	Boyer, 1971, DNA restriction and modification mechanisms in bacteria, Ann Rev Microbiol 25:153–76	
	29	Braasch, 2001, Locked nucleic acid (LNA): fine-tuning the recognition of DNA and RNA, Chemistry & Biology 8(1):1-7	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	30	Braslavsky, 2003, Sequence information can be obtained from single DNA molecules, PNAS 100:3960-4.	
	31	Brinkman, 2004, Splice Variants as Cancer Biomarkers, Clin Biochem 37:584	
	32	Brison, 1982, General method for cloning amplified DNA by differential screening, Mol Cell Biol 2(5):578-587	
	33	Brown, 1979, Chemical synthesis and cloning of a tyrosine tRNA gene, Methods Enzymol 68:109-51	
	34	Browne, 2002, Metal ion-catalyzed nucleic Acid alkylation and fragmentation, J Am Chem Soc 124(27):7950-7962	
	35	Brownstein, 2014, An international effort towards developing standards for best practices in analysis, interpretation and reporting of clinical genome sequencing results in the CLARITY Challenge, Genome Biol 15:R53	
	36	Bunyan, 2004, Dosage analysis of cancer predisposition genes by multiplex ligation-dependent probe amplification, British Journal of Cancer, 91(6):1155-59.	
	37	Burrow, 1994, A block-sorting lossless data compression algorithm, Technical Report 124, Digital Equipment Corporation, CA. (24 pages)	
	38	Carpenter, 2013, Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries, Am J Hum Genet 93(5):852-864	
	39	Caruthers, 1985, Gene synthesis machines: DNA chemistry and its uses, Science 230:281-285	
	40	Castellani, 2008, Consenses on the use of and interpretation of cystic fibrosis mutation analysis in clinical practice, J Cyst Fib 7:179-196	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	41	Challis, 2012, An integrative variant analysis suite for whole exome next-generation sequencing data, BMC Informatics 13(8):1-12	
	42	Chan, 2011, Natural and engineered nicking endonucleases-from cleavage mechanism to engineering of strand-specificity, Nucl Acids Res 39(1):1-18	
	43	Chen, 2010, Identification of racehorse and sample contamination by novel 24-plex STR system, Forensic Sci Int: Genetics 4:158-167	
	44	Chennagiri, 2013, A generalized scalable database model for storing and exploring genetic variations detected using sequencing data, American Society of Human Genetics 63rd Annual Meeting, Abstract, October 22, 2013	
	45	Chevreux, 1999, Genome sequence assembly using trace signals and additional sequence information, Proc GCB 99:45-56	
	46	Chirgwin, 1979, Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease, Biochemistry, 18:5294-99.	
	47	Choe, 2010, Novel CFTR Mutations in a Korean Infant with Cystic Fibrosis and Pancreatic Insufficiency, J Korean Med Sci 25:163-5.	
	48	Ciotti, 2004, Triplet repeat prmied PCR (TP PCR) in molecular diagnostic testing for Friedrich ataxia, J Mol Diag 6 (4):285-9	
	49	Cock, 2010, The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants, Nucleic Acids Res 38(6):1767-1771	
	50	Collins, 2004, Finishing the euchromatic sequence of the human genome, Nature 431(7011):931-45	

If you wish to add additional non-patent literature document citation information please click the Add button Add

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

EXAMINER SIGNATURE			
Examiner Signature	/Steven Pohnert/ (02/04/2021)	Date Considered	02/04/2021

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (**Not for submission under 37 CFR 1.99)**		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

See attached certification statement.

The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/Thomas C. Meyers/	Date (YYYY-MM-DD)	2021-01-15
Name/Print	Thomas C. Meyers	Registration Number	36,989

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

Doc code: IDS
Doc description: Information Disclosure Statement (IDS) Filed

PTO/SB/08a (02-18)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (**Not for submission under 37 CFR 1.99)**		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

U.S.PATENTS							Remove
Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	
	1						
If you wish to add additional U.S. Patent citation information please click the Add button.							Add

U.S.PATENT APPLICATION PUBLICATIONS							Remove
Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	
	1						
If you wish to add additional U.S. Published Application citation information please click the Add button.							Add

FOREIGN PATENT DOCUMENTS								Remove
Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1							
If you wish to add additional Foreign Patent Document citation information please click the Add button								Add

NON-PATENT LITERATURE DOCUMENTS			Remove
Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

/SP/	1	Maddalena, 2005, Technical standards and guidelines: molecular genetic testing for ultra-rare disorders, Genet Med 7:571-83	
/SP/	2	Malewicz, 2010, Pregel: a system for large-scale graph processing, Proc. ACM SIGMOD Int Conf Mgmt Data 135-46	
/SP/	3	Mamanova, 2010, Target-enrichment strategies for next-generation sequencing, Nat Meth 7(2):111-118	
/SP/	4	Margulies, 2005, Genome sequencing in micro-fabricated high-density picotiter reactors, Nature, 437:376-380	
/SP/	5	Marras, 1999, Multiplex detection of single-nucleotide variations using molecular beacons, Genetic Analysis: Biomolecular Engineering 14:151	
/SP/	6	Maxam, 1977, A new method for sequencing DNA, PNAS, 74:560-564	
/SP/	7	May, 1988, How Many Species Are There on Earth?, Science 241(4872):1441-9	
/SP/	8	McDonnell, 2007, Antisepsis, disinfection, and sterilization: types, action, and resistance, page 239	
/SP/	9	McKenna, 2010, The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data, Genome Research 20:1297-1303	
/SP/	10	Messiaen, 1999, Exon 10b of the NF1 gene represents a mutational hotspot and harbors a recurrent missense mutation Y489C associated with aberrant splicing, Genetics in Medicine, 1(6):248-253	
/SP/	11	Meyer, 2007, Targeted high-throughput sequencing of tagged nucleic acid samples, Nucleic Acids Research 35(15): e97 (5 pages).	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

/SP/	12	Meyer, 2008, Parallel tagged sequencing on the 454 platform, Nat Protocol 3(2):267-278	
/SP/	13	Miesenbock, 1998, Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins, Nature 394(6689):192-95	
/SP/	14	Miller, 2010, Assembly algorithms for next-generation sequencing data, Genomics 95:315-327	
/SP/	15	Mills, 2010, Mapping copy number variation by population-scale genome sequencing, Nature 470(7332):59-65	
/SP/	16	Miner, 2004, Molecular barcodes detect redundancy and contamination in hairpin-bisulfite PCR, Nucl Acids Res 32 (17):e135	
/SP/	17	Minton, 2011, Mutation Surveyor: software for DNA sequence analysis, Meth Mol Biol 688:143-53	
/SP/	18	Miyake, 2009, PIK3CA gene mutations and umplification in uterine cancers, Canc Lett 261:120-126	
/SP/	19	Miyazaki, 2009, Characterization of deletion breakpoints in patients with dystrophinopathy carrying a deletion of exons 45-55 of the Duchenne muscular dystrophy (DMD) gene, J Hum Gen 54:127-30	
/SP/	20	Mockler, 2005, Applications of DNA tiling arrays for whole-genome analysis, Genomics 85(1):1-15	
/SP/	21	Mohammed, 2012, DELIMINATE-a fast and efficient methods for loss-less compression of genomice sequences, Bioinformatics 28(19):2527-2529	
/SP/	22	Moudrianakis, 1965, Base Sequence Determination in Nucleic Acids with the Electron Microscope, III. Chemistry and Microscopy of Guanine-Labeled DNA, PNAS, 53:564-71	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

/SP/	23	Mullan, 2002, Multiple sequence alignment-the gateway to further analysis, Brief Bioinform 3(3):303-5	
/SP/	24	Munne, 2012, Preimplantation genetic diagnosis for aneuploidy and translocations using array comparative genomic hybridization, Curr Genomics 13(6):463-470	
/SP/	25	Nan, 2006, A novel CFTR mutation found in a Chinese patient with cystic fibrosis, Chinese Med J 119(2):103-9	
/SP/	26	Narang, 1979, Improved phosphotriester method for the synthesis of gene fragments, Meth Enz 68:90-98	
/SP/	27	Nelson, 1989, Bifunctional oligonucleotide probes synthesized using a novel CPG support are able to detect single base pair mutations, Nucl Acids Res 17(18):7187-7194	
/SP/	28	Ng, 2009, Targeted capture and massively parallel sequencing of 12 human exomes, Nature 461(7261):272-6	
/SP/	29	Nicholas, 2002, Strategies for multiple sequence alignment, Biotechniques 32:572-91	
/SP/	30	Nickerson, 1990, Automated DNA diagnostics using an ELISA-based oligonucleotide ligation assay, PNAS 87:8923-7	
/SP/	31	Nielsen, 1999, Peptide Nucleic Acids, Protocols and Applications (Norfolk: Horizon Scientific Press, 1-19)	
/SP/	32	Nilsson, 2006, Analyzing genes using closing and replicating circles, Trends in Biotechnology 24:83-8	
/SP/	33	Ning, 2001, SSAHA: a fast search method for large DNA databases, Genome Res 11(10):1725-9	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

/SP/	34	Nordhoff, 1993, Ion stability of nucleic acids in infrared matrix-assisted laser desorption/ ionization mass spectrometry, Nucl Acid Res 21(15):3347-57	
/SP/	35	Nuttle, 2013, Rapid and accurate large-scale genotyping of duplicated genes and discovery of interlocus gene conversions, Nat Meth 10(9):903-909	
/SP/	36	Nuttle, 2014, Resolving genomic disorder-associated breakpoints within segmental DNA duplications using massively parallel sequencing, Nat Prot 9(6):1496-1513	
/SP/	37	O'Roak, 2012, Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders, Science 338(6114):1619-1622	
/SP/	38	Oefner, 1996, Efficient random sub-cloning of DNA sheared in a recirculating point-sink flow system, Nucleic Acids Res 24(20):3879-3886	
/SP/	39	Oka, 2006, Detection of loss of heterozygosity in the p53 gene in renal cell carcinoma and bladder cancer using the polymerase chain reaction, Mol Carcinogenesis 4(1):10-13	
/SP/	40	Okoniewski, 2013, Precise breakpoint localization of large genomic deletions using PacBio and Illumina next-generation sequencers, Biotechniques 54(2):98-100	
/SP/	41	Okou, 2007, Microarray-based genomic selection for high-throughput reseuqencing, Nat Meth 4(11):907-909	
/SP/	42	Oliphant, 2002, BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping, Biotechniques Suppl:56-8, 60-1	
/SP/	43	Ordahl, 1976, Sheared DNA fragment sizing: comparison of techniques, Nucleic Acids Res 3:2985-2999	
/SP/	44	Ostrer, 2001, A genetic profile of contemporary Jewish populations, Nat Rev Genet 2(11):891-8	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

/SP/	45	Owens, 1998, Aspects of oligonucleotide and peptide sequencing with MALDI and electrospray mass spectrometry, Bioorg Med Chem 6:1547-1554	
/SP/	46	Parameswaran, 2007, A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing, Nucl Acids Rese 35:e130	
/SP/	47	Parkinson, 2012, Preparation of high-quality next-generation sequencing libraries from picogram quantities of target DNA, Genome Res 22:125-133	
/SP/	48	Pastor, 2010, Conceptual modeling of human genome mutations: a dichotomy between what we have and what we shoudl have, 2010 Proc BIOSTEC Bioinformatics, pp: 160-166	
/SP/	49	Paton, 2000, Conceptual modelling of genomic information, Bioinformatics 16(6):548-57	
/SP/	50	Pearson, 1988, Improved tools for biological sequence comparison, PNAS 85(8):2444-8	

If you wish to add additional non-patent literature document citation information please click the Add button Add

EXAMINER SIGNATURE

Examiner Signature	/Steven Pohnert/ (02/04/2021)	Date Considered	02/04/2021

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

See attached certification statement.

The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/Thomas C. Meyers/	Date (YYYY-MM-DD)	2021-01-15
Name/Print	Thomas C. Meyers	Registration Number	36,989

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

Doc code: IDS
Doc description: Information Disclosure Statement (IDS) Filed

PTO/SB/08a (02-18)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	16952764
Filing Date	2020-11-19
First Named Inventor	Gregory Porreca
Art Unit	N/A
Examiner Name	Not Yet Assigned
Attorney Docket Number	MLB-002/06US 34468/40

U.S.PATENTS Remove

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Patent citation information please click the Add button. Add

U.S.PATENT APPLICATION PUBLICATIONS Remove

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Published Application citation information please click the Add button. Add

FOREIGN PATENT DOCUMENTS Remove

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1							

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	1	Holland, 2008, BioJava: an open-source framework for bioinformatics, Bioinformatics 24(18):2096-2097	
	2	Homer, 2008, Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays. PLoS One 4(8):e1000167	
	3	Homer, 2009, BFAST: An alignment tool for large scale genome resequencing, PLoS ONE 4(11):e7767	
	4	Housley, 2009, SNP discovery and haplotype analysis in the segmentally duplicated DRD5 coding region, Ann Hum Genet 73(3):274-282	
	5	Huang, 2008, Comparative analysis of common CFTR polymorphisms poly-T, TGrepeats and M470V in a healthy Chinese population, World J Gastroenterol 14(12):1925-30	
	6	Husemann, 2009, Phylogenetic Comparative Assembly, Algorithms in Bioinformatics: 9th International Workshop, pp. 145-156, Salzberg & Warnow, Eds. Springer-Verlag, Berlin, Heidelberg	
	7	Illumina, 2010, De Novo assembly using Illumina reads, Technical Note (8 pages)	
	8	International Human Genome Sequencing Consortium, 2004, Finishing the euchromatic sequence of the human genome, Nature 431:931-945	
	9	Iqbal, 2012, De novo assembly and genotyping of variants using colored de Bruijn graphs, Nature Genetics 44:226-232	
	10	Isosomppi, 2009, Disease-causing mutations in the CLRN1 gene alter normal CLRN1 protien trafficking to the plasma membrane, Mol Vis 15:1806-1818	
	11	Jaijo, 2010, Microarray-based mutation analysis of 183 Spanish families with Usher syndrome, Invest Ophthalmol Vis Sci 51(3):1311-7	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	12	Jensen, 2001, Orthologs and paralogs—we need to get it right, Genome Biol 2(8):1002-1002.3	
	13	Jones, 2008, Core signaling pathways in human pancreatic cancers revealed by global genomic analyses, Science 321(5897):1801-1806	
	14	Kambara, 1988, Optimization of Parameters in a DNA Sequenator Using Fluorescence Detection, Nature Biotechnology 6:816-821	
	15	Kennedy, 2013, Accessing more human genetic variation with short sequencing reads, American Society of Human Genetics 63rd Annual Meeting, Abstract, October 22, 2013	
	16	Kent, 2002, BLAT--The BLAST-like alignment tool, Genome Res 12(4): 656-664.	
	17	Kerem, 1989, Identification of the cystic fibrosis gene: genetic analysis, Science 245:1073-1080	
	18	Kinde, 2012, FAST-SeqS: a simple an effective method for detection of aneuploidy by massively parallel sequencing, PLoS One 7(7):e41162	
	19	Kircher, 2010, High-througput DNA sequencing - concepts and limitations, Bioassays 32:524-36	
	20	Kirpekar, 1994, Matrix assisted laser desorption/ionization mass spectrometry of enzymatically synthesized RNA up to 150 kDa, Nucl Acids Res 22:3866-3870	
	21	Klein, 2011, LOCAS-A low coverage sequence assembly tool for re-sequencing projects, PLoS One 6(8):e23455	
	22	Kneen, 1998, Green fluorescent protein as a noninvasive intracellular pH indicator, Biophys J 74(3):1591-99	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	23	Koboldt, 2009, VarScan: variant detection in massively parallel sequencing of individual and pooled samples, Bioinformatics 25:2283-85	
	24	Krawitz, 2010, Microindel detection in short-read sequence data, Bioinformatics 26(6):722-729	
	25	Kreindler, 2010, Cystic fibrosis: exploiting its genetic basis in the hunt for new therapies, Pharmacol Ther 125 (2):219-229	
	26	Krishnakumar, 2008, A comprehensive assay for targeted multiplex amplification of human DNA sequences, PNAS 105:9296-301	
	27	Kumar, 2010, Comparing de novo assemblers for 454 transcriptome data, Genomics 11:571	
	28	Kurtz, 2004, Versatile and open software for comparing large genomes, Genome Biol 5:R12	
	29	Lam, 2008, Compressed indexing and local alignment of DNA, Bioinformatics 24(6):791-97	
	30	Langmead, 2009, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, Genome Biol 10:R25	
	31	Larkin, 2007, Clustal W and Clustal X version 2.0, Bioinformatics, 23(21):2947-2948	
	32	Lecompte, 2001, Multiple alignment of complete sequences (MACS) in the post-genomic era, Gene 270(1-2):17-30	
	33	Li, 2003, DNA binding and cleavage by the periplasmic nuclease Vvn: a novel structure with a known active site, EMBO J 22(15):4014-4025	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	34	Li, 2008, SOAP: short oligonucleotide alignment program, Bioinformatics 24(5):713-14	
	35	Li, 2009, Fast and accurate short read alignment with Burrows-Wheeler transform, Bioinformatics, 25 (14):1754-60	
	36	Li, 2009, SOAP2: an improved ultrafast tool for short read alignment, Bioinformatics 25(15):1966-67	
	37	Li, 2009, The Sequence Alignment/Map format and SAMtools, Bioinformatics 25(16):2078-9	
	38	Li, 2010, Fast and accurate long-read alignment with Burrows-Wheeler transform, Bioinformatics 26(5):589-95.	
	39	Li, 2011, Improving SNP discovery by base alignment quality, Bioinformatics 27:1157	
	40	Li, 2011, Single nucleotide polymorphism genotyping and point mutation detection by ligation on microarrays, J Nanosci Nanotechnol 11(2):994-1003	
	41	Li, 2012, A new approach to detecting low-level mutations in next-generation sequence data, Genome Biol 13:1-15	
	42	Li, 2014, HUGO: Hierarchical mUlti-reference Genome cOmpression for aligned reads, JAMIA 21:363-373	
	43	Lin, 2008, ZOOM! Zillions of Oligos Mapped, Bioinformatics, 24:2431	
	44	Lin, 2010, A molecular inversion prove assay for detecting alternative splicing, BMC Genomics 11(712):1-14	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	45	Lin, 2012, Development and evaluation of a reverse dot blot assay for the simultaneous detection of common alpha and beta thalassemia in Chinese, Blood Cells Molecules, and Diseases 48(2):86-90.	
	46	Lipman, 1985, Rapid and sensitive protein similarity searches, Science 227(4693):1435-41	
	47	Liu, 2012, Comparison of next-generation sequencing systems, J Biomed Biotech 2012:251364	
	48	Llopis, 1998, Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins, PNAS 95(12):6803-08	
	49	Ma, 2006, Application of real-time polymerase chain reaction (RT-PCR), J Am Soc 1-15	
	50	MacArthur, 2014, Guidelines for investigating causality of sequence variants in human disease, Nature 508:469-76	

If you wish to add additional non-patent literature document citation information please click the Add button Add

EXAMINER SIGNATURE

Examiner Signature	/Steven Pohnert/ (02/04/2021)	Date Considered	02/04/2021

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (**Not for submission under 37 CFR 1.99)**		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

See attached certification statement.

The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/Thomas C. Meyers/	Date (YYYY-MM-DD)	2021-01-15
Name/Print	Thomas C. Meyers	Registration Number	36,989

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

Doc code: IDS
Doc description: Information Disclosure Statement (IDS) Filed

PTO/SB/08a (02-18)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	16952764
Filing Date	2020-11-19
First Named Inventor	Gregory Porreca
Art Unit	N/A
Examiner Name	Not Yet Assigned
Attorney Docket Number	MLB-002/06US 34468/40

U.S.PATENTS Remove

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Patent citation information please click the Add button. Add

U.S.PATENT APPLICATION PUBLICATIONS Remove

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Published Application citation information please click the Add button. Add

FOREIGN PATENT DOCUMENTS Remove

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1							

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	1	Pertea, 2003, TIGR gene indices clustering tools (TGICL), Bioinformatics 19(5):651-52	
	2	Pieles, 1993, Matrix-assisted laser desorption ionization time-of-flight mass spectrometry: A powerful tool for the mass and sequence analysis of natural and modified oligonucleotides, Nucleic Acids Res 21:3191-3196	
	3	Pinho, 2013, MFCompress: a compression tool for FASTA and multi-FASTA data, Bioinformatics 30(1):117-8	
	4	Porreca, 2007, Multiplex amplificaiton of large sets of human exons, Nat Meth 4(11):931-936	
	5	Porreca, 2013, Analytical performance of a Next-Generation DNA sequencing-based clinical workflow for genetic carrier screening, American Society of Human Genetics 63rd Annual Meeting, Abstract, October 22, 2013	
	6	Pourmand, 2006, PathgoenMIPer: a tool for the design of molecular inversion probes, BMC informatics 7:500	
	7	Procter, 2006, Molecular diagnosis of Prader-Willi and Angelman syndromes by methylation-specific melting analysis and methylation-specific multiplex ligation-dependent probe amplification, Clin Chem 52(7):1276-83	
	8	Qiagen, 2011, Gentra Puregene handbook, 3d Ed. (72 pages)	
	9	Quail, 2010, DNA: Mechanical Breakage, In Encyclopedia of Life Sciences, John Wiley & Sons Ltd, Chicester (5 pages)	
	10	Rambaut, 1997, Seq-Gen:an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees, Bioinformatics 13:235-38	
	11	Richards, 2008 ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions, Genet Med 10(4):294-300	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	12	Richter, 2008, MetaSim-A Sequencing Simulator for Genomics and Metagenomics, PLoS ONE 3:e3373	
	13	Roberts, 1980, Restriction and modification enzymes and their recognition sequences, Nucleic Acids Res 8(1):r63–r80	
	14	Rodriguez, 2010, Constructions from Dots and Lines, Bull Am Soc Inf Sci Tech 36(6):35-41	
	15	Rosendahl, 2013, CFTR, SPINK1, CTRC and PRSS1 variants in chronic pancreatitis: is the role of mutated CFTR overestimated?, Gut 62:582-592	
	16	Rothberg, 2011, An integrated semiconductor device enabling non-optical genome sequencing, Nature 475:348-352	
	17	Rowntree, 2003, The phenotypic consequences of CFTR mutations, Ann Hum Gen 67:471-485	
	18	Saihan, 2009, Update on Usher syndrome, Curr Op Neurology 22(1):19-24	
	19	Sanger, 1977, DNA Sequencing with chain-terminating inhibitors, PNAS 74(12):5463-5467	
	20	Santa Lucia, 1998, A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics, PNAS 95(4):1460-5	
	21	Sargent, 1987, Isolation of differentially expressed genes, Meth Enzym 152:423-432	
	22	Sauro, 2004, How Do You Calculate a Z-Score/ Sigma Level?, https://www.measuringusability.com/zcalc.htm (online publication)	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	23	Sauro, 2004, What's a Z-score and Why Use It in Usability Testing?, https://www.measuringusability.com/z.htm (online publication)	
	24	Schadt, 2010, A window into third-generation sequencing, Human Mol Genet 19(R2):R227-40	
	25	Schatz, 2010, Assembly of large genomes using second-generation sequencing, Genome Res., 20:1165-1173	
	26	Schiffman, 2009, Molecular inversion probes reveal patterns of 9p21 deletion and copy number aberrations in childhood leukemia, Cancer Genetics and Cytogenetics 193:9-18.	
	27	Schneeberger, 2011, Reference-guided assembly of four diverse Arabidopsis thaliana genomes, PNAS 108 (25):10249-10254	
	28	Schouten, 2002, Relative Quantification of 40 Nucleic Acid Sequences by Multiplex Ligation-Dependent Probe Amplification, Nucle Acids Res 30 (12):257	
	29	Schrijver, 2005, Diagnostic testing by CFTR gene mutation analysis in a large group of Hispanics, J Mol Diag 7 (2):289-299	
	30	Schuette, 1995, Sequence analysis of phosphorothioate oligonucleotides via matrix-assisted laser desorption ionization time-of-flight mass spectrometry, J Pharm Biomed Anal 13:1195-1203	
	31	Schwartz, 2009, Identification of cystic fibrosis variants by polymerase chain reaction/oligonucleotide ligation assay, J Mol Diag 11(3):211-15	
	32	Schwartz, 2011, Clinical utility of single nucleotide polymorphism arrays, Clin Lab Med 31(4):581-94	
	33	Sequeira, 1997, Implementing generic, object-oriented models in biology, Ecological Modeling 94.1:17-31	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	34	Shagin, 2002, A novel method for SNP detection, Genome Res 12:1935	
	35	Shen, 2011, High quality DNA sequence capture of 524 disease candidate genes, PNAS 108(16):6549-6554	
	36	Shen, 2013, Multiplex capture with double-stranded DNA probes, Genome Medicine 5(50):1-8	
	37	Sievers, 2011, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, Mol Syst Biol 7:539	
	38	Simpson, 2009, ABySS: A parallel assembler for short read sequence data, Genome Res., 19(6):1117-23	
	39	Slater, 2005, Automated generation of heuristics for biological sequence comparison, BMC Bioinformatics 6:31	
	40	Smirnov, 1996, Sequencing oligonucleotides by exonuclease digestion and delayed extraction matrix-assisted laser desorption ionization time-of-flight mass spectrometry, Anal Biochem 238:19-25	
	41	Smith, 1985, The synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis, Nucl Acid Res 13:2399-2412	
	42	Smith, 2010, Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples, Nucleic Acids Research 38(13):e142 (8 pages).	
	43	Soni, 2007, Progress toward ultrafast DNA sequencing using solid-state nanopores, Clin Chem 53(11):1996-2001	
	44	Spanu, 2010, Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism, Science 330(6010):1543-46	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16952764
	Filing Date	2020-11-19
	First Named Inventor	Gregory Porreca
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	MLB-002/06US 34468/40

	45	Sproat, 1987, The synthesis of protected 5'-mercapto-2',5'-dideoxyribonucleoside-3'-O-phosphoramidites; uses of 5'-mercapto-oligodeoxyribonucleotides, Nucl Acid Res 15:4837-4848	
	46	Streit, 2003, CFTR gene: molecular analysis in patients from South Brazil, Molecular Genetics and Metabolism 78:259-264	
	47	Strom, 2005, Mutation detection, interpretation, and applications in the clinical laboratory setting, Mutat Res 573:160-67	
	48	Summerer, 2009, Enabling technologies of genomic-scale sequence enrichment for targeted high-throughput sequencing, Genomics 94(6):363-8	
	49	Summerer, 2010, Targeted High Throughput Sequencing of a Cancer-Related Exome Subset by Specific Sequence Capture With a Fully Automated Microarray Platform, Genomics 95(4):241-246	
	50	Sunnucks, 1996, Microsatellite and chromosome evolution of parthenogenetic sitobion aphids in Australia, Genetics 144:747-756	

If you wish to add additional non-patent literature document citation information please click the Add button Add

EXAMINER SIGNATURE

Examiner Signature	/Steven Pohnert/ (02/04/2021)	Date Considered	02/04/2021

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /SP/

Electronic Acknowledgement Receipt	
EFS ID:	42101217
Application Number:	16952764
International Application Number:	
Confirmation Number:	7812
Title of Invention:	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS
First Named Inventor/Applicant Name:	Gregory Porreca
Customer Number:	21710
Filer:	Thomas C. Meyers/Sandra Richards
Filer Authorized By:	Thomas C. Meyers
Attorney Docket Number:	MLB-002/06US 34468/40
Receipt Date:	05-MAR-2021
Filing Date:	19-NOV-2020
Time Stamp:	14:01:06
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Affidavit-traversing rejectns or objectns rule 132	MLB-002-06US_Inv-Declaration.pdf	124463 278c9fe3eab9abb60e6ca7a6c71964f5ae569d71	no	4

Warnings:

Information:					
2		MLB-002-06US-Response-Feb2021.pdf	56087 8d8cda80433d2f0cfe9d19c3be872198a74dfb73	yes	12

	Multipart Description/PDF files in .zip description		
	Document Description	**Start**	**End**
	Amendment/Req. Reconsideration-After Non-Final Reject	1	1
	Claims	2	4
	Applicant Arguments/Remarks Made in an Amendment	5	12

Warnings:	
Information:	
Total Files Size (in bytes):	180550

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.
National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.
New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANT:	Molecular Loop Biosolutions, LLC	ART UNIT:	1634
SERIAL NUMBER:	16/952,764	CONF. NO.:	7812
FILING DATE:	November 19, 2020	EXAMINER:	Steven C. Pohnert
TITLE:	**METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS**		

FILED ELECTRONICALLY

Commissioner for Patents
P. O. Box 1450
Alexandria, VA 22313-1450

Declaration under 37 C.F.R. §1.132

I, Gregory Porreca, declare as follow:

1. I received a PhD in Genetics from Harvard University and a BS in computer science and biology from The College of New Jersey. I am the Founder, President, and Chief Executive Officer of Molecular Loop Biosolutions, LLC, a company focused on producing rapid production sequencer-ready DNA libraries. I have served as a lecturer at Harvard Medical School as well as the Personal Genome Project Director of Technology. I co-developed high throughout DNA sequencing system technology that is now part of the Applied Biosystems SOLiD nucleic acid sequencing platform. I am the named inventor on more than 30 patents or patent applications world-wide. I have authored or co-authored over 14 reviewed publications and have presented numerous invited talks and posters on aspects of the next-generation-sequencing and medical genetics.

2. I am the inventor of the above captioned patent application. I have reviewed the above-captioned patent application, including the pending claims.

3. I make this declaration in order to present facts known to me and my technical opinion with respect to the above-captioned patent application.

4. The patent application provides preparative and analytical methods for identifying target nucleic acid molecules while correcting for error or bias introduced during the workflow. *As-filed application*, pg. 6 (lines 20-23); pg. 39 (13-16). The patent application discloses tagging individual target nucleic acid molecules with differentiator tags before any amplification or sequencing steps. *Id.*, pg. 6 (21-24); pg. 40 (7-9); pg. 42 (14-27). Each differentiator tag has its own unique sequence so that each individual target nucleic acid molecule is tagged, and therefore associated, with a differentiator tag. *Id.*, pg. 41 (10-22). Accordingly, each target:differentiator tag combination uniquely identifies a single nucleic acid molecule through subsequence amplification and sequencing steps. *Id.*, pg. 39 (21-26); pg. 41 (3-7), (9-10); pg. 42

(13-20). Because each target:differentiator tag combination can be associated back to its original, individual nucleic molecule based on the unique combination of individual nucleic acid molecules and their respective differentiator tags, the invention corrects for bias in subsequent steps. *Id.*, pg. 41 (7-9), (11-12), (23-24); pg. 42 (12), (17-20). For example, the overrepresentation of sequence reads having a combination of nucleic acid molecule sequence and its differentiator tag sequence is indicative of bias in the protocol at one or more steps between the incorporation of the differentiator tag sequence and the sequencing of the combination. *Id.* Accordingly, the bias may be corrected to provide an accurate representation of the individual nucleic acid molecule in the original pool of nucleic acid molecules. *Id.* For example, it can be recognized that target:differentiator tag combinations having the same sequence are derived from the same original individual nucleic acid molecule and are not representative of two copies of the nucleic acid molecule in the original sample or genome. *Id.*, pg. 7 (11-19), pg. 27 (27-33), pg. 28 (1-2). Target:differentiator tag combinations observed more than once can then be collapsed into a single read to cancel the effect of bias on consensus basecall accuracy. *Id.*, pg. 79, lines 8-15.

5. I have reviewed the Office Action mailed February 10, 2021, from the United States Patent and Trademark Office. I understand that the claims pending in the captioned application are alleged to be obvious by a 2007 Nature Methods publication by myself as first named author, entitled "Multiplex amplification or large sets of human exons." I also understand that the pending claims are alleged to be obvious over my 2007 Nature Methods publication in combination with a 2007 Genome Biology publication by Wang et al. entitled "Analysis of molecular inversion probe performance for allele copy number determination" and a 2007 Journal of Clinical Oncology publication by Schiffman et al. entitled "Adapting molecular inversion probe (MIP) technology for allele quantification in childhood leukemia." I have reviewed each of those cited references.

6. A distinguishing feature of the claims in my patent application is that each individual nucleic acid molecule is tagged with a unique differentiator tag prior to any amplification or sequencing steps. *As filed application*, pg. 40 (lines 7-9); pg. 42 (14-27). My claimed methods overcome systematic and random errors in genome capture, amplification, and sequencing. *Id.*, g. 6 (31-33); pg. 7 (1-4); pg. 28 (1-7); pg. 38 (25-28). In particular, my claimed invention in this patent application allows for a reduction in errors associated with the capture and amplification of nucleic acids by "collapsing" duplicate sequences, thereby providing a significant improvement in the art. *Id.*, pg. 42 (2-3); pg. 45 (2-6); pg. 80 (11-12); and FIG. 5, final "step".

7. As first named author of the 2007 Nature Methods publication, I understand what the publication discloses. The 2007 Nature Methods publication discloses a method of exon capture in which two 20 nucleotide targeting arms of a molecular inversion probe are designed to have unique sequences that hybridize immediately upstream and downstream, respectively, of an exon of interest. *Nature Methods, Figure 1.* Enzymatic gap filling between the two targeting arms results in a circular nucleic acid molecule that comprises the complement to the exon of interest and the two unique targeting arms. *Id.*

8. As first named author of the 2007 Nature Methods publication, I can affirm that the targeting arms of the molecular inversion probes disclosed in the 2007 Nature Methods

publication are not the same as differentiator tags disclose by the invention claimed in this patent application. This is because the unique targeting arms of the molecular inversion probes disclosed in the 2007 Nature Methods publication must be able to hybridize to nucleic acid sequences flanking exons of interest. The becomes a handicap in comparison to the differentiator tags of the claimed invention, for example when determining gene copy number polymorphisms (CNPs), in which the gene as well as its flanking sequences have become duplicated one or more times in a genome. Because the targeting arms of the molecular inversion probes disclosed in my 2007 Nature Methods publication will hybridize to the flanking sequences of *all* copies of the gene, the unique targeting arms cannot be used to differentiate between the identical copies of the gene as separate nucleic acid molecules.

9. As first named author of the 2007 Nature Methods publication, I can affirm that the concept of tagging nucleic molecules with differentiator tags prior to amplification had not been envisioned at the time of the 2007 Nature Methods publication and that I first conceived the claimed invention in 2009 consonant with the filing of the earliest application to which this patent application claims priority.

10. It is further my expert opinion that the additional references cited in the office action do not remedy the deficiencies in the methods reported in my 2007 Nature Methods publication. For example, Wang teaches only the use of molecular inversion probes for the improved measurement of allele copy number, that is the number of different variants of an allele of a gene at a chromosomal location. *Wang, R246.7, Discussion.* However, because the unique targeting arms of the molecular inversion probes needed to capture different alleles of the gene are actually identical, the different copies of the allele must be identified and differentiated by their own sequence rather than by the sequence of the unique targeting arms.

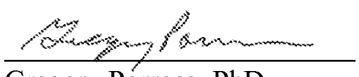
11. Additionally, it is my expert opinion that the molecular inversion probes of Wang suffer from the same handicap as the molecular inversion probes disclosed in my 2007 Nature Methods publication when dealing with copy number polymorphisms. That is, because the targeting arms of the molecular inversion probes will hybridize to the flanking sequences of *any* copies of the gene, the unique targeting arms cannot be used to differentiate between the identical copies of the gene as separate nucleic acid molecules. Wang makes clear this limitation in stating that copy number polymorphisms are likely attributable to identified false positives in their data that are actually true additional copies of the gene. *Id., 246.3, Systematic false positives.* Wang specifically states that one of the shortcomings of their workflow is the presence of CNPs in the autosomes. *Id., 246.3, Trade off between resolution and performance.* Even with the solutions proposed for "ameliorating the effect of miscalling two adjacent markers in a CNP as a false positive", Wang states that doing so is only valid for the assessment of the technology performance "and cannot generate any meaningful biological findings." *Id.*

12 It if further my opinion that Schiffman discloses only the use of molecular inversion probes to analyze loss of heterozygosity events, that is allele quantification or allele copy number as disclosed in Wang. Schiffman does not disclose the use of differentiator probes, and therefore cannot remedy the deficiencies of my 2007 Nature Methods publication or Wang.

13. The assertion that the cited references teach or suggest the presently claimed method is simply unsupported, as none of the references report tagging individual nucleic acid molecules each with a unique barcode sequence.

14. I declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine, imprisonment, or both, under §1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Date: **March 4, 2021**



Gregory Porreca, PhD

Attorney Docket No.: MLB-002/06US 34468/40

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANT:	Molecular Loop Biosolutions, LLC	ART UNIT:	1634
SERIAL NUMBER:	16/952,764	CONF. NO.:	7812
FILING DATE:	November 19, 2020	EXAMINER:	Steven C. Pohnert
TITLE:	**METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS**		

FILED ELECTRONICALLY

Commissioner for Patents
P. O. Box 1450
Alexandria, VA 22313-1450

Response

This paper is in response to the non-final Office Action mailed from the United States Patent and Trademark Office on February 10, 2021.

A **Listing of the Claims** begin on page 2.

Remarks begin on page 5.

Applicant believes that no fees are due with respect to this Response.

LISTING OF THE CLAIMS

1-67. (Canceled)

68. (Previously presented) A method for correcting for errors or bias introduced during nucleic acid analysis workflow, the method comprising the steps of:

obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin;

introducing a set of differentiator tags, wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag;

amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons;

sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and

correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count.

69. (Previously presented) The method of claim 68, wherein said correcting step further comprises collapsing all sequence reads comprising a same differentiator tag into a single read before determining a consensus base call for the target nucleic acid molecule sequences.

Petitioner Illumina, Inc., EX1004, p. 969

70. (Previously presented) The method of claim 69, further comprising the step of determining the presence of a sequencing or amplification error based upon a number of said differentiator tag sequences.

71. (Previously presented) The method of claim 68, wherein said biological sample is selected from the group consisting of blood and biopsy tissue.

72. (Previously presented) The method of claim 68, wherein said plurality of target nucleic acid molecules comprises bacterial or viral nucleic acids isolated from said biological sample.

73. (Previously presented) The method of claim 68, wherein said plurality of target nucleic acid molecules are circulating tumor nucleic acid molecules.

74. (Previously presented) The method of claim 68, further comprising the step of determining the copy number of a genomic region or transcript in a patient from whom the sample is obtained.

75. (Previously presented) The method of claim 68, wherein said plurality of target nucleic acid molecules are maternally-circulating fetal or placental nucleic acid molecules.

76. (Previously presented) A method for correcting for errors or bias introduced during nucleic acid analysis workflow, the method comprising the steps of:

obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin;

introducing a set of unique differentiator tags, wherein members of said set of unique differentiator tags are randomly associated with members of said plurality of target nucleic acid molecules, such that any given tag is uniquely associated with said locus of origin;

amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons;

sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and

correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule.

Remarks

Claims 68-69, 71, 73-74, and 76 are pending and under examination with claims 70, 72, and 75 withdrawn from consideration.

Rejections under 35 U.S.C. §112(b)

On pages 6-7 of the Office Action, claims 68-69, 71, 73-74, and 76 were rejected under 35 U.S.C. §112(b) as being indefinite.

With regard to claim 68, the Office Action stated that the specification "provides no guidance on how to perform collapsing, why collapsing is done or how collapsing corrects errors or bias." With regard to claim 74, the Office Action stated that the specification provides "no guidance on how to determine a copy number of genomic regions or transcript based on collapsing reads of target:differentiator tag combinations." With regard to claim 76, the Office Action stated that the specification "provides no guidance on how to perform recognizing, how recognizing is done or how recognizing corrects errors or bias."

As an initial matter, Applicant submits that the terms "collapsing reads" to correct errors or bias, "determining the copy number of a genomic region" from collapsed reads, and "recognizing target/differentiator tags having the same sequence as being derived from the same input molecule" are terms fully supported and defined in view of the as-filed specification, for example, as detailed in the table below. Not only are those terms in the Specification as filed but the surrounding text makes the application of those term clear.

Support for claim terms rejected in error under 35 U.S.C. §112	
Term	**Support is as-filed specification of U.S. Serial No. 15/231,687**
"collapsing reads" to correct errors or bias *and* "recognizing	"In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve the use of differentiator tag sequences to uniquely tag individual nucleic acid molecules. In some embodiments, the differentiator tag sequence permit the detection of bias based on the occurrence of combinations of differentiator tag and target sequences observed in a sequencing reaction. In other embodiments, the methods reduce errors caused by bias, or the risk of bias, associated with the capture, amplification and sequencing of genetic

target/differentiator tags having the same sequence as being derived from the same input molecule"	loci, e.g., for diagnostic purposes. *Pg. 6, lines 21-28* Aspects of the invention relate to providing sequence tags (referred to as differentiator tags) that are useful to determine whether target nucleic acid sequences identified in an assay are from independently isolated target nucleic acids or from multiple copies of the same target nucleic acid molecule (e.g., due to bias in a preparative step, for example, amplification). This information can be used to help analyze a threshold number of independently isolated target nucleic acids from a biological sample in order to obtain sequence information that is reliable and can be used to make a genotype conclusion (e.g., call) with a desired degree of confidence. This information also can be used to detect bias in one or more nucleic acid preparative steps." *Pg. 6, line 29 to page 7, line 4* "[S]ince any particular target:differentiator tag mapping will not occur more than once by chance, and therefore will only be observed if bias was present somewhere in the sample preparation process. For any target:differentiator tag combination observed more than once, all such reads are **'collapsed'** into a single read before consensus basecalls are determined. This will cancel the effect of bias on consensus basecall accuracy. FIG. 5 depicts a method for making diploid genotype calls in which repeat target:differentiator tag combination are collapsed." *Pg. 79, lines 8-15* "Aspects of the invention are based, in part, on the identification of different sources of ambiguity and error in genetic analyses, and, in part, on the identification of one or more approaches to avoid, reduce, recognize, and/or resolve these errors and ambiguities at different stages in a genetic analysis. Aspects of the invention relate to methods and compositions for addressing bias and/or stochastic variation associated with one or more preparative and/or analytical steps of a nucleic acid evaluation technology." *Pg. 27, line 27 to page 28, line 2*
"determining the copy number of a genomic region"	"Other examples include cancer, which has been associated with microsatellite instability (MSI) involving an increase or decrease in the genomic copy number of nucleic acid repeats at one or more microsatellite loci (e.g., BAT-25 and/or BAT-26). The are currently many sequencing-based assays for determining the number of nucleic acid sequence repeats at a particular locus and identifying the presence of nucleic acid insertions or deletions. However, such techniques are not useful in a high throughput multiplex analysis where the entire length of a region may not be sequenced." *Pg. 58, lines 3-9.* "Accordingly, in some aspects, methods of genotyping a subject are provided. In some embodiments, the methods comprise determining the sequence of at least a threshold number of independently isolated nucleic acids, wherein the sequence of each isolated nucleic acid comprises a target nucleic acid sequence and a differentiator tag sequence, wherein the threshold number is a number of unique combinations of target nucleic acid and differentiator tag sequences, wherein the isolated nucleic acids are identified as independently isolated if they comprise unique combinations of target nucleic acid and differentiator tag sequences, and wherein the target nucleic acid sequence is the sequence of a genomic locus of a subject." *Pg. 7, lines 11-19*

Moreover, in addition to the clear guidance provided for the claim terms in the specification as detailed in the table above, Examples 2 and 3 and Figures 5 and 6 of the as-filed specification provide experimental implementation of the claimed invention, including recognizing target/differentiator tags having the same sequence as being derived from the same input molecule and collapsing the reads to correct for errors or bias. Any fair reading of the application would make clear that Applicant has provided substantial guidance and detail regarding the claimed invention involving collapsing sequence reads to correct of error or bias.

Additionally, Applicant submits that the Office Action has failed to consider Dr. Porreca's Declaration under 37 C.F.R. §1.132 previously submitted with Applicant's response filed February 18, 2020 in the parent application to this continuation application, U.S. Serial No. 15/231,687. The claims examined in U.S. Serial No. 15/231,687, prior to the filing of the instant continuation application, were directed to the same invention currently claimed, namely a method comprising tagging each individual nucleic acid molecule with a differentiator tag prior to amplification or sequencing steps.

Dr. Porreca's declaration represents the understanding of the claims and prior art by a person of skill in the art. Dr. Porreca's qualifications in this area are unimpeachable. The Office Action has provided no basis whatsoever to question Dr. Porreca's objective statements and the Examiner has provided no assertion that he too is an expert in the field and has a contrary opinion to that of a scientist who has been at the forefront of this work for two decades. Dr. Porreca's declaration, at paragraphs 4 and 6, explains in detail how a person of skill in the art would understand that the as-filed specification details collapsing target:differentiator tag combinations to correct for errors and bias. The Federal Circuit has made clear that 35 U.S.C. 112(b) requires only that "the claims read in light of the specification, reasonable apprise those skilled in the art and are as precise as the subject matter permits. As a matter of law, no court can demand more." *Hybritech, Inc. v. Monoclonal Antibodies, Inc.*, 802 F.2d 1367 (Fed. Cir. 1986) (emphasis added). The Office Action's failure to consider Dr. Porreca's understanding of the

claims in light of the specification is in error. Applicant submits that the Examiner must provide evidence to substantiate technical conclusions, or, if the Examiner is subject matter expert, he or she should provide an affidavit providing evidence of expertise and the technical basis for the conclusions made. *In re Kahn*, 441 F.3d 997, 998 (Fed. Cir. 2006).

In view of the above remarks and the Declaration of Dr. Porreca submitted by Applicant on February 18, 2020 in U.S. Serial No. 15/231,687, applicant submits that the rejection under 35 U.S.C. §112(b) has been overcome and applicant requests withdrawal of the rejection.

Rejections under 35 U.S.C. §101

On pages 7-10 of the Office Action, claims 68-69, 71, 73-74, and 76 were rejected under 35 U.S.C. §101 as being directed to a mental step.

The Supreme Court has held that everything under the sun made by a human is patentable under §101, subject to three narrow judicial exceptions: natural phenomena, laws of nature, and abstract ideas. *Diamond v. Chakrabarty*, 447 U.S. 303, 308-309 (1980). Even if the claim is directed to one of the judicial exceptions, then the second part of the analysis provides that the claim is still patent eligible if it contains "significantly more" than the judicial exception. *Mayo Collaborative Servs. v. Prometheus Labs.*, Inc., 556 U.S. 66 (2012); *Alice Corp. v. CLS Bank Int'l*, 573 U.S. 208 (2014).

The USPTO's 2019 Revised Patent Subject Matter Eligibility Guidance (the "*Revised Guidance*") explains that the judicial category of mental concepts seeks to exclude from patentability any process that would include an individual performing the claimed steps mentally or with a pencil and paper. *Revised Guidance on third page.* As a corollary, this judicial category does not include inventions that "could not, as a practical matter, be performed entirely in a human's mind." *CyberSource Corp. v. Retail Decisions, Inc.*, 654 F.3d 1366, 1375 (Fed. Cir. 2015)(emphasis added).

The USPTO's October 2019 Subject Matter Eligibility Update ("*October Update*")

explains that a claim is not "directed to" a judicial exception, and is thus patent eligible, if the judicial exception is integrated into a practical application of the judicial exception. *October Update, page 10.* A claim integrates a judicial exception if it improves the functioning of a technology. *Id., page 12.* The claim itself does not need to explicitly recite the improvement, if the derived improvement is clearly set forth in the specification. *Id.* This analysis is undertaken without reference to what is well-understood, routine, or conventional. *Id.* Moreover, the USPTO's 2019 Revised Patent Subject Matter Eligibility Guidance (*"2019 Guidance"*) states that when considering whether a judicial exception is integrated into a practical application, the analysis "specifically excludes consideration of whether the additional elements represent well-understood, routine, conventional activity . . . [e]xaminers are reminded that a claim that includes conventional elements may still integrate an exception into a practical application, thereby satisfying the subject matter eligibility requirement of Section 101." *2019 Guidance, page 55.*

The claimed invention is not directed to a mental step. Rather, the claimed invention directed to methods comprising numerous active laboratory steps. For example, the claims require introducing differentiator tags to a biological sample. It's hard to imagine how that would be done as a "mental step". Next the claims require sequencing. It is not possible to sequence genomic material mentally. That, as the Examiner must know, has to physically be done in a laboratory. In addition, the claims require amplifying nucleic acid. Again, it is difficult to imagine sequencing in one's head. Each of the claims requires multiple physical laboratory steps, including tagging nucleic acid molecules with differentiator tags, amplifying the tagged nucleic acid molecules to generate amplicons, and sequencing the amplicons. The further steps of correcting for error or bias provide the benefits of the improved workflow provided by the claimed invention.

In view of the above remarks, applicant submits that the pending claims are not directed to a mental step and are patent eligible under 35 U.S.C. §101. Applicant requests withdrawal of the rejection under 35 U.S.C. §101.

Rejections under 35 U.S.C. §103

On pages 10-16 of the Office Action, claims 68-69 and 76 were rejected under 35 U.S.C. §103 in view of Porreca 2007 (Nature Methods, 2007, 4:931-36), claim 74 was rejected under 35 U.S.C. §103 in view of Porreca 2007 in light of Wang (Genome Biology, 2017, 8:R246.1-.14), and claims 71 and 73 were rejected under 35 U.S.C. §103 in view of Porreca 2007 in light of Schiffman (Journal of Clinical Oncology (2007) 25:530).

In making these rejections under 35 U.S.C. §103 the Office Action relies on a factual error. This is because none of the cited reference teach tagging each individual nucleic acid molecule with a differentiator tag prior to amplification or sequencing steps.

In order to establish a prima facie case of obviousness, the prior art must provide motivation to modify a reference or to combine multiple references in a manner that would lead one of ordinary skill in the art to an Applicant's claimed invention with a reasonable expectation of success. *See e.g., KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398 (2007); *see also MPEP § 2143.* The Federal Circuit has made clear that "while 'common sense' can be invoked, even potentially to supply a limitation missing from the prior art, it must still be supported by evidence and a reasoned explanation . . . [T]his is particularly true where the missing limitation goes to the heart of an invention." *Arendi v. Apple*, 832 F.3d 1355, 1363 (Fed. Cir. 2016).

As an initial matter, Applicant submits that the rejection of claims 68-69 and 76 under 35 U.S.C. §103 in view of Porreca 2007 alone is an improper rejection under 35 U.S.C. §103 relying on a single prior-art reference. Page 12 of the Office Action acknowledged that Porreca 2007 did not teach "correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count." The Office Action does not, and cannot, point to a motivation provided in Porreca 2007 itself to modify Porreca 2007 to lead to the claimed invention. Moreover, the Office Action has not provided evidence or a reasoned explanation as to why common sense would lead one to modify Porreca

2007 to provide the missing limitation, when Porreca 2007 itself recognized that PCR amplification steps were a source of nonuniformity error but did not suggest such a solution. *Porreca 2007, pg. 4, Discussion*. The elements of the claims are simply missing from Porreca 2007.

Applicant submits herewith a second Declaration under 37 C.F.R. §1.132 of Dr. Gregory Porreca, the inventor of the instant application and first named author of Porreca 2007. As stated in Dr. Porreca's Declaration, none of the cited reference, including his 2007 paper disclose a method comprising tagging individual nucleic acid molecules with differentiator tags prior to amplification or sequencing steps or the use of such methods for error or bias correction. *Porreca dec., par. 6-12.* The Office Action, based on his above-quoted statement, would appear to agree with this conclusion.

Specifically, Dr. Porreca states that at the time of his 2007 publication, he had not yet envisioned tagging nucleic acid molecules with differentiator tags. *Id., par. 9*. Moreover, Dr. Porreca makes clear that unique targeting arms of the molecular inversion probes disclosed in Porreca 2007, Wang, and Shiffman are not differentiator probes at all. *Id. pars. 8, 11, 12.* This is, because the targeting arms of the molecular inversion probes need to hybridize to the flanking sequences of *any* copies of a gene, and therefore the unique targeting arms cannot be used to differentiate between the identical copies of the gene as separate nucleic acid molecules. *Id.* As a result, collapsing identical reads from the methods of Porreca 2007, Wang, and Shiffman into a single count, rather than correcting for error or bias introduced during their workflow, would only further introduce error or bias by incorrectly counting multiple copies of a gene present in an original sample as a single copy. Applicant respectfully submits that since key elements are missing from all of the cited references, it would be improper to maintain the prior art rejections. If the Examiner believes that he has technical knowledge that contradicts that of Dr. Porreca, the recognized expert in this area, the Examiner is requested to provide a basis for his expertise and his own declaration refuting the conclusions of Dr. Porreca.

In view of the above remarks and accompanying Declaration of Dr. Porreca, applicant submits that the pending claims are non-obvious of the cited art and request withdrawal of the rejection under 35 U.S.C. §103.

Summary

Applicant submits that the pending claims are in condition for allowance, which is respectfully requested. If there are any questions regarding these remarks, the Examiner is invited to contact Applicant's representative.

Dated: March 5, 2021

Respectfully submitted,

BROWN RUDNICK LLP

BROWN RUDNICK LLP
One Financial Center
Boston, MA 02111
Tel: (617) 856-8483
Email: tmeyers@brownrudnick.com

/Thomas C. Meyers/
Thomas C. Meyers, Reg. No. 36,989
Attorney for Applicant

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/952,764	11/19/2020	Gregory Porreca	MLB-002/06US 34468/40	7812

21710 7590 04/09/2021

BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

EXAMINER	
POHNERT, STEVEN C	

ART UNIT	PAPER NUMBER
1634	

NOTIFICATION DATE	DELIVERY MODE
04/09/2021	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

ip@brownrudnick.com
usactions@brownrudnick.com

Office Action Summary	**Application No.** 16/952,764	**Applicant(s)** Porreca et al.	
	Examiner STEVEN C POHNERT PhD	**Art Unit** 1634	**AIA (FITF) Status** No

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☑ Responsive to communication(s) filed on 3/5/2021.

☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2a)☑ This action is **FINAL.** 2b) ☐ This action is non-final.

3)☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

4)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims*

5) ☑ Claim(s) 68-76 is/are pending in the application.

5a) Of the above claim(s) 70,72 and 75 is/are withdrawn from consideration.

6) ☐ Claim(s) _____ is/are allowed.

7) ☑ Claim(s) 68-69,71,73-74 and 76 is/are rejected.

8) ☐ Claim(s) _____ is/are objected to.

9) ☐ Claim(s) _____ are subject to restriction and/or election requirement

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to **PPHfeedback@uspto.gov.**

Application Papers

10)☑ The specification is objected to by the Examiner.

11)☑ The drawing(s) filed on 11/19/20 is/are: a)☑ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a)☐ All b)☐ Some** c)☐ None of the:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)

2) ☐ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b) Paper No(s)/Mail Date _____.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date _____.

4) ☐ Other: _____.

Petitioner Illumina, Inc., EX1004, p. 981

DETAILED ACTION

Notice of Pre-AIA or AIA Status

1. The present application is being examined under the pre-AIA first to invent provisions.

Claim Status and Formal matters

The instant action is in response to papers filed 3/5/2021.

No claims are amended.

Claims 68-76 are pending.

Applicant's election without traverse of bias, blood as the biological sample, circulating tumor cells and copy number of a genomic region in the reply filed on 2/1/2021 is acknowledged.

Claims 70, 72, 75 are withdrawn from further consideration pursuant to 37 CFR 1.142(b) as being drawn to a nonelected species, there being no allowable generic or linking claim. Election was made **without** traverse in the reply filed on 2/1/2021.

Claims 68-69,71, 73-74, 76 are being examined.

Priority

2. The instant application was filed 11/19/2020 is a continuation of 15231687, filed 08/08/2016 which is a continuation of 13266862, filed 03/13/2012 which is a national stage entry of PCT/US10/01293 having an international filing date: 04/30/2010 which claims priority from provisional application 61182089, filed 05/28/2009, claims priority from provisional application 61179358, filed 05/18/2009 claims priority from provisional application 61178923, filed 05/15/2009 and claims priority from provisional application 61174470, filed 04/30/2009.

Applicant's claim for the benefit of a prior-filed application under 35 U.S.C. 119(e) or under 35 U.S.C. 120, 121, 365(c), or 386(c) is acknowledged. Applicant has not complied with one or more conditions for receiving the benefit of an earlier filing date under 35 U.S.C. 35 U.S.C. 119 (e)120, 121, 365(c), or 386(c) as follows: The later-filed application must be an application for a patent for an invention which is also disclosed in the prior application (the parent or original nonprovisional application or provisional application). The disclosure of the invention in the parent application and in the later-filed application must be sufficient to comply with the requirements of 35 U.S.C. 112(a) or the first paragraph of pre-AIA 35 U.S.C. 112, except for the best mode requirement. See *Transco Products, Inc. v. Performance Contracting, Inc.*, 38 F.3d 551, 32 USPQ2d 1077 (Fed. Cir. 1994)

The disclosure of the prior-filed application, Application No. 61182089, 61179358, 61178923, 61174470, fails to provide adequate support or enablement in the manner provided by 35 U.S.C. 112(a) or pre-AIA 35 U.S.C. 112, first paragraph for one or more claims of this application.

The instant claims were submitted as a preliminary amendment on 11/19/20. The instant claims require, "introducing a set of differentiator tags, wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag; amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons; sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence;

and correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count." Provisional application 61/179358 is the only application that discloses collapsing sequencing reads. However it requires each target loci has a different differentiator tags. Review of provisional applications, 61182089, 61178923, 61174470 did not reveal support for the collapsing or introducing step. Thus the claims are being afforded priority to the instant filing date.

Response to Arguments

The response does not address the priority issue

Response to Amendment

3. The declaration filed by Porecca under 37 CFR 1.132 filed 3/5/2021 is insufficient to overcome the rejection of claims 68-69,71, 73-74, 76 based upon obviousness as set forth in the last Office action because: The declaration indicates it is by the inventor and employee and thus has a vested interest in the outcome. The declaration begins by providing the declarants background and review of the declarant's review of the invention.

MPEP 716.01(c) states:

III. OPINION EVIDENCE

Although factual evidence is preferable to opinion testimony, such testimony is entitled to consideration and some weight so long as the opinion is not on the ultimate legal conclusion at issue. While an opinion as to a legal conclusion is not entitled to any weight, the underlying basis for the opinion may be persuasive. *In re Chilowsky,* 306 F.2d 908, 134 USPQ 515 (CCPA 1962) (expert opinion that an application meets the requirements of 35 U.S.C. 112 is not entitled to any weight; however, facts supporting a basis for deciding that the specification complies with 35 U.S.C. 112 are entitled to some weight); *In re Lindell,* 385 F.2d 453, 155 USPQ 521 (CCPA 1967) (Although an affiant's or declarant's opinion on the ultimate legal issue is not evidence in the case, "some weight ought to be given to a persuasively supported statement of one skilled in the art on what was not obvious to him." 385 F.2d at 456, 155 USPQ at 524 (emphasis in original)).

In assessing the probative value of an expert opinion, the examiner must consider the nature of the matter sought to be established, the strength of any opposing evidence, the interest of the expert in the

outcome of the case, and the presence or absence of factual support for the expert's opinion. *Ashland Oil, Inc. v. Delta Resins & Refractories, Inc.,* 776 F.2d 281, 227 USPQ 657 (Fed. Cir. 1985), *cert. denied,* 475 U.S. 1017 (1986). See also *In re Oelrich,* 579 F.2d 86, 198 USPQ 210 (CCPA 1978) (factually based expert opinions on the level of ordinary skill in the art were sufficient to rebut the *prima facie* case of obviousness); *Ex parte Gray,* 10 USPQ2d 1922 (Bd. Pat. App. & Inter. 1989) (statement in publication dismissing the "preliminary identification of a human b-NGF-like molecule" in the prior art, even if considered to be an expert opinion, was inadequate to overcome the rejection based on that prior art because there was no factual evidence supporting the statement); *In re Carroll,* 601 F.2d 1184, 202 USPQ 571 (CCPA 1979) (expert opinion on what the prior art taught, supported by documentary evidence and formulated prior to the making of the claimed invention, received considerable deference); *In re Beattie,* 974 F.2d 1309, 24 USPQ2d 1040 (Fed. Cir. 1992) (declarations of seven persons skilled in the art offering opinion evidence praising the merits of the claimed invention were found to have little value because of a lack of factual support); *Ex parte George*, 21 USPQ2d 1058 (Bd. Pat. App. & Inter. 1991) (conclusory statements that results were "unexpected," unsupported by objective factual evidence, were considered but were not found to be of substantial evidentiary value).

In paragraph five the declaration begins by providing the inventors opinion on the teachings of Porecca(Nature methods (2007) volume 4, pages 931-936)by providing applicant's opinion that the differentiator tags of the instant claims are different than the MIPS of Porecca (Nature methods (2007) volume 4, pages 931-936). This argument has been thoroughly reviewed but is not considered persuasive as the specification teaches, "As used herein, a "differentiator tag sequence" is a sequence of a nucleic acid (a preparative nucleic acid), which in the context of a plurality of different isolated nucleic acids, identifies a unique, independently isolated nucleic acid."

The declaration continues by asserting each individual nucleic acid is tagged with a unique differentiator tag. This argument has been thoroughly reviewed but is not considered persuasive as this is not a limitation of the claim. Although the claims are interpreted in light of the specification, limitations from the specification are not read into the claims. See *In re Van Geuns*, 988 F.2d 1181, 26 USPQ2d 1057 (Fed. Cir. 1993).

The declaration continues by providing the inventors opinion that the teachings of Porecca (Nature methods) does not envision the differentiator tags of the claims. This argument has been thoroughly reviewed but considered persuasive as the specification teaches, "As used herein, a "differentiator tag sequence" is a sequence of a nucleic acid

(a preparative nucleic acid), which in the context of a plurality of different isolated nucleic acids, identifies a unique, independently isolated nucleic acid." Thus this broadly encompasses the targeting arms of a molecular inversion probe (MIP).

The declaration continues by asserting, "Because the targeting arms of the molecular inversion probes disclosed in my 2007 Nature Methods publication will hybridize to the flanking sequences of all copies of the gene, the unique targeting arms cannot be used to differentiate between the identical copies of the gene as separate nucleic acid molecules." This argument has been thoroughly reviewed but is not considered persuasive as the claims do not specifically require the differentiator tag be used to differentiate between the identical copies of the gene as separate nucleic acid molecules.

The declaration continues by asserting the secondary references do not overcome the deficiencies of Porecca (Nature methods). This argument has been thoroughly reviewed but is not considered persuasive as the alleged deficiencies are due to a narrow interpretation of the claims, which is not consistent with the broadest reasonable interpretation of the claims.

Specification

4. The specification is objected to as failing to provide proper antecedent basis for the claimed subject matter. See 37 CFR 1.75(d)(1) and MPEP § 608.01(o). Correction of the following is required: Claim 73 has been added by amendment and recites, "wherein said plurality of target nucleic acid molecules are circulating tumor nucleic acid molecules." Review and searching of the specification did not reveal antecedent basis for this amendment.

Response to Arguments

The response provides no arguments to the objection to the specification.

Claim Rejections - 35 USC § 112

5. The following is a quotation of the first paragraph of 35 U.S.C. 112(a):

> (a) IN GENERAL.—The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor or joint inventor of carrying out the invention.

The following is a quotation of the first paragraph of pre-AIA 35 U.S.C. 112:

> The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor of carrying out his invention.

6. Claims 68-69,71, 73-74, 76 rejected under 35 U.S.C. 112(a) or 35 U.S.C. 112 (pre-AIA), first paragraph, as failing to comply with the written description requirement. The claim(s) contains subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor or a joint inventor, or for applications subject to pre-AIA 35 U.S.C. 112, the inventor(s), at the time the application was filed, had possession of the claimed invention.

Claim 68 recites, "correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count." The specification recites collapsing three times. The specification in figure 5 demonstrates 4 bases with one base difference in sequences. It demonstrates in a very simple example that sequences can be preferentially amplified. The specification does not any guidance on what is required of collapsed, how collapsing is done or how

collapsing corrects errors. How to determine to collapse sequences of greater complexity, or what is require of a target sequence to identify it to be collapsed. Thus the claims lack adequate written description.

Further claim 74 recites, "further comprising the step of determining the copy number of a genomic region or transcript in a patient from whom the sample is obtained.." The specification and claims provide no guidance on how to determine a copy number of genomic region or transcript based on collapsing reads of target: differentiator tag combinations. Thus the claims lack adequate written description.

Claim 76 recites, "correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule.." The specification recites recognize six times. The specification does not recite, in the context of the limitations of claim 76.The specification in figure 5 demonstrates 4 bases with one base difference in sequences. It demonstrates in a very simple example that sequences can be preferentially amplified. The specification does not any guidance on what is required of recognizing, how recognizing is done or how recognizing corrects errors. How to determine to recognizing sequences of greater complexity, or what is require of a target sequence to identify it to be corrected. Thus the claims lack adequate written description.

Response to Arguments

The response traverses the 112 rejections with respect to collapsing and recognizing by citing pages 6, lines 21-28; pages6 line 29-page 7, line 4; page 79, lines 8-15; and page 27, lines 27 to page 28 line 2. This argument has been thoroughly reviewed but is not considered persuasive as the single recitation of collapse in the

cited section merely states collapsing into a single read, but does not disclose how such collapsing is done. While the cited sections recite "recognize" they do not provide recite recognize in the context of claim 76 or provide how such recognizing is done to correct errors.

The response traverses the rejection with respect to claim 74 by citing page 58, lines 3-9, page 7, lines 11-19. This argument has been thoroughly reviewed but is not considered persuasive as while page 58 provide antecedent basis for copy number it does not provide how copy number is determined in the context of claim 74 as it depends from claim 68. Further the teachings of page 7 do not recite copy number of provide any specific guidance.

The response continues by asserting there is clear guidance in examples 2 and 3 and figure 5. The specification in examples 2 and 3 figure 5 demonstrates 4 bases with one base difference in sequences. It demonstrates in a very simple example that sequences can be preferentially amplified. The specification does not any guidance on what is required of recognizing, how recognizing is done or how recognizing corrects errors.

The response continues by providing arguments to the declaration of Porecca filed 2/18/2020. Applicant refers to an affidavit or declaration filed in the prior application. Affidavits or declarations, such as those submitted under 37 CFR 1.130, 1.131 and 1.132, filed during the prosecution of the prior application do not automatically become a part of this application. Where it is desired to rely on an earlier-filed affidavit or declaration, the applicant should make the remarks of record in this

application and include a copy of the original affidavit or declaration filed in the prior application.

Further the instant declaration is not persuasive for the reasons of record.

7. The following is a quotation of 35 U.S.C. 112(b):

> (b) CONCLUSION.—The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the inventor or a joint inventor regards as the invention.

The following is a quotation of 35 U.S.C. 112 (pre-AIA), second paragraph:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

8. Claims 68-69,71, 73-74, 76 rejected under 35 U.S.C. 112(b) or 35 U.S.C. 112 (pre-AIA), second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which the inventor or a joint inventor (or for applications subject to pre-AIA 35 U.S.C. 112, the applicant), regards as the invention.

Claim 68 recites, "correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count." The specification recites collapsing three times. The specification provides no guidance on how to perform collapsing, why collapsing is done or how collapsing corrects errors of bias. Thus the recitation is incomplete and unclear as it provides no guidance on how to perform collapsing, why collapsing is done or how collapsing corrects errors of bias.

Further claim 74 recites, "further comprising the step of determining the copy number of a genomic region or transcript in a patient from whom the sample is obtained.." The recitation is incomplete and unclear as the specification and claims provide no guidance on how to determine a copy number of genomic region or transcript based on collapsing reads of target: differentiator tag combinations.

Claim 76 recites, "correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule.." The specification recites recognize six times. The specification provides no guidance on how to perform recognizing, why recognizing is done or how recognizing corrects errors of bias. Thus the recitation is incomplete and unclear as it provides no guidance on how to perform recognizing, why recognizing is done or how recognizing corrects errors of bias.

Response to Arguments

The response appears to traverse the rejections for the reasons presented with respect to the 112(a) rejection. This argument is not persuasive for the reasons of record.

Claim Rejections - 35 USC § 101

9. 35 U.S.C. 101 reads as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

10. Claims 68-69, 71, 73-74, 76 are rejected under 35 U.S.C. 101 because the claimed invention is directed to a mental step or abstract idea without significantly more. The claim(s) recite(s) the abstract idea or mental step of correcting. This judicial exception is not integrated into a practical application because if there is no steps which depend from or otherwise flow from the correcting step. The claim(s) does/do not include additional elements that are sufficient to amount to significantly more than the judicial exception because if there are no additional steps integrating the correcting.

Claim analysis

The instant claim 68 is directed towards a method for correcting for errors or bias introduced during nucleic acid analysis workflow, the method comprising the steps of: obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin; introducing a set of differentiator tags, wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag; amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons; sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and correcting for error or bias introduced during said workflow. The correcting step is a mental step or abstract idea.

The instant claim 76 is directed towards method for correcting for errors or bias introduced during nucleic acid analysis workflow, the method comprising the steps of: obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin; introducing a set of unique differentiator tags, wherein members of said set of unique differentiator tags are randomly associated with members of said plurality of target nucleic acid molecules, such that any given tag is uniquely associated with said locus of origin; amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons; sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule

sequence and a differentiator tag sequence; and correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule.. The correcting step is a mental step or abstract idea.

The steps of obtaining, introducing, amplifying and sequencing are active steps.

Dependent claims set forth further limitations to about the reference level, method of detecting expression, subject and source of cancer.

According to the 2019 Patent Eligibility Guidance an initial two step analysis is required for determining statutory eligibility.

Step 1. Is the claim directed to a process, machine, manufacture, or composition of matter? In the instant case the Step 1 requirement is satisfied as the claims are directed towards a process.

Step 2A Prong one. Does the claim recite a law of nature, a natural phenomenon or an abstract idea? Yes, abstract idea and law.

With regards to claim 68, the claim recites, "and correcting for error or bias introduced during said workflow." This is an abstract idea or mental step.

With regards to claim 76, the claim recites, "correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule." This is an abstract idea or mental step.

Step 2A prong two. Does the claim recite additional elements that integrate the judicial exception into a practical application? The answer is no as the claims provide no active step in which the correcting is relied upon or otherwise integrated.

Step 2B. Does the claim recite additional elements that are significantly more than the judicial exceptions? No

Claims 68 and 76 require the steps of obtaining, introducing, amplifying and sequencing are active steps.. The specification teaches:

Sample preparative methods may fall into three classes: 1) single- or several-target amplification (e.g., uniplex PCR, 'multiplex' PCR), 2) multi-target hybridization enrichment (e.g., Agilent SureSelect 'hybrid capture' [Gnirke et al 2009, Nature methods 27:182-9], Roche/Nimblegen 'sequence capture' [Hodges et al 2007, Nature genetics 39:1522-7], and 3) multi-target circularization selection (e.g. molecular inversion probes or padlock probes, [Porreca et al 2007, Nature methods 4:931-6, Turner et al 2009, Nature methods 6:315-6], 'selectors' [Dahl et al 2005, Nucleic acids research 33:e71]). arrays, single nucleotide polymorphism (SNP) arrays (Affymetrix, Santa Clara, Calif.).

Response to Arguments

The response traverses the rejection asserting the claim is not to a mental step, but laboratory methods. This argument has been thoroughly reviewed but is not considered persuasive as the preamble of the claim recites," method for correcting for errors or bias introduced during nucleic acid analysis workflow." Further the last active step of the claim recites, "correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count." Thus while the claims provide routine and conventional assaying steps the preamble and last active step are drawn to mental steps or abstract ideas.

Thus the rejection is maintained.

Claim Rejections - 35 USC § 103

11. In the event the determination of the status of the application as subject to AIA 35 U.S.C. 102 and 103 (or as subject to pre-AIA 35 U.S.C. 102 and 103) is incorrect, any correction of the statutory basis for the rejection will not be considered a new ground of

rejection if the prior art relied upon, and the rationale supporting the rejection, would be the same under either status.

12. The following is a quotation of pre-AIA 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

13. Claim 68-69, 76 is/are rejected under pre-AIA 35 U.S.C. 103(a) as being unpatentable over Porecca(Nature methods (2007) volume 4, pages 931-936).

The specification provides no guidance on how, "correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count" is done thus this is considered broad.

Further the specification teaches, "As used herein, a "differentiator tag sequence" is a sequence of a nucleic acid (a preparative nucleic acid), which in the context of a plurality of different isolated nucleic acids, identifies a unique, independently isolated nucleic acid." Thus this broadly encompasses the targeting arms of a molecular inversion probe (MIP).

Porecca teaches figure 1:

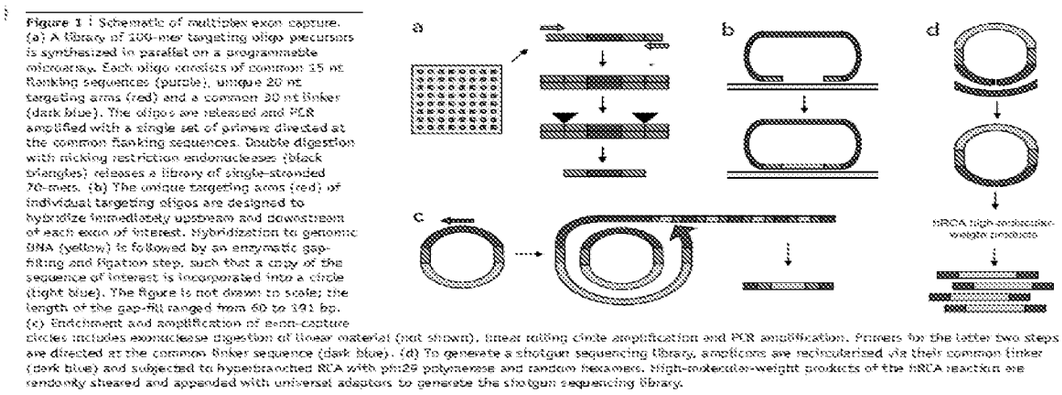


Figure 1 | Schematic of multiplex exon capture. (a) A library of 100-mer targeting oligo precursors is synthesized in parallel on a programmable microarray. Each oligo consists of common 15 nt flanking sequences (purple), unique 20 nt targeting arms (red) and a common 30 nt linker (dark blue). The oligos are released and PCR amplified with a single set of primers directed at the common flanking sequences. Double digestion with nicking restriction endonucleases (black triangles) releases a library of single-stranded 70-mers. (b) The unique targeting arms (red) of individual targeting oligos are designed to hybridize immediately upstream and downstream of each exon of interest. Hybridization to genomic DNA (yellow) is followed by an enzymatic gap-filling and ligation step, such that a copy of the sequence of interest is incorporated into a circle (light blue). The figure is not drawn to scale; the length of the gap-fill ranged from 60 to 191 bp. (c) Enrichment and amplification of exon-capture circles includes exonuclease digestion of linear material (not shown), linear rolling circle amplification and PCR amplification. Primers for the latter two steps are directed at the common linker sequence (dark blue). (d) To generate a shotgun sequencing library, amplicons are recircularized *via* their common linker (dark blue) and subjected to hyperbranched RCA with phi29 polymerase and random hexamers. High-molecular-weight products of the hRCA reaction are randomly sheared and appended with universal adaptors to generate the shotgun sequencing library.

On page 935, 2nd column Porrecca teaches genomic DNA from lyphoblastoid cell lines were hybridized to targeting oligos (MIPS via unique 20 nucleotide targeting arms), ligated and amplified and then sequenced.

Porecca does not specifically teach, "correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count"

However, Porecca teaches, "All evaluated positions at which the genotype is homozygous were dominated by the correct base-call (n ¼ 388). The data on heterozygous positions (n ¼ 90) are more ambiguous. For example, if we were to call positions with allele frequencies between 0.2 and 0.8 as heterozygous, then only 25 of 90 heterozygous positions (31%) would be called correctly. An alternative perspective is that because the goal of variation discovery is often to interrogate for rare variants, the key task is to identify positions that deviate from the reference genome with high sensitivity and specificity. If we call positions with observed allele frequencies of o0.8 as likely to be mutated relative to the reference genome, we can estimate a sensitivity of 63% for detecting rare mutations while maintaining high specificity (sensitivity, 57 of 90

heterozygous positions had a reference allele frequency of less than 0.8; specificity, 0 of 313 homozygous positions had a reference allele frequency of less than 0.8)." (page 934, 1st column)

Therefore it would have been prima facie obvious to one of ordinary skill in the art at the time the invention was made to sequence the arms of the MIPS and collapse (or identify) sequences including arms with greater than 0.8 allele frequency and identify they them as homozygous. The artisan would have a reasonable expectation of success as the artisan is merely using the guidance of Porreca to identify and collapse sequences that are homozygous. The artisan would have a reasonable expectation of success as the artisan is using techniques demonstrated by the art to work.

With regards to claim 69, Porecca teaches homozygous positions were dominated by the correct base call. Thus it is obvious to collapse all sequence reads before making a base call.

The specification provides no guidance on how, "correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule" is done thus this is considered broad.

Further the specification teaches, "As used herein, a "differentiator tag sequence" is a sequence of a nucleic acid (a preparative nucleic acid), which in the context of a plurality of different isolated nucleic acids, identifies a unique, independently isolated nucleic acid." Thus this broadly encompasses the targeting arms of a molecular inversion probe (MIP).

However, Porecca teaches, "All evaluated positions at which the genotype is homozygous were dominated by the correct base-call (n ¼ 388). The data on heterozygous positions (n ¼ 90) are more ambiguous. For example, if we were to call positions with allele frequencies between 0.2 and 0.8 as heterozygous, then only 25 of 90 heterozygous positions (31%) would be called correctly. An alternative perspective is that because the goal of variation discovery is often to interrogate for rare variants, the key task is to identify positions that deviate from the reference genome with high sensitivity and specificity. If we call positions with observed allele frequencies of o0.8 as likely to be mutated relative to the reference genome, we can estimate a sensitivity of 63% for detecting rare mutations while maintaining high specificity (sensitivity, 57 of 90 heterozygous positions had a reference allele frequency of less than 0.8; specificity, 0 of 313 homozygous positions had a reference allele frequency of less than 0.8)." (page 934, 1st column)

Therefore it would have been prima facie obvious to one of ordinary skill in the art at the time the invention was made to sequence the arms of the MIPS and recognize (or identify) sequences including arms with greater than 0.8 allele frequency and identify they them as homozygous. The artisan would motivated merely using the guidance of Porreca to recognize sequences with the same target sequences and MIP targeting arms as homozygous. The artisan would have a reasonable expectation of success as the artisan is using techniques demonstrated by the art to work.

Response to Arguments

The response traversing the rejection by asserting the prior art of record provides no specific motivation. This argument has been thoroughly reviewed but is not

considered persuasive as Porecca teaches, "All evaluated positions at which the genotype is homozygous were dominated by the correct base-call (n ¼ 388). The data on heterozygous positions (n ¼ 90) are more ambiguous. For example, if we were to call positions with allele frequencies between 0.2 and 0.8 as heterozygous, then only 25 of 90 heterozygous positions (31%) would be called correctly. An alternative perspective is that because the goal of variation discovery is often to interrogate for rare variants, the key task is to identify positions that deviate from the reference genome with high sensitivity and specificity. If we call positions with observed allele frequencies of o0.8 as likely to be mutated relative to the reference genome, we can estimate a sensitivity of 63% for detecting rare mutations while maintaining high specificity (sensitivity, 57 of 90 heterozygous positions had a reference allele frequency of less than 0.8; specificity, 0 of 313 homozygous positions had a reference allele frequency of less than 0.8)." (page 934, 1st column)

The response continues traversing the rejection in view of the declaration of Porecca. This argument is not persuasive for the reasons of record.

The response continues by asserting Porecca had not yet envisioned differentiator tags. This argument has been thoroughly reviewed but is not considered persuasive as in view of the breadth of the teachings of the specification the arms of the MIPs can be construed as differentiator tags. The specification teaches, "As used herein, a "differentiator tag sequence" is a sequence of a nucleic acid (a preparative nucleic acid), which in the context of a plurality of different isolated nucleic acids, identifies a unique, independently isolated nucleic acid." Thus this broadly encompasses the targeting arms of a molecular inversion probe (MIP).

14. Claim74 is rejected under pre-AIA 35 U.S.C. 103(a) as being unpatentable over Porecca(Nature methods (2007) volume 4, pages 931-936).as applied to claim 68-69, 76 above, and further in view of Wang (Genome Biology (2007) volume 8, R246.1-246.14).

The teachings of Porecca are set forth above and teach the detection of homozygous alleles by use of unique targeting arms of MIPS and recognizing or collapsing the combination of unique target arms and alleles (or target sequences).

Porecca does not specifically teach determination of copy number.

However, Wang teaches, "One of the techniques that have previously been described for the measurement of ACN is molecular inversion probes (MIPs) [10-12]. Briefly, MIP probes are circularizable oligonucleotides, where the two ends carry two sequences that are complementary to two sequences on the genome separated by one nucleotide (exactly where the variant to be genotyped is). After hybridization to the genomic DNA, the reaction is split into four tubes where a single nucleotide is added to each tube. Upon the addition of the nucleotide, the MIP probe is ligated closed (but this only occurs in the tube with the nucleotide that is complementary to the allele on the genome), turning the probe into a circle. This structure can be selected for by the use of exonucleases, allowing for minimal cross talk' between probes and making it possible to obtain high quality data from highly multiplexed assays (>50,000- plex)" (page 1-2).

Wang teaches, "We describe in this manuscript significant improvements we have made to the MIP-based measurements of ACN. By increasing the proportion of genomic targets that are hybridized to the MIP probes, we have improved the

performance while requiring a smaller amount of DNA. Additionally, for copy number measurements there are substantial advantages in uniformity and robustness when utilizing one-color readouts, especially at high levels of multiplexing" (page 7). Wang teaches, "These observations led us to suspect that it should be possible to genotype mixed DNA populations, such as occurs in tumor samples contaminated with normal tissue."

Therefore it would have been prima facie obvious to one of skill in the art at the time the invention was made to use MIPs to determine copy number. The artisan would be motivated as it would allow detection using less DNA and would allow genotype of mixed populations of cells to be determined. The artisan would have a reasonable expectation of success as the artisan is merely using MIPs as the art suggests.

Response to Arguments

The response provides no arguments to the instant rejection.

15. Claims 71, 73 are rejected under pre-AIA 35 U.S.C. 103(a) as being unpatentable over Porecca(Nature methods (2007) volume 4, pages 931-936).as applied to claim 68-69, 76 above, and further in view of Schiffman (Journal of Clinical Oncology (2007) volume 25, p530).

The teachings of Porecca are set forth above and teach the detection of homozygous alleles by use of unique targeting arms of MIPS and recognizing or collapsing the combination of unique target arms and alleles (or target sequences).

Porecca does not specifically teach determination of blood samples or circulating tumor nucleic acids.

However, Schiffman teaches the use of MIPs for analysis of blood samples subjects with leukemia (leukemia is a circulating cancer or tumor).

Therefore it would have been prima facie obvious to one of ordinary skill in the art at the time the invention was made to use blood samples from subjects with leukemia for examination by MIPs. The artisan would be motivated to examine gene copy number changes and LOH in blood samples to characterize the leukemia. The artisan would have a reasonable expectation of success as the artisan is using known samples in know methods.

Response to Arguments

The response provides no arguments to the instant rejection.

Summary

No claims are allowed.

Conclusion

THIS ACTION IS MADE FINAL. Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to STEVEN C POHNERT PhD whose telephone number is (571)272-3803. The examiner can normally be reached on Monday- Friday about 6:00 AM-5:30 PM, every second Friday off.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Dave Nguyen can be reached on 571-272-0731. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see https://ppair-my.uspto.gov/pair/PrivatePair. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Steven Pohnert/
Primary Examiner, Art Unit 1634

Electronic Acknowledgement Receipt	
EFS ID:	42849210
Application Number:	16952764
International Application Number:	
Confirmation Number:	7812
Title of Invention:	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS
First Named Inventor/Applicant Name:	Gregory Porreca
Customer Number:	21710
Filer:	Thomas C. Meyers/Jeanne Waters
Filer Authorized By:	Thomas C. Meyers
Attorney Docket Number:	MLB-002/06US 34468/40
Receipt Date:	28-MAY-2021
Filing Date:	19-NOV-2020
Time Stamp:	14:12:54
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Assignee showing of ownership per 37 CFR 3.73	MLB-002-06US_373-Statement.pdf	156159 32e71deb7bc711ea67bc365d958e154b5608d095	no	2
Warnings:					

Information:					
2	Power of Attorney	MLB_POA.pdf	53693 7e5904fa9919c8522c3cb6a204d5df2e293e5698	no	1
Warnings:					
Information:					
			Total Files Size (in bytes):	209852	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

PTO/AIA/96 (08-12)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

STATEMENT UNDER 37 CFR 3.73(c)

Applicant/Patent Owner: Molecular Loop Biosolutions, LLC

Application No./Patent No.: 16/952,764 Filed/Issue Date: November 19, 2020

Titled: METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS

Molecular Loop Biosolutions, LLC , a Limited Liability Company
(Name of Assignee) (Type of Assignee, e.g., corporation, partnership, university, government agency, etc.)

states that, for the patent application/patent identified above, it is (choose **one** of options 1, 2, 3 or 4 below):

1. [x] The assignee of the entire right, title, and interest.

2. [ ] An assignee of less than the entire right, title, and interest (check applicable box):

   [ ] The extent (by percentage) of its ownership interest is _____ %. Additional Statement(s) by the owners holding the balance of the interest must be submitted to account for 100% of the ownership interest.

   [ ] There are unspecified percentages of ownership. The other parties, including inventors, who together own the entire right, title and interest are:

   Additional Statement(s) by the owner(s) holding the balance of the interest must be submitted to account for the entire right, title, and interest.

3. [ ] The assignee of an undivided interest in the entirety (a complete assignment from one of the joint inventors was made). The other parties, including inventors, who together own the entire right, title, and interest are:

   Additional Statement(s) by the owner(s) holding the balance of the interest must be submitted to account for the entire right, title, and interest.

4. [ ] The recipient, via a court proceeding or the like (*e.g.*, bankruptcy, probate), of an undivided interest in the entirety (a complete transfer of ownership interest was made). The certified document(s) showing the transfer is attached.

The interest identified in option 1, 2 or 3 above (not option 4) is evidenced by either (choose **one** of options A or B below):

A. [ ] An assignment from the inventor(s) of the patent application/patent identified above. The assignment was recorded in the United States Patent and Trademark Office at Reel _____ , Frame _____ , or for which a copy thereof is attached.

B. [x] A chain of title from the inventor(s), of the patent application/patent identified above, to the current assignee as follows:

1. From: Inventors To: Good Start Genetics, Inc.
   The document was recorded in the United States Patent and Trademark Office at
   Reel 054561 , Frame 0315 , or for which a copy thereof is attached.

2. From: Good Start Genetics, Inc. To: Molecular Loop Biosolutions, LLC
   The document was recorded in the United States Patent and Trademark Office at
   Reel 054561 , Frame 0341 , or for which a copy thereof is attached.

PTO/AIA/96 (08-12)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

STATEMENT UNDER 37 CFR 3.73(c)

3. From: _____ To: _____
The document was recorded in the United States Patent and Trademark Office at
Reel _____ , Frame _____ , or for which a copy thereof is attached.

4. From: _____ To: _____
The document was recorded in the United States Patent and Trademark Office at
Reel _____ , Frame _____ , or for which a copy thereof is attached.

5. From: _____ To: _____
The document was recorded in the United States Patent and Trademark Office at
Reel _____ , Frame _____ , or for which a copy thereof is attached.

6. From: _____ To: _____
The document was recorded in the United States Patent and Trademark Office at
Reel _____ , Frame _____ , or for which a copy thereof is attached.

☐ Additional documents in the chain of title are listed on a supplemental sheet(s).

☒ As required by 37 CFR 3.73(c)(1)(i), the documentary evidence of the chain of title from the original owner to the assignee was, or concurrently is being, submitted for recordation pursuant to 37 CFR 3.11.

[NOTE: A separate copy (i.e., a true copy of the original assignment document(s)) must be submitted to Assignment Division in accordance with 37 CFR Part 3, to record the assignment in the records of the USPTO. See MPEP 302.08]

The undersigned (whose title is supplied below) is authorized to act on behalf of the assignee.

/Thomas C. Meyers/	May 28, 2021
Signature	Date
Thomas C. Meyers	36,989
Printed or Typed Name	Title or Registration Number

64081982 v1-WorkSiteUS-034468/0040

PTO/AIA/80 (07-17)
Approved for use through 01/31/2018. OMB 0651-0035
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no person is required to respond to a collection of information unless it displays a valid OMB control number

POWER OF ATTORNEY TO PROSECUTE APPLICATIONS BEFORE THE USPTO

I hereby revoke all previous powers of attorney given in the application identified in the attached statement under 37 CFR 3.73(c).

I hereby appoint:

[X] Practitioners associated with Customer Number: 21710

OR

[ ] Practitioner(s) named below (if more than ten patent practitioners are to be named, then a customer number must be used):

Name	Registration Number	Name	Registration Number

As attorney(s) or agent(s) to represent the undersigned before the United States Patent and Trademark Office (USPTO) in connection with any and all patent applications assigned only to the undersigned according to the USPTO assignment records or assignment documents attached to this form in accordance with 37 CFR 3.73(c).

Please change the correspondence address for the application identified in the attached statement under 37 CFR 3.73(c) to:

[X] The address associated with Customer Number: 21710

OR

[ ]

Firm or individual name		
Address		
City	State	Zip
Country		
Telephone	Email	

Assignee name and address: Molecular Loop Biosolutions, LLC
57 Reservoir Street #1
Cambridge, MA 02138

A copy of this form, together with a statement under 37 CFR 3.73(c) (Form PTO/AIA/96 or equivalent) is required to be filed in each application in which this form is used. The statement under 37 CFR 3.73(c) may be completed by one of the practitioners appointed in this form, and must identify the application in which this Power of Attorney is to be filed.

SIGNATURE of Assignee of Record

The individual whose signature and title is supplied below is authorized to act on behalf of the assignee.

Signature	Date 11/6/2018
Name Gregory Porreca	Telephone 609 457 4431
Title Managing Member and CEO	

This collection of information is required by 37 CFR 1.31, 1.32, and 1.33. The information is required to obtain or retain a benefit by the public, which is to update (and by the USPTO to process) the file of a patent or reexamination proceeding. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 18 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
16/952,764	11/19/2020	Gregory Porreca	MLB-002/06US 34468/40

CONFIRMATION NO. 7812

21710
BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

POA ACCEPTANCE LETTER

OC000000126000534

Date Mailed: 06/02/2021

NOTICE OF ACCEPTANCE OF POWER OF ATTORNEY

This is in response to the Power of Attorney filed 05/28/2021.

The Power of Attorney in this application is accepted. Correspondence in this application will be mailed to the above address as provided by 37 CFR 1.33.

Questions about the contents of this notice and the requirements it sets forth should be directed to the Office of Data Management, Application Assistance Unit, at **(571) 272-4000** or **(571) 272-4200** or **1-888-786-0101**.

/agizaw/

Electronic Acknowledgement Receipt	
EFS ID:	43270668
Application Number:	16952764
International Application Number:	
Confirmation Number:	7812
Title of Invention:	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS
First Named Inventor/Applicant Name:	Gregory Porreca
Customer Number:	21710
Filer:	Thomas C. Meyers/Jeanne Waters
Filer Authorized By:	Thomas C. Meyers
Attorney Docket Number:	MLB-002/06US 34468/40
Receipt Date:	16-JUL-2021
Filing Date:	19-NOV-2020
Time Stamp:	16:18:06
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	DA
Payment was successfully received in RAM	$530
RAM confirmation Number	E20217FG18293201
Deposit Account	
Authorized User	
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Pre-Brief Conference request	MLB-002-06US_Appeal-Remarks.pdf	172702 ba2b64507131618155 7b8d4e251ade2903138908	no	5
Warnings:					
Information:					
2	Pre-Brief Conference request	MLB-002-06US_Pre-appeal-form.pdf	223918 9e9c7e100deb3d5aea92db000db1d1d9bce9d83f	no	2
Warnings:					
Information:					
3	Notice of Appeal Filed	MLB-002-06US_Notice-of-Appeal-form.pdf	256730 1a06003e84d52daa7a7ee9c2752fd7fe77e727ee	no	2
Warnings:					
Information:					
4	Extension of Time	MLB-002-06US_EOT.pdf	157517 4bc3e80a89c817c659927e0ea74ed3fabd73be96	no	1
Warnings:					
Information:					
5	Fee Worksheet (SB06)	fee-info.pdf	32597 0455dd6961d3b03871d9e529c1bfcdf33a9def39	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	843464		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

PRE-APPEAL REVIEW REQUEST

The pending claims in this U.S. Application No. 16/952,764 were rejected on February 10, 2021, and April 9, 2021, and the application is eligible for appeal. Applicant requests pre-appeal review. The pending claims were originally filed on October 2, 2020, in *parent* Application No. 15/231,687 ('687 parent) with remarks that both showed the claims to be patentable and asked that due weight be given to a March 12, 2020, 37 C.F.R. 1.132 expert declaration by Dr. Gregory Porreca. As show in (i) the October 2, 2020, remarks and the March 12, 2020, expert declaration in the '687 parent, and (ii) in the April 9, 2021, remarks and a March 5, 2021, expert declaration in this application, the written record provides substantial evidence that the claims are patentable.

The Examiner in this matter has continually abused his authority, ignored clear scientific evidence offered by a recognized expert in the field, and failed to properly apply the relevant laws and implementing regulations, improperly refusing to allow the claims. Examiner Pohnert has repeatedly suggested specific actions to the Applicant regarding the claims both in this application and the'687 parent. Despite Applicant's compliance with the Examiner's various suggestions, Examiner Pohnert continues to reject claims over his apparent disagreement with technical issues that are clearly addressed by Dr. Porreca in the 2021 expert declaration. Moreover, Examiner Pohnert has repeatedly failed to properly substantiate his rejections under 35 U.S.C. §§112, 101, and 103, most recently in the February 10, 2021, Office Action. In an effort to avoid bringing these clear errors to the Patent Trial and Appeal Board, Applicant requests this review and allowance of the pending claims.

1. <u>The Examiner made obvious factual errors in rejecting the claims under 35 USC §112</u>

In the April 9, 2021, Office Action, at pages 6-11, Examiner Pohnert rejected claims 68-69, 71, 73-74, and 76 under 35 U.S.C. §112(a) as lacking written description and under §112(b) as being indefinite. Those rejections have no basis in fact or law.

With regard to claim 68, Examiner Pohnert stated that the specification "provides no guidance on how to perform collapsing, why collapsing is done or how collapsing corrects errors or bias." Explicit guidance is provided at page 6, lines 21 through page 7, line 4 of the as-filed Application. With regard to claim 74, Examiner Pohnert stated that the specification provides "no guidance on how to determine a copy number of genomic regions or transcript based on collapsing reads of target:differentiator tag combinations." Again, explicit guidance is provided at page 79, lines 8-15. With regard to claim 76, Examiner Pohnert stated that the specification "provides no guidance on how to perform recognizing, how recognizing is done or how recognizing corrects errors or bias." However, Applicant notes that the terms "collapsing reads" to correct errors or bias, "determining the copy number of a genomic region" from collapsed reads, and "recognizing target/differentiator tags having the same sequence as being derived from the same input molecule" are **verbatim** disclosed and fully supported in the as-filed specification. There is a clear explanation of how the invention works at, for example, pages 6, 7, 27, 79 and in examples 2 and 3 and Figures 5 and 6. At least those indicated passages provide verbatim support and enablement for the claim terms. See also Remarks and Table of support in Applicant's response dated March 5, 2021, in the present application.

Examiner Pohnert's assertion that the specification provides "no guidance" for the rejected claim terms is in clear error and is contrary to the explicit words of the specification.

Additionally, the Examiner erred by refusing to give due weight to the technical opinions and factual findings of Dr. Porreca provided his expert declaration filed with the response dated March 5, 2021. Dr. Porreca's declaration represents the understanding of the cited art and the instant application by a recognized expert in the field. Dr. Porreca's declaration, at paragraphs 4 and 6, explains in detail how a person of skill in the art would understand that the as-filed specification teaches collapsing target:differentiator tag combinations to correct for errors and bias. The Federal Circuit has made clear that 35 U.S.C. 112(b) requires only that "the claims read in light of the specification, reasonably apprise those skilled in the art and are as precise as the

subject matter permits. As a matter of law, no court can demand more." *Hybritech, Inc. v. Monoclonal Antibodies, Inc.*, 802 F.2d 1367 (Fed. Cir. 1986) (emphasis added). The Examiner's dismissal of Dr. Porreca's factual assertions regarding the specification is another clear error. If the Examiner disagrees with Dr. Porreca's expert conclusions, he must provide documentary evidence to substantiate the rebuttal, or, if the Examiner is a subject matter expert, he should provide an affidavit providing evidence of expertise and the technical basis for the conclusions made. *In re Kahn*, 441 F.3d 997, 998 (Fed. Cir. 2006). It is absurd, to say the least, for the Examiner to ignore or, worse, contradict the conclusions of an expert in the field without any evidence to support his position.

2. The Examiner made key errors in rejecting the claims under 35 U.S.C. §101

In the Office Action mailed February 10, 2021, at pages 7-10, Examiner Pohnert rejected claims 68-69, 71, 73-74, and 76 under 35 U.S.C. §101 as being directed to a mental step.

Examiner Pohnert has again committed a clear error in making this rejection. The claimed invention is directed to methods comprising tagging individual nucleic acid molecules with a differentiator tag prior to amplification and sequencing steps. Each of the claims require multiple physical laboratory steps to tag nucleic acid molecules with differentiator tags, amplify the tagged molecules, and sequence the amplicons. The step of collapsing tag combinations goes further and improves the workflow by correcting for any error or bias that is introduced. There is simply no way in which the steps of the claims could be carried out absent physical manipulation in the laboratory. One wonders how amplifying and/or sequencing could ever be considered a mental process. Constructing a building provides an analogy. If one were to merely "think" of constructing the building, the building would not be built. The process requires multiple physical steps in the natural world. Similarly, Applicant has claimed actual physical laboratory procedures, none of which is met by a mental step.

3. The Examiner made factual and legal errors in rejecting the claims under 35 U.S.C. §103

In the February 10 Office Action, on pages 10-16, Examiner Pohnert rejected claims 68-69 and 76 under 35 U.S.C. §103 over Porreca 2007; claim 74 over Porreca 2007 in view of Wang; and claims 71 and 73 over Porreca 2007 in view of Schiffman.

In making those rejections, Examiner Pohnert has again made clear factual and legal errors. The Examiner has ignored the factual statements made in the present application, the prior art, and in Dr. Porreca's declaration. Tellingly, the rejection of claims 68-69 and 76 under §103 is based on a single reference that Examiner Pohnert admits fails to disclose an element of the claims. Not only does the Examiner misinterpret the technical detail, he ignores legal standard for a prima facie case for obviousness. In order to establish a prima facie case of obviousness, the prior art must provide motivation to modify a reference or to combine multiple references in a manner that would lead one of ordinary skill in the art to an applicant's claimed invention with a reasonable expectation of success. *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398 (2007).

On page 12 of the Office Action, Examiner Pohnert concedes that Porreca 2007 does not teach "correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count." Examiner Pohnert does not, and cannot, identify any motivation provided by Porreca 2007 to modify the teachings of that reference itself to lead to the claimed invention. Moreover, Examiner Pohnert has not provided evidence or a reasoned explanation as to why common sense would lead once to modify Porreca 2007 to provide the missing limitation, when **Porreca 2007 itself recognized that PCR amplification steps were a source of nonuniformity error but did not suggest such a solution**. *Porreca 2007, pg. 4, Discussion.*

Moreover, the March 5, 2021, expert declaration of Dr. Gregory Porreca gives record evidence that the claims are not obvious. In that declaration, Dr. Porreca—inventor of the instant claims and first author of Porreca 2007—shows that none of the cited references, including his 2007 paper, disclose a method comprising tagging individual nucleic acid molecules with differentiator tags prior to amplification or sequencing steps. *Porreca dec., par. 6-12.*

Specifically, Dr. Porreca states that at the time of his 2007 publication, he had not yet envisioned tagging nucleic acid molecules with differentiator tags. *Id., par. 9*. Moreover, Dr. Porreca reports that the unique targeting arms of the molecular inversion probes disclosed in Porreca 2007, Wang, and Schiffman are not differentiator probes at all. *Id. pars. 8, 11, 12.* This is because each molecular inversion probe is designed to uniquely capture an exon of interest with targeting arms that hybridize to sequences flanking the exon. *Id., para. 7*. The targeting arms are unique only in the sense that PCR primers are unique; unique designs each synthesized in the millions (this point is settled where Supplementary Figure 1 in Porreca 2007 states that the 480 probes were used at a "concentration of ~167 pM for each species"). Because the probes are only "unique" to their target exons, the expert declaration concludes that their targeting arms cannot differentiate separate single molecules. *Id.* The Examiner has erred by ignoring that testimony from the author of Porreca 2007. In fact, none of Porreca 2007, Wang, and Schiffman teach or suggest on their face tags that differentiate separate individual molecules. Because that element is missing from **all of the cited references**, a conclusion of obviousness is unsupported.

As detailed herein, Examiner Pohnert has continually failed to properly substantiate his rejections under 35 U.S.C. §§112, 101, and 103. Applicant believes that Examiner Pohnert's conduct in examining this application and the '687 parent does not represent a fair and impartial examination, and Examiner Pohnert has abused his authority and discretion. In an effort to avoid bringing these issues to the Patent Trial and Appeal Board, Applicant requests review by this panel, and a finding that the application is allowed on the existing claims and that prosecution remains closed.

Dated: July 16, 2021

Respectfully submitted,
BROWN RUDNICK LLP

/Thomas C. Meyers/
Thomas C. Meyers, Reg. No. 36,989
Attorney for Applicant
Email: tmeyers@brownrudnick.com

BROWN RUDNICK LLP
One Financial Center
Boston, MA 02111
Tel: 617-856-8483

64116926 v1

Petitioner Illumina, Inc., EX1004, p. 1017

Doc Code: AP.PRE.REQ

PTO/AIA/33 (03-13)
Approved for use through 01/31/2021. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PRE-APPEAL BRIEF REQUEST FOR REVIEW

Docket Number (Optional): MLB-002/06US 34468/40

I hereby certify that this correspondence is being facsimile transmitted to the USPTO, EFS-Web transmitted to the USPTO, or deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to "Mail Stop AF, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450" [37 CFR 1.8(a)]
on July 16, 2021

Signature /Aditya Ramanathan/

Typed or printed name Aditya Ramanathan

Application Number	Filed
16/952,764	November 19, 2020

First Named Inventor: Gregory Porreca

Art Unit	Examiner
1634	Steven C. Pohnert

Applicant requests review of the final rejection in the above-identified application. No amendments are being filed with this request.

This request is being filed with a notice of appeal.

The review is requested for the reason(s) stated on the attached sheet(s).
Note: No more than five (5) pages may be provided.

I am the

☐ applicant.

☑ attorney or agent of record. Registration number 36,989.

☐ attorney or agent acting under 37 CFR 1.34.
Registration number if acting under 37 CFR 1.34 _____

/Thomas C. Meyers/
Signature

Thomas C. Meyers
Typed or printed name

(617) 856-8483
Telephone number

July 16, 2021
Date

NOTE: This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4 for signature requirements and certifications. Submit multiple forms if more than one signature is required, see below*.

☑ *Total of 1 forms are submitted.

This collection of information is required by 35 U.S.C. 132. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11, 1.14 and 41.6. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Mail Stop AF, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

Privacy Act Statement

The **Privacy Act of 1974 (P.L. 93-579)** requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether disclosure of these records is required by the Freedom of Information Act.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (*i.e.*, GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspection or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

PTO/AIA/31 (03-14)
Approved for use through 12/31/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

NOTICE OF APPEAL FROM THE EXAMINER TO THE PATENT TRIAL AND APPEAL BOARD

Docket Number (Optional): MLB-002/06US 34468/40

I hereby certify that this correspondence is being facsimile transmitted to the USPTO, EFS-Web transmitted to the USPTO, or deposited with the United States Postal Service with sufficient postage in an envelope addressed to "Commissioner for Patents, P.O. Box 1450, Alexandria, on Alexandria, VA 22313-1450" [37 CFR 1.8(a)] on July 16, 2021.

Signature /Aditya Ramanathan/

Typed or printed name Aditya Ramanathan

In re Application of: Molecular Loop Biosolutions, LLC

Application Number	Filed
16/952,764	November 19, 2020

For: METHOD AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS

Art Unit	Examiner
1634	Steven C. Pohnert

Applicant hereby **appeals** to the Patent Trial and Appeal Board from the last decision of the examiner.

The fee for this Notice of Appeal is (37 CFR 41.20(b)(1)) $ 840.00

☑ Applicant asserts small entity status. See 37 CFR 1.27. Therefore, the fee shown above is reduced by 50%, and the resulting fee is: $ 420.00

☐ Applicant certifies micro entity status. See 37 CFR 1.29. Therefore, the fee shown above is reduced by 75%, and the resulting fee is: $ _____
Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously.

☐ A check in the amount of the fee is enclosed.

☐ Payment by credit card. Form PTO-2038 is attached.

☑ The Director is hereby authorized to charge any fees which may be required, or credit any overpayment to Deposit Account No. 500369.

☑ Payment made via EFS-Web.

☑ A petition for an extension of time under 37 CFR 1.136(a) (PTO/AIA/22 or equivalent) is enclosed.
For extensions of time in reexamination proceedings, see 37 CFR 1.550.

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

I am the

☐ applicant

☑ attorney or agent of record, Registration number 36,989

☐ attorney or agent acting under 37 CFR 1.34, Registration number _____

Signature /Thomas C. Meyers/

Typed or printed name Thomas C. Meyers

Telephone Number (617)856-8483

Date July 16, 2021

NOTE: This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4 for signature requirements and certifications. Submit multiple forms if more than one signature is required, see below*.

☑ * Total of 1 forms are submitted.

This collection of information is required by 37 CFR 41.20(b)(1) and 41.31. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11, 1.14 and 41.6. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

Privacy Act Statement

The **Privacy Act of 1974 (P.L. 93-579)** requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether disclosure of these records is required by the Freedom of Information Act.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (*i.e.*, GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspection or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

PTO/AIA/22 (10-20)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PETITION FOR EXTENSION OF TIME UNDER 37 CFR 1.136(a)	Docket Number (Optional) MLB-002/06US
Application Number 16/952,764-Conf. #7812	Filed November 19, 2020
For METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS	
Art Unit 1634	Examiner S. C. Pohnert

This is a request under the provisions of 37 CFR 1.136(a) to extend the period for filing a reply in the above-identified application.

The requested extension and fee are as follows (check time period desired and enter the appropriate fee below):

	Fee	Small Entity Fee	Micro Entity Fee	
[X] One month (37 CFR 1.17(a)(1))	$220	$110	$55	$ 110.00
[ ] Two months (37 CFR 1.17(a)(2))	$640	$320	$160	$
[ ] Three months (37 CFR 1.17(a)(3))	$1,480	$740	$370	$
[ ] Four months (37 CFR 1.17(a)(4))	$2,320	$1,160	$580	$
[ ] Five months (37 CFR 1.17(a)(5))	$3,160	$1,580	$790	$

[X] Applicant asserts small entity status. See 37 CFR 1.27.

[ ] Applicant certifies micro entity status. See 37 CFR 1.29.
Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously.

[ ] A check in the amount of the fee is enclosed.

[ ] Payment by credit card. Form PTO-2038 is attached.

[X] The Director has already been authorized to charge fees in this application to a Deposit Account.

[X] The Director is hereby authorized to charge any fees which may be required, or credit any overpayment, to Deposit Account Number 50-0369 .

[X] Payment made via EFS-Web.

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

I am the

[ ] applicant.

[X] attorney or agent of record. Registration number 36,989 .

[ ] attorney or agent acting under 37 CFR 1.34. Registration number _____ .

/Thomas C. Meyers/	July 16, 2021
Signature	Date
Thomas C. Meyers	(617) 856-8200
Typed or printed name	Telephone Number

NOTE: This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4 for signature requirements and certifications. Submit multiple forms if more than one signature is required, see below*.

[ ] * Total of 1 forms are submitted.

64117950 v1-WorkSiteUS-034468/0040

Electronic Patent Application Fee Transmittal

Application Number:	16952764
Filing Date:	19-Nov-2020
Title of Invention:	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS
First Named Inventor/Applicant Name:	Gregory Porreca
Filer:	Thomas C. Meyers/Jeanne Waters
Attorney Docket Number:	MLB-002/06US 34468/40

Filed as Small Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
NOTICE OF APPEAL	2401	1	420	420
Post-Allowance-and-Post-Issuance:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Extension-of-Time:				
Extension - 1 month with $0 paid	2251	1	110	110
Miscellaneous:				
		Total in USD ($)		**530**

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/952,764	11/19/2020	Gregory Porreca	MLB-002/06US 34468/40	7812

21710 7590 09/10/2021

BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

EXAMINER	
POHNERT, STEVEN C	

ART UNIT	PAPER NUMBER
1634	

NOTIFICATION DATE	DELIVERY MODE
09/10/2021	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

ip@brownrudnick.com
usactions@brownrudnick.com

Notice of Panel Decision from Pre-Appeal Brief Review	Application No. 16/952,764	Applicant(s) Porreca et al.	
	Examiner DAVE T NGUYEN	Art Unit 1634	AIA (FITF) Status No

This is in response to the Pre-Appeal Brief Request for Review filed 16 July 2021 .

1. ☐ **Improper Request** - The Request is improper and a conference will not be held for the following reason(s):

☐ The Notice of Appeal has not been filed concurrent with the Pre-Appeal Brief Request.
☐ The request does not include reasons why a review is appropriate.
☐ A proposed amendment is included with the Pre-Appeal Brief request.
☐ Other: _____ .

The time period for filing a response continues to run from the receipt date of the Notice of Appeal or from the mail date of the last Office communication, if no Notice of Appeal has been received.

2. ☑ **Proceed to Board of Patent Appeals and Interferences** - A Pre-Appeal Brief conference has been held. The application remains under appeal because there is at least one actual issue for appeal. Applicant is required to submit an appeal brief in accordance with 37 CFR 41.37. The time period for filing an appeal brief will be reset to be one month from mailing this decision, or the balance of the two-month time period running from the receipt of the notice of appeal, whichever is greater. Further, the time period for filing of the appeal brief is extendible under 37 CFR 1.136 based upon the mail date of this decision or the receipt date of the notice of appeal, as applicable.

☐ The panel has determined the status of the claim(s) is as follows:
Claim(s) allowed: _____ .
Claim(s) objected to: _____ .
Claim(s) rejected: _____ .
Claim(s) withdrawn from consideration: _____ .

3. ☐ **Allowable application** - A conference has been held. The rejection is withdrawn and a Notice of Allowance will be mailed. Prosecution on the merits remains closed. No further action is required by applicant at this time.

4. ☐ **Reopen Prosecution** - A conference has been held. The rejection is withdrawn and a new Office action will be mailed. No further action is required by applicant at this time.

All participants:

(1) DAVE T NGUYEN .	(3) STEVE POHNERT .
(2) ASHWIN MEHTA .	(4) _____ .

/DAVE T NGUYEN/ Supervisory Patent Examiner, Art Unit 1634		

Electronic Acknowledgement Receipt

EFS ID:	43917948
Application Number:	16952764
International Application Number:	
Confirmation Number:	7812
Title of Invention:	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS
First Named Inventor/Applicant Name:	Gregory Porreca
Customer Number:	21710
Filer:	Thomas C. Meyers/Kelley Warren
Filer Authorized By:	Thomas C. Meyers
Attorney Docket Number:	MLB-002/06US 34468/40
Receipt Date:	30-SEP-2021
Filing Date:	19-NOV-2020
Time Stamp:	17:47:52
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Application Data Sheet	Application_Data_Sheet.pdf	58027 f83942176d9f834f805fef9d5d56685e1dc196a8	no	4

Warnings:

Information:					
This is not an USPTO supplied ADS fillable form					
2	Request for Corrected Filing Receipt	Request_for_Corrected_Filing_Receipt.pdf	20434 0dcc88bd7be00697be322ee550b3d15631ddc164	no	2
Warnings:					
Information:					
3	Assignee showing of ownership per 37 CFR 3.73	Statement_By_Assignee_to_Establish_Ownership_37_CFR_373c.pdf	29511 e1c766087a6327079112e1121ce0cef8dde6cd21	no	2
Warnings:					
Information:					
4	Power of Attorney	Molecular-Loop-Inc_POA-USPTO_002.pdf	150293 6c830a7bd5bbb04655b4d98d19fe06d52a166a14	no	1
Warnings:					
Information:					
		Total Files Size (in bytes):	258265		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Corrected Application Data Sheet

Inventor Information

Inventor Number::	1
Given Name::	Gregory
Family Name::	Porreca
City of Residence::	Cambridge
State or Province of Residence::	MA
Country of Residence::	US
Street of mailing address::	57 Reservoir Street
City of mailing address::	Cambridge
State or Province of mailing address::	MA
Country of mailing address::	US
Postal or Zip Code of mailing address::	02138

Inventor Number::	2
Given Name::	Uri
Family Name::	Laserson
City of Residence::	Boston
State or Province of Residence::	MA
Country of Residence::	US
Street of mailing address::	645 Beacon Street
	#215
City of mailing address::	Boston

State or Province of mailing address::	MA
Country of mailing address::	US
Postal or Zip Code of mailing address::	02215

Correspondence Information

Correspondence Customer Number::	21710
E-Mail address::	ip@brownrudnick.com

Application Information

Application Number::	16/952,764
Filing Date::	11/19/20
Application Type::	Nonprovisional
Subject Matter::	Utility
CD-ROM or CD-R?::	None
Sequence submission?::	None
Computer Readable Form (CRF)?::	No
Title::	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS
Attorney Docket Number::	MLB-002/06 34468/40US
Request for Early Publication?::	No
Request for Non-Publication?::	Yes
Total Drawing Sheets::	16
Small Entity?::	Yes

Petition included?::	No
Portions or all of the application associated with this Application Data Sheet may fall under a Secrecy Order pursuant to 37 CFR 5.2::	No
This application (1) claims priority to or the benefit of an application filed before March 16, 2013 and (2) also contains, or contained at any time, a claim to a claimed invention that has an effective filing date on or after March 16, 2013::	No

Representative Information

Representative Customer Number::	21710

Domestic Priority Information

Application::	Continuity Type::	Parent Application::	Parent Filing Date::	Prior Appl Status::
This Application	Continuation of	15/231687	08/08/16	Pending
15/231687	Continuation of	13/266862	03/13/12	Pending
13/266862	a 371 of international	US10/01293	04/30/10	
US10/01293	Claims benefit of provisional	61/182089	05/28/09	Expired
US10/01293	Claims benefit of provisional	61/179358	05/18/09	Expired
US10/01293	Claims benefit of provisional	61/178923	05/15/09	Expired
US10/01293	Claims benefit of provisional	61/174470	04/30/09	Expired

Petitioner Illumina, Inc., EX1004, p. 1031

Foreign Priority Information

Applicant Information

Applicant Number::	1
Applicant Type::	Assignee
Organization Name::	~~Molecular Loop Biosolutions, LLC~~ Molecular Loop Biosciences, Inc.
Street of mailing address::	57 Reservoir Street #1
City of mailing address::	Cambridge
State or Province of mailing address::	MA
Country of mailing address::	US
Postal or Zip Code of mailing address::	02138

Signature:

NOTE: This Application Data Sheet must be signed in accordance with 37 CFR 1.33(b). **However, if this Application Data Sheet is submitted with the INITIAL filing of the application and either box A or B is not checked in subsection 2 of the "Authorization or Opt-Out of Authorization to Permit Access" section, then this form must also be signed in accordance with 37 CFR 1.14(c).**

This Application Data Sheet **must** be signed by a patent practitioner if one or more of the applicants is a **juristic entity** (e.g., corporation or association). If the applicant is two or more joint inventors, this form must be signed by a patent practitioner, **all** joint inventors who are the applicant, or one or more joint inventor-applicants who have been given power of attorney (e.g., see USPTO Form PTO/AIA/81) on behalf of **all** joint inventor-applicants.

See 37 CFR 1.4(d) for the manner of making signatures and certifications.

Signature	/Thomas C. Meyers/	Date (YYYY-MM-DD)	2021-09-30
Name	Thomas C. Meyers	Registration Number	36,989

64166127 v1-WorkSiteUS-034468/0040

Petitioner Illumina, Inc., EX1004, p. 1032

Docket No.: MLB-002/06US
34468/40
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of:
Molecular Loop Biosolutions, LLC, Good Start Genetics, Inc., and Molecular Loop Biosciences, Inc.

Application No.: 16/952,764 Confirmation No.: 7812

Filed: November 19, 2020 Art Unit: 1634

For: METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS Examiner: S. C. Pohnert

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

REQUEST FOR CORRECTED FILING RECEIPT

Applicant herewith submits a request to update the name of the applicant for the above identified application from Molecular Loop Biosolutions, LLC to Molecular Loop Biosciences, Inc. Accompanying this request is an updated application data sheet marked to reflect the updated applicant name and a Statement 3.73(c) showing ownership by Molecular Loop Biosciences, Inc.

Applicant additionally requests issuance of an updated Filing Receipt reflecting the updated applicant name, and requests that all pertinent U.S. Patent and Trademark Office records relating to the subject application be changed to reflect this correction.

64166581

Applicant believes no fee is due with this request.

Dated: September 30, 2021

Respectfully submitted,

Electronic signature: /Thomas C. Meyers/
Thomas C. Meyers
Registration No.: 36,989
BROWN RUDNICK LLP
Brown Rudnick LLP
One Financial Center
Boston, Massachusetts 02111
(617) 856-8200
Attorney for Applicant

64166581 v1-WorkSiteUS-034468/0040

PTO/AIA/96 (08-12)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

STATEMENT UNDER 37 CFR 3.73(c)

Applicant/Patent Owner: Molecular Loop Biosciences, Inc.

Application No./Patent No.: 16/952,764 Filed/Issue Date: November 19, 2020

Titled: METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS

Molecular Loop Biosciences, Inc. , a Limited Liability Company
(Name of Assignee) (Type of Assignee, e.g., corporation, partnership, university, government agency, etc.)

states that, for the patent application/patent identified above, it is (choose **one** of options 1, 2, 3 or 4 below):

1. [X] The assignee of the entire right, title, and interest.

2. [ ] An assignee of less than the entire right, title, and interest (check applicable box):

   [ ] The extent (by percentage) of its ownership interest is _____ %. Additional Statement(s) by the owners holding the balance of the interest must be submitted to account for 100% of the ownership interest.

   [ ] There are unspecified percentages of ownership. The other parties, including inventors, who together own the entire right, title and interest are:

   Additional Statement(s) by the owner(s) holding the balance of the interest must be submitted to account for the entire right, title, and interest.

3. [ ] The assignee of an undivided interest in the entirety (a complete assignment from one of the joint inventors was made). The other parties, including inventors, who together own the entire right, title, and interest are:

   Additional Statement(s) by the owner(s) holding the balance of the interest must be submitted to account for the entire right, title, and interest.

4. [ ] The recipient, via a court proceeding or the like (*e.g.*, bankruptcy, probate), of an undivided interest in the entirety (a complete transfer of ownership interest was made). The certified document(s) showing the transfer is attached.

The interest identified in option 1, 2 or 3 above (not option 4) is evidenced by either (choose **one** of options A or B below):

A. [ ] An assignment from the inventor(s) of the patent application/patent identified above. The assignment was recorded in the United States Patent and Trademark Office at Reel _____ , Frame _____ , or for which a copy thereof is attached.

B. [X] A chain of title from the inventor(s), of the patent application/patent identified above, to the current assignee as follows:

1. From: Gregory Porreca To: Good Start Genetics, Inc.
   The document was recorded in the United States Patent and Trademark Office at
   Reel 054561 , Frame 0315 , or for which a copy thereof is attached.

2. From: Uri Laserson To: Good Start Genetics, Inc.
   The document was recorded in the United States Patent and Trademark Office at
   Reel 054561 , Frame 0315 , or for which a copy thereof is attached.

64165959

Petitioner Illumina, Inc., EX1004, p. 1035

PTO/AIA/96 (08-12)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

STATEMENT UNDER 37 CFR 3.73(c)

3. From: Good Start Genetics, Inc. To: Molecular Loop Biosolutions, LLC
The document was recorded in the United States Patent and Trademark Office at
Reel 054561 , Frame 0341 , or for which a copy thereof is attached.

4. From: Molecular Loop Biosolutions, LLC To: Molecular Loop Biosciences, Inc.
The document was recorded in the United States Patent and Trademark Office at
Reel 057634 , Frame 0714 , or for which a copy thereof is attached.

5. From: _____ To: _____
The document was recorded in the United States Patent and Trademark Office at
Reel _____ , Frame _____ , or for which a copy thereof is attached.

6. From: _____ To: _____
The document was recorded in the United States Patent and Trademark Office at
Reel _____ , Frame _____ , or for which a copy thereof is attached.

☐ Additional documents in the chain of title are listed on a supplemental sheet(s).

☒ As required by 37 CFR 3.73(c)(1)(i), the documentary evidence of the chain of title from the original owner to the assignee was, or concurrently is being, submitted for recordation pursuant to 37 CFR 3.11.

[NOTE: A separate copy (i.e., a true copy of the original assignment document(s)) must be submitted to Assignment Division in accordance with 37 CFR Part 3, to record the assignment in the records of the USPTO. See MPEP 302.08]

The undersigned (whose title is supplied below) is authorized to act on behalf of the assignee.

/Thomas C. Meyers/	September 30, 2021
Signature	Date
Thomas C. Meyers	36,989
Printed or Typed Name	Title or Registration Number

64165959 v1-WorkSiteUS-034468/0040

PTO/AIA/80 (07-17)
Approved for use through 03/31/2021. OMB 0651-0035
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no person is required to respond to a collection of information unless it displays a valid OMB control number

POWER OF ATTORNEY TO PROSECUTE APPLICATIONS BEFORE THE USPTO

I hereby revoke all previous powers of attorney given in the application identified in the attached statement under 37 CFR 3.73(c).

I hereby appoint:

[x] Practitioners associated with Customer Number: 21710

OR

[ ] Practitioner(s) named below (if more than ten patent practitioners are to be named, then a customer number must be used):

Name	Registration Number	Name	Registration Number

As attorney(s) or agent(s) to represent the undersigned before the United States Patent and Trademark Office (USPTO) in connection with any and all patent applications assigned only to the undersigned according to the USPTO assignment records or assignments documents attached to this form in accordance with 37 CFR 3.73(c).

Please change the correspondence address for the application identified in the attached statement under 37 CFR 3.73(c) to:

[ ] The address associated with Customer Number:

OR

[ ]

Firm or individual name		
Address		
City	State	Zip
Country		
Telephone	Email	

Assignee name and address:

Molecular Loop Biosciences, Inc.

A copy of this form, together with a statement under 37 CFR 3.73(c) (Form PTO/AIA/96 or equivalent) is required to be filed in each application in which this form is used. The statement under 37 CFR 3.73(c) may be completed by one of the practitioners appointed in this form, and must identify the application in which this Power of Attorney is to be filed.

SIGNATURE of Assignee of Record

The individual whose signature and title is supplied below is authorized to act on behalf of the assignee.

Signature	Date **June 22, 2021**
Name **Gregory Porreca**	Telephone
Title **President and CEO**	

64069697

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
16/952,764	11/19/2020	Gregory Porreca	MLB-002/06US 34468/40

CONFIRMATION NO. 7812

21710
BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

POA ACCEPTANCE LETTER

OC000000128856963

Date Mailed: 10/05/2021

NOTICE OF ACCEPTANCE OF POWER OF ATTORNEY

This is in response to the Power of Attorney filed 09/30/2021.

The Power of Attorney in this application is accepted. Correspondence in this application will be mailed to the above address as provided by 37 CFR 1.33.

Questions about the contents of this notice and the requirements it sets forth should be directed to the Office of Data Management, Application Assistance Unit, at **(571) 272-4000** or **(571) 272-4200** or **1-888-786-0101**.

/byemane/

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING or 371(c) DATE	GRP ART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
16/952,764	11/19/2020	1634	830	MLB-002/06US 34468/40	9	2

CONFIRMATION NO. 7812

21710
BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

CORRECTED FILING RECEIPT



Date Mailed: 10/05/2021

Receipt is acknowledged of this non-provisional utility patent application. The application will be taken up for examination in due course. Applicant will be notified as to the results of the examination. Any correspondence concerning the application must include the following identification information: the U.S. APPLICATION NUMBER, FILING DATE, NAME OF FIRST INVENTOR, and TITLE OF INVENTION. Fees transmitted by check or draft are subject to collection.

Please verify the accuracy of the data presented on this receipt. If an error is noted on this Filing Receipt, please submit a written request for a corrected Filing Receipt, including a properly marked-up ADS showing the changes with strike-through for deletions and underlining for additions. If you received a "Notice to File Missing Parts" or other Notice requiring a response for this application, please submit any request for correction to this Filing Receipt with your reply to the Notice. When the USPTO processes the reply to the Notice, the USPTO will generate another Filing Receipt incorporating the requested corrections provided that the request is grantable.

Inventor(s)

Gregory Porreca, Cambridge, MA;
Uri Laserson, Boston, MA;

Applicant(s)

Molecular Loop Biosolutions, Inc., Cambridge, MA;

Power of Attorney: The patent practitioners associated with Customer Number 21710

Domestic Priority data as claimed by applicant

This application is a CON of 15/231,687 08/08/2016 ABN
which is a CON of 13/266,862 03/13/2012
which is a 371 of PCT/US10/01293 04/30/2010
which claims benefit of 61/182,089 05/28/2009
and claims benefit of 61/179,358 05/18/2009
and claims benefit of 61/178,923 05/15/2009
and claims benefit of 61/174,470 04/30/2009

Foreign Applications for which priority is claimed (You may be eligible to benefit from the **Patent Prosecution Highway** program at the USPTO. Please see http://www.uspto.gov for more information.) - None.
Foreign application information must be provided in an Application Data Sheet in order to constitute a claim to foreign priority. See 37 CFR 1.55 and 1.76.

Permission to Access Application via Priority Document Exchange: Yes

Permission to Access Search Results: Yes

Applicant may provide or rescind an authorization for access using Form PTO/SB/39 or Form PTO/SB/69 as appropriate.

If Required, Foreign Filing License Granted: 11/30/2020

The country code and number of your priority application, to be used for filing abroad under the Paris Convention, is **US 16/952,764**

Projected Publication Date: Request for Non-Publication Acknowledged

Non-Publication Request: Yes

**********WARNING**********

Applicant has requested non-publication, certifying that the invention disclosed in this application has not been and will not be the subject of an application filed in another country, or under a multilateral agreement, that requires publication at eighteen months after filing; **and has also claimed** the benefit of an application filed in another country or under a multilateral agreement. Because the non-publication request and corresponding certification appears to be inconsistent with the priority claim, applicant is reminded that he/she has an obligation to promptly notify the Office and **rescind the non-publication request** if the invention disclosed in this application is the subject of an application filed in another country, or under a multilateral agreement, that requires eighteen-month publication.

**********WARNING**********

Early Publication Request: No

**** SMALL ENTITY ****

Title

METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS

Preliminary Class

506

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications: No

PROTECTING YOUR INVENTION OUTSIDE THE UNITED STATES

Since the rights granted by a U.S. patent extend only throughout the territory of the United States and have no effect in a foreign country, an inventor who wishes patent protection in another country must apply for a patent in a specific country or in regional patent offices. Applicants may wish to consider the filing of an international application under the Patent Cooperation Treaty (PCT). An international (PCT) application generally has the same effect as a regular national patent application in each PCT-member country. The PCT process **simplifies** the filing of patent applications on the same invention in member countries, but **does not result** in a grant of "an international patent" and does not eliminate the need of applicants to file additional documents and fees in countries where patent protection is desired.

Almost every country has its own patent law, and a person desiring a patent in a particular country must make an application for patent in that country in accordance with its particular laws. Since the laws of many countries differ in various respects from the patent law of the United States, applicants are advised to seek guidance from specific foreign countries to ensure that patent rights are not lost prematurely.

Applicants also are advised that in the case of inventions made in the United States, the Director of the USPTO must issue a license before applicants can apply for a patent in a foreign country. The filing of a U.S. patent application

Petitioner Illumina, Inc., EX1004, p. 1040

serves as a request for a foreign filing license. The application's filing receipt contains further information and guidance as to the status of applicant's license for foreign filing.

Applicants may wish to consult the USPTO booklet, "General Information Concerning Patents" (specifically, the section entitled "Treaties and Foreign Patents") for more information on timeframes and deadlines for filing foreign patent applications. The guide is available either by contacting the USPTO Contact Center at 800-786-9199, or it can be viewed on the USPTO website at http://www.uspto.gov/web/offices/pac/doc/general/index.html.

For information on preventing theft of your intellectual property (patents, trademarks and copyrights), you may wish to consult the U.S. Government website, http://www.stopfakes.gov. Part of a Department of Commerce initiative, this website includes self-help "toolkits" giving innovators guidance on how to protect intellectual property in specific countries such as China, Korea and Mexico. For questions regarding patent enforcement issues, applicants may call the U.S. Government hotline at 1-866-999-HALT (1-866-999-4258).

LICENSE FOR FOREIGN FILING UNDER

Title 35, United States Code, Section 184

Title 37, Code of Federal Regulations, 5.11 & 5.15

GRANTED

The applicant has been granted a license under 35 U.S.C. 184, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" followed by a date appears on this form. Such licenses are issued in all applications where the conditions for issuance of a license have been met, regardless of whether or not a license may be required as set forth in 37 CFR 5.15. The scope and limitations of this license are set forth in 37 CFR 5.15(a) unless an earlier license has been issued under 37 CFR 5.15(b). The license is subject to revocation upon written notification. The date indicated is the effective date of the license, unless an earlier license of similar scope has been granted under 37 CFR 5.13 or 5.14.

This license is to be retained by the licensee and may be used at any time on or after the effective date thereof unless it is revoked. This license is automatically transferred to any related applications(s) filed under 37 CFR 1.53(d). This license is not retroactive.

The grant of a license does not in any way lessen the responsibility of a licensee for the security of the subject matter as imposed by any Government contract or the provisions of existing laws relating to espionage and the national security or the export of technical data. Licensees should apprise themselves of current regulations especially with respect to certain countries, of other agencies, particularly the Office of Defense Trade Controls, Department of State (with respect to Arms, Munitions and Implements of War (22 CFR 121-128)); the Bureau of Industry and Security, Department of Commerce (15 CFR parts 730-774); the Office of Foreign AssetsControl, Department of Treasury (31 CFR Parts 500+) and the Department of Energy.

NOT GRANTED

No license under 35 U.S.C. 184 has been granted at this time, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" DOES NOT appear on this form. Applicant may still petition for a license under 37 CFR 5.12, if a license is desired before the expiration of 6 months from the filing date of the application. If 6 months has lapsed from the filing date of this application and the licensee has not received any indication of a secrecy order under 35 U.S.C. 181, the licensee may foreign file the application pursuant to 37 CFR 5.15(b).

Petitioner Illumina, Inc., EX1004, p. 1041

SelectUSA

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The U.S. offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to promote and facilitate business investment. SelectUSA provides information assistance to the international investor community; serves as an ombudsman for existing and potential investors; advocates on behalf of U.S. cities, states, and regions competing for global investment; and counsels U.S. economic development organizations on investment attraction best practices. To learn more about why the United States is the best country in the world to develop technology, manufacture products, deliver services, and grow your business, visit http://www.SelectUSA.gov or call +1-202-482-6800.

Petitioner Illumina, Inc., EX1004, p. 1042

Electronic Acknowledgement Receipt

EFS ID:	43949938
Application Number:	16952764
International Application Number:	
Confirmation Number:	7812
Title of Invention:	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS
First Named Inventor/Applicant Name:	Gregory Porreca
Customer Number:	21710
Filer:	Thomas C. Meyers/Kelley Warren
Filer Authorized By:	Thomas C. Meyers
Attorney Docket Number:	MLB-002/06US 34468/40
Receipt Date:	05-OCT-2021
Filing Date:	19-NOV-2020
Time Stamp:	17:23:34
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Request for Corrected Filing Receipt	34468_40_Request_for_Corrected_Filing_Receipt.pdf	24914 e8e2485dbfa0d00266f5851c0e9f7b47c40f5bd5	no	2

Warnings:

Information:	
Total Files Size (in bytes):	24914

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.
National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.
New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Docket No.: MLB-002/06US
34468/40
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of:
Molecular Loop Biosolutions, LLC, Good Start Genetics, Inc., and Molecular Loop Biosciences, Inc.

Application No.: 16/952,764 Confirmation No.: 7812

Filed: November 19, 2020 Art Unit: 1634

For: METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS Examiner: S. C. Pohnert

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

REQUEST FOR CORRECTED FILING RECEIPT

Applicant herewith submits a request to update the name of the applicant for the above identified application from Molecular Loop Biosolutions, LLC to Molecular Loop Biosciences, Inc. Applicant submitted a Corrected Application Data Sheet reflecting this change on September 30, 2021, however, the Corrected Filing Receipt issued on October 5, 2021 did not reflect the requested change.

Applicant additionally requests issuance of an updated Filing Receipt reflecting the updated applicant name, and requests that all pertinent U.S. Patent and Trademark Office records relating to the subject application be changed to reflect this correction.

64166581

Applicant believes no fee is due with this request.

Dated: October 5, 2021

Respectfully submitted,

Electronic signature: /Thomas C. Meyers/
Thomas C. Meyers
Registration No.: 36,989
BROWN RUDNICK LLP
Brown Rudnick LLP
One Financial Center
Boston, Massachusetts 02111
(617) 856-8200
Attorney for Applicant

64166581 v1-WorkSiteUS-034468/0040

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING or 371(c) DATE	GRP ART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
16/952,764	11/19/2020	1634	830	MLB-002/06US 34468/40	9	2

CONFIRMATION NO. 7812

CORRECTED FILING RECEIPT

21710
BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

OC000000128948672

Date Mailed: 10/08/2021

Receipt is acknowledged of this non-provisional utility patent application. The application will be taken up for examination in due course. Applicant will be notified as to the results of the examination. Any correspondence concerning the application must include the following identification information: the U.S. APPLICATION NUMBER, FILING DATE, NAME OF FIRST INVENTOR, and TITLE OF INVENTION. Fees transmitted by check or draft are subject to collection.

Please verify the accuracy of the data presented on this receipt. If an error is noted on this Filing Receipt, please submit a written request for a corrected Filing Receipt, including a properly marked-up ADS showing the changes with strike-through for deletions and underlining for additions. If you received a "Notice to File Missing Parts" or other Notice requiring a response for this application, please submit any request for correction to this Filing Receipt with your reply to the Notice. When the USPTO processes the reply to the Notice, the USPTO will generate another Filing Receipt incorporating the requested corrections provided that the request is grantable.

Inventor(s)

Gregory Porreca, Cambridge, MA;
Uri Laserson, Boston, MA;

Applicant(s)

Molecular Loop Biosciences, Inc., Cambridge, MA;

Power of Attorney: The patent practitioners associated with Customer Number 21710

Domestic Priority data as claimed by applicant

This application is a CON of 15/231,687 08/08/2016 ABN
which is a CON of 13/266,862 03/13/2012
which is a 371 of PCT/US10/01293 04/30/2010
which claims benefit of 61/182,089 05/28/2009
and claims benefit of 61/179,358 05/18/2009
and claims benefit of 61/178,923 05/15/2009
and claims benefit of 61/174,470 04/30/2009

Foreign Applications for which priority is claimed (You may be eligible to benefit from the **Patent Prosecution Highway** program at the USPTO. Please see http://www.uspto.gov for more information.) - None.
Foreign application information must be provided in an Application Data Sheet in order to constitute a claim to foreign priority. See 37 CFR 1.55 and 1.76.

Permission to Access Application via Priority Document Exchange: Yes

Petitioner Illumina, Inc., EX1004, p. 1047

Permission to Access Search Results: Yes

Applicant may provide or rescind an authorization for access using Form PTO/SB/39 or Form PTO/SB/69 as appropriate.

If Required, Foreign Filing License Granted: 11/30/2020

The country code and number of your priority application, to be used for filing abroad under the Paris Convention, is **US 16/952,764**

Projected Publication Date: Request for Non-Publication Acknowledged

Non-Publication Request: Yes

**********WARNING**********

Applicant has requested non-publication, certifying that the invention disclosed in this application has not been and will not be the subject of an application filed in another country, or under a multilateral agreement, that requires publication at eighteen months after filing; **and has also claimed** the benefit of an application filed in another country or under a multilateral agreement. Because the non-publication request and corresponding certification appears to be inconsistent with the priority claim, applicant is reminded that he/she has an obligation to promptly notify the Office and **rescind the non-publication request** if the invention disclosed in this application is the subject of an application filed in another country, or under a multilateral agreement, that requires eighteen-month publication.

**********WARNING**********

Early Publication Request: No

**** SMALL ENTITY ****

Title

METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS

Preliminary Class

506

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications: No

PROTECTING YOUR INVENTION OUTSIDE THE UNITED STATES

Since the rights granted by a U.S. patent extend only throughout the territory of the United States and have no effect in a foreign country, an inventor who wishes patent protection in another country must apply for a patent in a specific country or in regional patent offices. Applicants may wish to consider the filing of an international application under the Patent Cooperation Treaty (PCT). An international (PCT) application generally has the same effect as a regular national patent application in each PCT-member country. The PCT process **simplifies** the filing of patent applications on the same invention in member countries, but **does not result** in a grant of "an international patent" and does not eliminate the need of applicants to file additional documents and fees in countries where patent protection is desired.

Almost every country has its own patent law, and a person desiring a patent in a particular country must make an application for patent in that country in accordance with its particular laws. Since the laws of many countries differ in various respects from the patent law of the United States, applicants are advised to seek guidance from specific foreign countries to ensure that patent rights are not lost prematurely.

Applicants also are advised that in the case of inventions made in the United States, the Director of the USPTO must issue a license before applicants can apply for a patent in a foreign country. The filing of a U.S. patent application

Petitioner Illumina, Inc., EX1004, p. 1048

serves as a request for a foreign filing license. The application's filing receipt contains further information and guidance as to the status of applicant's license for foreign filing.

Applicants may wish to consult the USPTO booklet, "General Information Concerning Patents" (specifically, the section entitled "Treaties and Foreign Patents") for more information on timeframes and deadlines for filing foreign patent applications. The guide is available either by contacting the USPTO Contact Center at 800-786-9199, or it can be viewed on the USPTO website at http://www.uspto.gov/web/offices/pac/doc/general/index.html.

For information on preventing theft of your intellectual property (patents, trademarks and copyrights), you may wish to consult the U.S. Government website, http://www.stopfakes.gov. Part of a Department of Commerce initiative, this website includes self-help "toolkits" giving innovators guidance on how to protect intellectual property in specific countries such as China, Korea and Mexico. For questions regarding patent enforcement issues, applicants may call the U.S. Government hotline at 1-866-999-HALT (1-866-999-4258).

LICENSE FOR FOREIGN FILING UNDER

Title 35, United States Code, Section 184

Title 37, Code of Federal Regulations, 5.11 & 5.15

GRANTED

The applicant has been granted a license under 35 U.S.C. 184, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" followed by a date appears on this form. Such licenses are issued in all applications where the conditions for issuance of a license have been met, regardless of whether or not a license may be required as set forth in 37 CFR 5.15. The scope and limitations of this license are set forth in 37 CFR 5.15(a) unless an earlier license has been issued under 37 CFR 5.15(b). The license is subject to revocation upon written notification. The date indicated is the effective date of the license, unless an earlier license of similar scope has been granted under 37 CFR 5.13 or 5.14.

This license is to be retained by the licensee and may be used at any time on or after the effective date thereof unless it is revoked. This license is automatically transferred to any related applications(s) filed under 37 CFR 1.53(d). This license is not retroactive.

The grant of a license does not in any way lessen the responsibility of a licensee for the security of the subject matter as imposed by any Government contract or the provisions of existing laws relating to espionage and the national security or the export of technical data. Licensees should apprise themselves of current regulations especially with respect to certain countries, of other agencies, particularly the Office of Defense Trade Controls, Department of State (with respect to Arms, Munitions and Implements of War (22 CFR 121-128)); the Bureau of Industry and Security, Department of Commerce (15 CFR parts 730-774); the Office of Foreign AssetsControl, Department of Treasury (31 CFR Parts 500+) and the Department of Energy.

NOT GRANTED

No license under 35 U.S.C. 184 has been granted at this time, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" DOES NOT appear on this form. Applicant may still petition for a license under 37 CFR 5.12, if a license is desired before the expiration of 6 months from the filing date of the application. If 6 months has lapsed from the filing date of this application and the licensee has not received any indication of a secrecy order under 35 U.S.C. 181, the licensee may foreign file the application pursuant to 37 CFR 5.15(b).

Petitioner Illumina, Inc., EX1004, p. 1049

SelectUSA

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The U.S. offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to promote and facilitate business investment. SelectUSA provides information assistance to the international investor community; serves as an ombudsman for existing and potential investors; advocates on behalf of U.S. cities, states, and regions competing for global investment; and counsels U.S. economic development organizations on investment attraction best practices. To learn more about why the United States is the best country in the world to develop technology, manufacture products, deliver services, and grow your business, visit http://www.SelectUSA.gov or call +1-202-482-6800.

Petitioner Illumina, Inc., EX1004, p. 1050

Electronic Acknowledgement Receipt	
EFS ID:	43989498
Application Number:	16952764
International Application Number:	
Confirmation Number:	7812
Title of Invention:	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS
First Named Inventor/Applicant Name:	Gregory Porreca
Customer Number:	21710
Filer:	Thomas C. Meyers/Sandra Richards
Filer Authorized By:	Thomas C. Meyers
Attorney Docket Number:	MLB-002/06US 34468/40
Receipt Date:	08-OCT-2021
Filing Date:	19-NOV-2020
Time Stamp:	17:30:31
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Appeal Brief Filed	MLB-002-06US-APPEALBRIEF.pdf	4143483 135435babce6f5f3ba38c15cbaf37f4234169af2	no	29

Warnings:

Information:	
Total Files Size (in bytes):	4143483

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.
National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.
New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Attorney Docket No.: MLB-002/06US 34468/40

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPELLANT:	Molecular Loop Biosciences, Inc.	ART UNIT:	1634
SERIAL NUMBER:	16/952,764	CONF. NO.:	7812
FILING DATE:	November 19, 2020	EXAMINER:	Steven C. Pohnert
TITLE:	**METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS**		

FILED ELECTRONICALLY

Commissioner for Patents
P. O. Box 1450
Alexandria, VA 22313-1450

APPEAL BRIEF

This appeal is taken from the Examiner's final rejection of claims 68-69, 71, 73-74, and 76 in the Final Office Action mailed on April 9, 2021 from the United States Patent and Trademark Office.

On July 16, 2021, Appellant filed a Pre-Appeal Brief Request for Review. A Notice of Panel Decision from Pre-Appeal Brief Review was mailed on September 10, 2021 from the Office, indicating that the application remains under appeal because there is at least one actual issue for appeal.

Table of Contents

I. Real Party in Interest 3
II. Related Appeals, Interferences and Trials 4
III. Summary of Claimed Subject Matter 5
IV. Argument 6
4.1 Summary of the Examination History 6
4.2 The Examiner made key factual errors in rejecting the claims under 35 U.S.C. §112(a) 9
4.3 The Examiner made key factual errors in rejecting the claims under 35 U.S.C. §112(b) 12
4.4 The Examiner made key errors in rejecting the claims under 35 U.S.C. §101 15
4.5 The Examiner made key factual errors in rejection the claims under 35 U.S.C. §103 .. 18
4.6 Summary 21
V. APPENDIX A: Claims Appendix 23
VI. APPENDIX B: March 4, 2021 Declaration of Dr. Gregory Porreca under 37 C.F.R. §1.132 26

I. Real Party in Interest

The real party in interest is Molecular Loop Biosciences, Inc., formerly Molecular Loop Biosolutions, LLC, having a place of business at 57 Reservoir Street #1, Cambridge, Massachusetts 02138.

Molecular Loop Biosciences, Inc. is the owner of the subject application by virtue of an assignment from inventors Gregory Porreca and Uri Laserson to Good Start Genetics, Inc., recorded on December 7, 2020 with the U.S. Patent and Trademark Office at Reel/Frame 054561/0315, and an assignment from Good Start Genetics, Inc. to Molecular Loop Biosolutions, LLC, recorded on December 7, 2020 with the U.S. Patent and Trademark Office at Reel/Frame 054561/0341 and a corporate conversion of Molecular Loop Biosolutions LLC to Molecular Loop Biosciences, Inc recorded on September 28, 2021 at the United States Patent and Trademark Office at Reel/Frame 057634/0714.

II. Related Appeals, Interferences and Trials

There are no other appeals, interferences, or trials known to Appellants that may be related to, directly affect, or be directly affected by or having a bearing on the Board's decision in this appeal.

III. Summary of Claimed Subject Matter

The subject matter common to all of the pending claims is a method for correcting errors or bias introduced during nucleic acid sequencing workflows. The claimed methods comprise introducing a set of differentiator tags to nucleic acid molecules from multiple loci, such that nucleic acid molecules from one or more loci of origin are associated with a different tag. After amplifying and sequencing the differentiator tagged nucleic acid molecules, sequencing reads from copies of a nucleic acid molecule with the same differentiator tag are collapsed into a single count.

All examined dependent claims recite the method for correcting errors or bias introduced during sequencing.

Support for independent claim 1 is found in the as-filed specification, *inter alia*, at page 2, lines 18-28, page 3, lines 7-20; page 6, line 21 to page 7, line 19; page 27, line 27 to page 28, line 2; page 58, lines 3-9, and page 79, lines 8-15. Support is also be found in the Drawings, for example in FIG. 5 and FIG. 6.

Petitioner Illumina, Inc., EX1004, p. 1057

IV. Argument

4.1 Summary of the Examination History

The pending application is a continuation of U.S. Serial No. 15/231,687. The claims examined in U.S. Serial No. 15/231,687 prior to the filing of the instant continuation application were drawn to the claimed subject matter invention, namely a method comprising tagging nucleic acid molecule with a differentiator tag prior to amplification or sequencing steps. The claims of U.S. Serial No. 15/231,687 were first rejected by Examiner Pohnert in Office Actions mailed on June 6, 2017, October 27, 2017, and October 12, 2018 under 35 U.S.C. §§102 and 103 in view of primary reference Pourmand (BMC informatics, 2006, 7(500)).

On November 15, 2018, inventor Dr. Gregory Porreca, representatives for Appellant, and Examiner Pohnert conducted an examiner interview in which Examiner Pohnert suggested amending the claims as an after-final amendment to differentiate them from Pourmand. On December 5, 2018, Appellant amended the claims in view of Examiner Pohnert's comments along with a Request for Consideration Under the After Final Consideration Pilot Program 2.0 (AFCP). Despite the AFCP request being deemed a proper request, and despite Appellant's implementation of Examiner Pohnert's own suggestion, Examiner Pohnert failed to enter the amendments to the claims, requiring Appellant to file a Request for Continued Examination (RCE), which Appellant filed on December 26, 2018.

On May 2, 2019, Appellant's representative followed up with Examiner Pohnert regarding the status of Application No. 15/231,687. Examiner Pohnert informed Appellant's representative that although the RCE had been processed, "the instant application was not one of the applications indicated by the office to be at the top of the examiner's list" and that "it would probably be at least a couple of months before he would work on the application." *Appellant-Initiated Interview Summary mailed May 8, 2019 in U.S. Serial No. 15/231,687.*

On July 17, 2019, despite Appellant's amendment to the claims, Examiner Pohnert again rejected the claims of U.S. Serial No. 15/231,687 in view of Pourmand. On October 22, 2019,

inventor Dr. Gregory Porreca, representatives for Appellant, and Examiner Pohnert again conducted an examiner interview. On October 24, 2019, Appellants again amended the claims in view of Examiner Pohnert's comments. Nevertheless, on December 20, 2019, Examiner Pohnert again rejected the amended claims in view of the primary reference Pourmand.

On February 18, 2020, Appellant again amended the claims of U.S. Serial No. 15/231,687, along with an AFCP request, and a Declaration under 37 C.F.R. §1.132 by inventor Dr. Porreca explaining in detail how a person having skill in the art would understand that the references cited by Examiner Pohnert fail to teach tagging each nucleic acid molecule with a unique barcode as recited by the claimed invention.

On February 26, 2020, Examiner Pohnert once again did not enter the amendments to the claims, requiring Appellant to file an RCE, which Appellant filed on March 12, 2020. On July 14, 2020, Examiner Pohnert again rejected the amended claims in view of the Pourmand reference. Stunningly, Examiner Pohnert rejected Dr. Porreca's explanation of the scientific basis for the differences between the claimed invention and prior art. Examiner Pohnert failed to provide even a single scientific fact, counter-argument, or explanation as to why he did not consider scientific facts and explanations by one of the leading experts on the subject matter of the claimed invention.

On October 2, 2020, Appellant amended the claims to address Examiner Pohnert's remarks provided in the rejection mailed February 26, 2020. On November 18, 2020, Dr. Gregory Porreca and representatives for Appellant again conducted an examiner interview with Examiner Pohnert. In the examiner interview, Examiner Pohnert stated that the amended claims were drawn to a new invention for correcting errors or bias and would not be entered in Application No. 15/231,687. Although Appellant did not agree with the Examiner's interpretation, on November 18, 2020, Appellant filed the instant continuation application, U.S. Application No. 16/952,764 in yet another attempt to accommodate the Examiner.

On February 10, 2021, Examiner Pohnert rejected claims 68-69, 71, 73-74, and 76 now at issue on appeal under 35 U.S.C. §112(a), 35 U.S.C. §112(b), and 35 U.S.C. §101. Examiner Pohnert further rejected claims 68-69 and 76 under 35 U.S.C. §103 in view of primary reference Porreca 2007 (Nature Methods, 2007, 4:931-36), claim 74 in view of Porreca 2007 in light of Wang (Genome Biology, 2017, 8:R246.1-.14), and claims 71 and 73 in view of Porreca 2007 in light of Schiffman (Journal of Clinical Oncology (2007) 25:530).

On March 5, 2021, Appellant submitted a new Declaration under 37 C.F.R. §1.132 (attached as Appendix B) by Dr. Porreca. In his second declaration, Dr. Porreca explained that he is the first named author of the Porreca 2007 article cited by the Examiner and explained why the targeting arms of the molecular inversion probes disclosed in Porreca 2007 are not the same as differentiator tags as disclosed in the claimed invention. Dr. Porreca explained that the targeting arms of the molecular inversion probes disclosed in Porreca 2007 cannot be used to differentiate between identical copies of a gene because the arms hybridize to the flanking sequences of all copies of a gene, and therefore cannot differentiate between those copies. Dr. Porreca further explained that as the author of Porreca 2007, the concept of tagging nucleic molecules with differentiator tags had not been envisioned at the time of the Porreca 2007 publication.

Again, ignoring an expert in the field (who also happens to the lead author of a reference cited in the rejection) without scientific counter-argument or explanation, in the Office Action, mailed April 9, 2021, Examiner Pohnert maintained each of the rejections of from the February 10, 2021 Office Action, including the rejection under 35 U.S.C. §103 in view of Porreca 2007.

Appellant submits that the rejections of the claims of the parent application and this continuation application show that Examiner Pohnert has failed to conduct a principled examination of the claims. Examiner Pohnert has repeatedly failed to properly substantiate his rejections and has improperly dismissed the sworn affidavits of a person having skill in the art without explanation or counter-argument. Moreover, Examiner Pohnert has repeatedly made

suggestions to the Appellant, all of which were accepted by Appellant, only to subsequently maintain rejections as if the interviews never took place.

Appellant notes that the public record indicates that Examiner Pohnert has been an examiner for 15 years, yet he is the record examiner on only 69 issued patents. According to industry data collection services, Examiner Pohnert has a 3-year grant rate of just 4% over 338 cases, placing Examiner Pohnert in the bottom 2% of examiners in having an application granted once assigned to his docket. *See* PatentBots, accessed September 27, 2021, *<https://www.patentbots.com/stats/examiner/1634-POHNERT-STEVEN-C>*.

4.2 The Examiner made key factual errors in rejecting the claims under 35 U.S.C. §112(a)

On pages 7-9 of the final Office Action, mailed April 9, 2021, Examiner Pohnert rejected claims 68-69, 71, 73-74, and 76 under 35 U.S.C. §112(a) as failing to comply with the written description requirement.

With regard to claim 68, Examiner Pohnert stated that the specification does not provide any guidance on what is required of collapsing reads, or "how collapsing is done or how collapsing corrects errors." With regard to claim 74, Examiner Pohnert stated that the specification provides "no guidance on how to determine a copy number of genomic regions or transcript based on collapsing reads of target:differentiator tag combinations." With regard to claim 76, Examiner Pohnert stated that the specification "provides no guidance on how to perform recognizing, how recognizing is done or how recognizing corrects errors or bias."

Examiner Pohnert is clearly incorrect and Appellant requests that the Patent Trial and Appeal Board (PTAB) reverse the rejection of claims 68-69, 71, 73-74, and 76 under 35 U.S.C. §112(a).

Appellant notes that the terms "collapsing reads" to correct errors or bias, "determining the copy number of a genomic region" from collapsed reads, and "recognizing target/differentiator tags having the same sequence as being derived from the same input

molecule" are disclosed and fully-supported in the as-filed specification. The table below shows clear support for the rejected claim terms.

Table 1: Claim terms rejected in error under 35 U.S.C. §112(a)

Term	Support is as-filed specification of U.S. Serial No. 15/231,687
"collapsing reads" to correct errors or bias *and* "recognizing target/differentiator tags having the same sequence as being derived from the same input molecule"	"[S]ince any particular target:differentiator tag mapping will not occur more than once by chance, and therefore will only be observed if bias was present somewhere in the sample preparation process. **For any target:differentiator tag combination observed more than once, all such reads are 'collapsed' into a single read** before consensus basecalls are determined. This will cancel the effect of bias on consensus basecall accuracy. **FIG. 5 depicts a method for making diploid genotype calls in which repeat target:differentiator tag combination are collapsed**." *Pg. 79, lines 8-15* "In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve the use of differentiator tag sequences to uniquely tag individual nucleic acid molecules. In some embodiments, the differentiator tag sequence permit the detection of bias based on the occurrence of combinations of differentiator tag and target sequences observed in a sequencing reaction. In other embodiments, the methods reduce errors caused by bias, or the risk of bias, associated with the capture, amplification and sequencing of genetic loci, e.g., for diagnostic purposes. *Pg. 6, lines 21-28* Aspects of the invention relate to providing sequence tags (referred to as differentiator tags) that are useful to determine whether target nucleic acid sequences identified in an assay are from independently isolated target nucleic acids or from multiple copies of the same target nucleic acid molecule (e.g., due to bias in a preparative step, for example, amplification). This information can be used to help analyze a threshold number of independently isolated target nucleic acids from a biological sample in order to obtain sequence information that is reliable and can be used to make a genotype conclusion (e.g., call) with a desired degree of confidence. This information also can be used to detect bias in one or more nucleic acid preparative steps." *Pg. 6, line 29 to page 7, line 4* "Aspects of the invention are based, in part, on the identification of different sources of ambiguity and error in genetic analyses, and, in part, on the identification of one or more approaches to avoid, reduce, recognize, and/or resolve these errors and ambiguities at different stages in a genetic analysis. Aspects of the invention relate to methods and compositions for addressing bias and/or stochastic variation associated with one or more preparative and/or analytical steps of a nucleic acid evaluation technology." *Pg. 27, line 27 to page 28, line 2*
"determining the copy number of a genomic region"	"**Other examples include cancer, which has been associated with microsatellite instability (MSI) involving an increase or decrease in the genomic copy number of nucleic acid repeats at one or more microsatellite loci** (e.g., BAT-25 and/or BAT-26). The are currently many sequencing-based assays for determining the number of nucleic acid sequence repeats at a particular locus and identifying the presence of nucleic acid insertions or deletions. However, such techniques are not

	useful in a high throughput multiplex analysis where the entire length of a region may not be sequenced." *Pg. 58, lines 3-9.* "Accordingly, in some aspects, methods of genotyping a subject are provided. In some embodiments, the methods comprise determining the sequence of at least a threshold number of independently isolated nucleic acids, wherein the sequence of each isolated nucleic acid comprises a target nucleic acid sequence and a differentiator tag sequence, wherein the threshold number is a number of unique combinations of target nucleic acid and differentiator tag sequences, wherein the isolated nucleic acids are identified as independently isolated if they comprise unique combinations of target nucleic acid and differentiator tag sequences, and wherein the target nucleic acid sequence is the sequence of a genomic locus of a subject." *Pg. 7, lines 11-19*

Moreover, in addition to the clear guidance provided for the claim terms in the specification as detailed in the table above, Examples 2 and 3 and Figures 5 and 6 of the as-filed specification provide experimental implementation of the claimed invention, including recognizing target/differentiator tags having the same sequence as being derived from the same input molecule and collapsing the reads to correct for errors or bias. Examiner Pohnert's assertion that the specification provides "no guidance" for the rejected claim terms is in clear error and is contrary to the explicit words of the specification.

Additionally, the Examiner erred by refusing to give due weight to the technical opinions and factual findings of Dr. Porreca provided in his expert declaration filed with the response dated March 5, 2021. Although any fair reading of the Specification would make it clear that the subject claim terms are supported therein, Dr. Porreca's declaration goes further in representing the understanding what is clearly written in the Specification by a recognized expert in the field. Dr. Porreca's declaration, at paragraphs 4 and 6, explains in detail how a person of skill in the art would understand that the as-filed specification teaches collapsing target:differentiator tag combinations to correct for errors and bias. The Examiner's dismissal without explanation of Dr. Porreca's factual assertions regarding the specification is another clear error. If the Examiner disagrees with Dr. Porreca's expert conclusions, he must provide documentary evidence to substantiate the rebuttal, or, if the Examiner is a subject matter expert, he should have provided an affidavit providing evidence of expertise and the technical basis for the conclusions made. *In*

re Kahn, 441 F.3d 997, 998 (Fed. Cir. 2006). In view of the above remarks, Appellant requests that the Patent Trial and Appeal Board (PTAB) reverse the rejection of claims 68-69, 71, 73-74, and 76 under 35 U.S.C. §112(a).

4.3 The Examiner made key factual errors in rejecting the claims under 35 U.S.C. §112(b)

On pages 10-11 of the final Office Action, Examiner Pohnert rejected claims 68-69, 71, 73-74, and 76 under 35 U.S.C. §112(b) as being indefinite.

Examiner Pohnert, in three paragraphs, in-part reiterating his comments in rejecting claims 68-69, 71, 73-74, and 76 under 35 U.S.C. §112(a), once again stated that the specification does not provide any guidance on what is required of collapsing reads, or "how collapsing is done or how collapsing corrects errors," provides "no guidance on how to determine a copy number of genomic regions or transcript based on collapsing reads of target:differentiator tag combinations," and "provides no guidance on how to perform recognizing, how recognizing is done or how recognizing corrects errors or bias."

As discussed in section 4.3 and Table 1 of this Appeal Brief, Examiner Pohnert is clearly incorrect and Appellant requests that the Patent Trial and Appeal Board (PTAB) reverse the rejection of claims 68-69, 71, 73-74, and 76 under 35 U.S.C. §112(b).

Clear and detailed explanations of the claim terms can be found in the as-filed specification, for example, as detailed in the table below.

Table 2: Claim terms rejected in error under 35 U.S.C. §112(b)

Term	**Support is as-filed specification of U.S. Serial No. 15/231,687**
"collapsing reads" to correct errors or bias *and* "recognizing target/differentiator tags having the same sequence as being derived from	"[S]ince any particular target:differentiator tag mapping will not occur more than once by chance, and therefore will only be observed if bias was present somewhere in the sample preparation process. **For any target:differentiator tag combination observed more than once, all such reads are 'collapsed' into a single read** before consensus basecalls are determined. This will cancel the effect of bias on consensus basecall accuracy. **FIG. 5 depicts a method for making diploid genotype calls in which repeat target:differentiator tag combination are collapsed**." *Pg. 79, lines 8-15* "In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve the use of differentiator tag

the same input molecule"	sequences to uniquely tag individual nucleic acid molecules. In some embodiments, the differentiator tag sequence permit the detection of bias based on the occurrence of combinations of differentiator tag and target sequences observed in a sequencing reaction. In other embodiments, the methods reduce errors caused by bias, or the risk of bias, associated with the capture, amplification and sequencing of genetic loci, e.g., for diagnostic purposes. *Pg. 6, lines 21-28* Aspects of the invention relate to providing sequence tags (referred to as differentiator tags) that are useful to determine whether target nucleic acid sequences identified in an assay are from independently isolated target nucleic acids or from multiple copies of the same target nucleic acid molecule (e.g., due to bias in a preparative step, for example, amplification). This information can be used to help analyze a threshold number of independently isolated target nucleic acids from a biological sample in order to obtain sequence information that is reliable and can be used to make a genotype conclusion (e.g., call) with a desired degree of confidence. This information also can be used to detect bias in one or more nucleic acid preparative steps." *Pg. 6, line 29 to page 7, line 4* "Aspects of the invention are based, in part, on the identification of different sources of ambiguity and error in genetic analyses, and, in part, on the identification of one or more approaches to avoid, reduce, recognize, and/or resolve these errors and ambiguities at different stages in a genetic analysis. Aspects of the invention relate to methods and compositions for addressing bias and/or stochastic variation associated with one or more preparative and/or analytical steps of a nucleic acid evaluation technology." *Pg. 27, line 27 to page 28, line 2*
"determining the copy number of a genomic region"	"**Other examples include cancer, which has been associated with microsatellite instability (MSI) involving an increase or decrease in the genomic copy number of nucleic acid repeats at one or more microsatellite loci** (e.g., BAT-25 and/or BAT-26). The are currently many sequencing-based assays for determining the number of nucleic acid sequence repeats at a particular locus and identifying the presence of nucleic acid insertions or deletions. However, such techniques are not useful in a high throughput multiplex analysis where the entire length of a region may not be sequenced." *Pg. 58, lines 3-9.* "Accordingly, in some aspects, methods of genotyping a subject are provided. In some embodiments, the methods comprise determining the sequence of at least a threshold number of independently isolated nucleic acids, wherein the sequence of each isolated nucleic acid comprises a target nucleic acid sequence and a differentiator tag sequence, wherein the threshold number is a number of unique combinations of target nucleic acid and differentiator tag sequences, wherein the isolated nucleic acids are identified as independently isolated if they comprise unique combinations of target nucleic acid and differentiator tag sequences, and wherein the target nucleic acid sequence is the sequence of a genomic locus of a subject." *Pg. 7, lines 11-19*

Moreover, in merely repeating the language from his rejection under 35 U.S.C. §112(a)

and in cursorily dismissing the language of the specification describing the claim terms and their operation in the claimed invention, Examiner Pohnert has deprived Appellant of an opportunity to meaningful respond to the rejection under 35 U.S.C. §112(b). Appellant submits that

As stated in M.P.E.P.:

> when rejecting a claim as *prima facie* indefinite under 35 U.S.C. 112(b) or pre-AIA 35 U.S.C. 112, second paragraph, the examiner should provide enough information in the Office action to permit applicant to make a meaningful response, as the indefiniteness rejection requires the applicant to explain or provide evidence as to why the claim language is not indefinite or amend the claim. For example, in making a *prima facie* case of indefiniteness, the examiner should point out the specific term or phrase that is indefinite, explain in detail why such term or phrase renders the metes and bounds of the claim scope unclear and, whenever practicable, indicate how the indefiniteness issues may be resolved to overcome the rejection.

M.P.E.P. §2173(II)(B). Examiner Pohnert has failed to provide any meaningful explanation as to why the claim terms are not definite and has taken no effort to indicate how the indefiniteness issues may be resolved.

Additionally, the Examiner erred by refusing to give due weight to the technical opinions and factual findings of Dr. Porreca provided in his expert declaration filed with the response dated March 5, 2021. Although any fair reading of the Specification would make it clear that the subject claim terms are supported therein, Dr. Porreca's declaration goes further in representing the understanding of what is clearly written in the Specification by a recognized expert in the field. Dr. Porreca's declaration, at paragraphs 4 and 6, explains in detail how a person of skill in the art would understand that the as-filed specification teaches collapsing target:differentiator tag combinations to correct for errors and bias. The Federal Circuit has made clear that 35 U.S.C. 112(b) requires only that "the claims read in light of the specification, reasonably apprise <u>those skilled in the art</u> and are as precise as the subject matter permits. As a matter of law, <u>no court can demand more</u>." *Hybritech, Inc. v. Monoclonal Antibodies, Inc.*, 802 F.2d 1367 (Fed. Cir. 1986)

(emphasis added). The Examiner's dismissal without explanation of Dr. Porreca's factual assertions regarding the specification is another clear error. Once again, if the Examiner disagrees with Dr. Porreca's expert conclusions, he must provide documentary evidence to substantiate the rebuttal, or, if the Examiner is a subject matter expert, he should have provided an affidavit providing evidence of expertise and the technical basis for the conclusions made. *In re Kahn*, 441 F.3d 997, 998 (Fed. Cir. 2006).

In view of the above remarks, Appellant requests that the Patent Trial and Appeal Board (PTAB) reverse the rejection of claims 68-69, 71, 73-74, and 76 under 35 U.S.C. §112(b).

4.4 The Examiner made key errors in rejecting the claims under 35 U.S.C. §101

On pages 11-14 of the Office Action, Examiner Pohnert rejected claims 68-69, 71, 73-74, and 76 under 35 U.S.C. §101 as being directed to a mental step.

Examiner Pohnert is in error in making this rejection and Appellant requests that the Patent Trial and Appeal Board (PTAB) reverse the rejection of claims 68-69, 71, 73-74, and 76 under 35 U.S.C. §101.

The Supreme Court has held that everything under the sun made by a human is patentable under §101, subject to three narrow judicial exceptions: natural phenomena, laws of nature, and abstract ideas. *Diamond v. Chakrabarty*, 447 U.S. 303, 308-309 (1980). Even if the claim is directed to one of the judicial exceptions, then the second part of the analysis provides that the claim is still patent eligible if it contains "significantly more" than the judicial exception. *Mayo Collaborative Servs. v. Prometheus Labs.*, Inc., 556 U.S. 66 (2012); *Alice Corp. v. CLS Bank Int'l*, 573 U.S. 208 (2014).

The USPTO's 2019 Revised Patent Subject Matter Eligibility Guidance (the "*Revised Guidance*") explains that the judicial category of mental concepts seeks to exclude from patentability any process that would include an individual performing the claimed steps mentally or with a pencil and paper. *Revised Guidance on third page.* As a corollary, this judicial category does not include inventions that "could not, as a practical matter, be performed entirely in a

human's mind." *CyberSource Corp. v. Retail Decisions, Inc.*, 654 F.3d 1366, 1375 (Fed. Cir. 2015)(emphasis added).

The claimed invention is directed to methods comprising tagging individual nucleic acid molecules with a differentiator tag prior to amplification and sequencing steps. Each of the claims require multiple physical laboratory steps to tag nucleic acid molecules with differentiator tags, amplify the tagged molecules, and sequence the amplicons. The claims specifically recite amplifying and sequencing steps. As the Examiner knows, or should know, it is impossible for those steps to be carried out as a mental process. There is simply no way in which the steps of the claims could be carried out absent physical manipulation in the laboratory. This is another example of a clear failure of the Examiner to provide a fair and reasoned examination of the subject application.

In addition to the obvious errors based on the claim language, the rejection under 35 U.S.C. §101 ignores clear guidance regarding subject matter eligibility provided by the Office.

The USPTO's October 2019 Subject Matter Eligibility Update ("*October Update"*) explains that a claim is not "directed to" a judicial exception, and is thus patent eligible, if the judicial exception is integrated into a practical application of the judicial exception. *October Update, page 10.* A claim integrates a judicial exception if it improves the functioning of a technology. *Id., page 12.* The claim itself does not need to explicitly recite the improvement, if the derived improvement is clearly set forth in the specification. *Id.* This analysis is undertaken without reference to what is well-understood, routine, or conventional. *Id.* Moreover, the USPTO's 2019 Revised Patent Subject Matter Eligibility Guidance ("*2019 Guidance"*) states that when considering whether a judicial exception is integrated into a practical application, the analysis "specifically excludes consideration of whether the additional elements represent well-understood, routine, conventional activity . . . [e]xaminers are reminded that a claim that includes conventional elements may still integrate an exception into a practical application, thereby satisfying the subject matter eligibility requirement of Section 101." *2019 Guidance, page 55.*

Each of the claims requires multiple physical laboratory steps, including tagging nucleic acid molecules with differentiator tags, amplifying the tagged nucleic acid molecules to generate amplicons, and sequencing the amplicons. The further steps of correcting for error or bias provide the benefits of the improved sequencing workflow provided by the claimed invention. Although it is difficult to imagine sequencing nucleic acids in one's head, even if this were considered in any way mental step, the claims are still subject matter eligible under 35 U.S.C. §101.

This is made clear in the October 7, 2019 guidance issued by the Office on subject matter eligibility, issued to assist examiners and the public in understanding how the Office applies its eligibility guidance in certain fact-specific situations. In the October 7, 2019 guidance, on pages 23-28 under Example 31, the Office provided guidance regarding the following claim:

> 85. A method for amplifying BRCA1 sequences comprising: amplifying by Cool-Melt PCR all or part of a BRCA1 gene from a tissue sample from a human subject using a set of primers to produce amplified nucleic acids; and sequencing the amplified nucleic acids.

The October 2019 guidance, on page 28, affirmed that the claim is subject matter eligible. The guidance explained that:

> although nature-based product limitations are recited in the claim (e.g., the primers and BRCA1 gene), analysis of the claim as a whole indicates that the claim is focused on a process of amplifying and sequencing a BRCA1 gene, and is not focused on the products *per se*. Thus, there is no need to perform the markedly different characteristics analysis on the recited nature-based product limitations. In addition, note that because the analysis of this claim ends with eligibility at Step 2A, the Step 2B analysis does not need to be performed. Thus, the examiner would not need to evaluate the significantly more considerations for this claim.

October 2019 USPTO Subject Matter guidance 2019 at page 28 (emphasis added). Like claim 85 in Example 31 of the October 2019 guidance, independent claim 1 as a whole indicates that

the claim is focused on a process of amplifying and sequence nucleic acids.

It is clear from the above remarks that claim 1 is subject matter eligible and the Examiner's rejection of the claims under 35 U.S.C. §101 further evidences the Examiners failure to provide a fair and reasoned examination of the subject application.

Appellant notes that Examiner Pohnert at no point rejected the claims of parent application U.S. Serial No. 15/231,687 under 35 U.S.C. §101, despite the claims being drawn to similar subject matter. Examiner Pohnert should know that the law has not changed in any way that would render the recognized patent eligible claims of U.S. Serial No. 15/231,687 now ineligible in this the instant application.

In view of the above remarks, Appellant requests that the Patent Trial and Appeal Board (PTAB) reverse the rejection of claims 68-69, 71, 73-74, and 76 under 35 U.S.C. 101.

4.5 The Examiner made key factual errors in rejection the claims under 35 U.S.C. §103

On pages 14-22 of the Office Action, Examiner Pohnert rejected claims 68-69 and 76 under 35 U.S.C. §103 in view of Porreca 2007 and claim 74 in view of Porreca 2007 in light of Wang and claims 71 and 73 in view of Porreca 2007 in light of Schiffman.

In making those rejections, Examiner Pohnert has again made clear factual and legal errors. The Examiner has ignored the factual statements made in the present application, the prior art, and in Dr. Porreca's declaration. Tellingly, the rejection of claims 68-69 and 76 under §103 is based on a single reference that Examiner Pohnert himself admits **fails to disclose an element of the claims**. Not only does the Examiner misinterpret the technical details of the cited art, he ignores legal standard for a *prima facie* case for obviousness.

Examiner Pohnert is once again in error in making this rejection and Appellant requests that the Patent Trial and Appeal Board (PTAB) reverse the rejection of claims 68-69, 71, 73-74, and 76 under 35 U.S.C. §103.

In order to establish a *prima facie* case of obviousness, the prior art must provide

motivation to modify a reference or to combine multiple references in a manner that would lead one of ordinary skill in the art to an applicant's claimed invention with a reasonable expectation of success. *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398 (2007).

On page 12 of the Office Action, Examiner Pohnert concedes that Porreca 2007 does not teach "correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count." Examiner Pohnert does not, and cannot, identify any motivation provided by Porreca 2007 to modify the teachings of that reference itself to lead to the claimed invention. Moreover, Examiner Pohnert has not provided evidence or a reasoned explanation as to why common sense would lead once to modify Porreca 2007 to provide the missing limitation, when **Porreca 2007 itself recognized that PCR amplification steps were a source of nonuniformity error but did not suggest such a solution**. *Porreca 2007, pg. 4, Discussion.*

Moreover, the March 5, 2021, expert declaration of Dr. Gregory Porreca gives record evidence supporting the non-obviousness of the claims. In that declaration, Dr. Porreca—inventor of the instant claims and first author of Porreca 2007—shows that none of the cited references, including his 2007 paper, disclose a method comprising tagging individual nucleic acid molecules with differentiator tags prior to amplification or sequencing steps. *Porreca dec., par. 6-12.*

Specifically, Dr. Porreca states that at the time of his 2007 publication, he had not yet envisioned tagging nucleic acid molecules with differentiator tags. *Id., par. 9.* Moreover, Dr. Porreca reports that the unique targeting arms of the molecular inversion probes disclosed in Porreca 2007 are not differentiator probes at all. *Id. pars. 8, 11, 12.* As Dr. Porreca points out, this is because each molecular inversion probe is designed to uniquely capture an exon of interest with targeting arms that hybridize to sequences flanking the exon. *Id., para. 7.* The targeting arms are unique only in the sense that PCR primers are unique; unique designs each synthesized in the millions (this point is settled where Supplementary Figure 1 in Porreca 2007 states that the

480 probes were used at a "concentration of ~167 pM for each species"). Because the probes disclosed in Porreca 2007 are only "unique" to their target exons, the expert declaration concludes that their targeting arms cannot differentiate separate single molecules. *Id.*

Dr. Porreca further explained that Wang teaches only the use of molecular inversion probes for the improved measurement of allele copy number, that is the number of different variants of an allele of a gene at a chromosomal location. *Id., pars 10-11.* However, because the unique targeting arms of the molecular inversion probes needed to capture different alleles of the gene are actually identical, the different copies of the allele must be identified and differentiated by their own sequence rather than by the sequence of the unique targeting arms. As a result, Wang suffers from the same handicap as the molecular inversion probes disclosed in Porreca 2007 when dealing with copy number polymorphisms. *Id.* That is, because the targeting arms of the molecular inversion probes will hybridize to the flanking sequences of any copies of the gene, the unique targeting arms cannot be used to differentiate between the identical copies of the gene as separate nucleic acid molecules. *Id.* Wang makes clear this limitation in stating that copy number polymorphisms are likely attributable to identified false positives in their data that are actually true additional copies of the gene. *Id.* Wang specifically states that one of the shortcomings of their workflow is the presence of CNPs in the autosomes. *Id.* Even with the solutions proposed for "ameliorating the effect of miscalling two adjacent markers in a CNP as a false positive", Wang states that doing so is only valid for the assessment of the technology performance "and cannot generate any meaningful biological findings." *Id.*

Dr. Porreca also explained that Schiffman does not disclose differentiator tags. *Id., para. 12.* Rather, Schiffman discloses only the use of molecular inversion probes to analyze loss of heterozygosity events, which are the same allele quantification or allele copy number events as disclosed in Wang. *Id.* As a result, Schiffman does not disclose the use of differentiator probes, and therefore cannot remedy the deficiencies of Porreca 2007 or Wang. *Id.*

Additionally, not only do Porreca 2007, Wang, and Schiffman fail to disclose

differentiator tags, modifying any of the references to collapse reads would only introduce errors and bias to their sequencing workflows. This is because rather than correcting for error or bias introduced during their workflow, collapsing reads would only further introduce error or bias by incorrectly counting multiple copies of a gene present in an original sample as a single copy.

A conclusion of obviousness is simply completely unsupported.

In view of the above remarks, Appellant requests that the Patent Trial and Appeal Board (PTAB) reverse the rejection of claims 68-69, 71, 73-74, and 76 under 35 U.S.C. §103.

4.6 Summary

As detailed herein, Examiner Pohnert has continuously and systematically failed to properly substantiate his rejections under 35 U.S.C. §§112(a), 112(b), 101, and 103. The Examiner has ignored the legal standards, the explicit language of the Specification and the claims, the declaration of a leading expert in the field and has utterly failed to conduct a proper reading of the prior art. Appellant believes that Examiner Pohnert's conduct in examining this application and the '687 parent does not represent a fair and reasoned examination, and that Examiner Pohnert has abused his authority and discretion. Appellant requests that the Patent Trial and Appeal Board (PTAB) reverse the rejections of claims 68-69, 71, 73-74, and 76.

Dated: October 8, 2021

Respectfully submitted,

BROWN RUDNICK LLP

BROWN RUDNICK LLP
One Financial Center
Boston, MA 02111
Tel: (617) 856-8483
Fax: (617) 856-8201

/Thomas C. Meyers/
Thomas C. Meyers, Reg. No. 36,989
Attorney for Appellant
Email: tmeyers@brownrudnick.com

63911551 v1

Petitioner Illumina, Inc., EX1004, p. 1074

V. APPENDIX A: Claims Appendix

1-67. (Canceled)

68. (Previously presented) A method for correcting for errors or bias introduced during nucleic acid analysis workflow, the method comprising the steps of:

obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin;

introducing a set of differentiator tags, wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag;

amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons;

sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and

correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count.

69. (Previously presented) The method of claim 68, wherein said correcting step further comprises collapsing all sequence reads comprising a same differentiator tag into a single read before determining a consensus base call for the target nucleic acid molecule sequences.

70. (Previously presented) The method of claim 69, further comprising the step of determining the presence of a sequencing or amplification error based upon a number of said differentiator tag sequences.

Petitioner Illumina, Inc., EX1004, p. 1075

71. (Previously presented) The method of claim 68, wherein said biological sample is selected from the group consisting of blood and biopsy tissue.

72. (Previously presented) The method of claim 68, wherein said plurality of target nucleic acid molecules comprises bacterial or viral nucleic acids isolated from said biological sample.

73. (Previously presented) The method of claim 68, wherein said plurality of target nucleic acid molecules are circulating tumor nucleic acid molecules.

74. (Previously presented) The method of claim 68, further comprising the step of determining the copy number of a genomic region or transcript in a patient from whom the sample is obtained.

75. (Previously presented) The method of claim 68, wherein said plurality of target nucleic acid molecules are maternally-circulating fetal or placental nucleic acid molecules.

76. (Previously presented) A method for correcting for errors or bias introduced during nucleic acid analysis workflow, the method comprising the steps of:

obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin;

introducing a set of unique differentiator tags, wherein members of said set of unique differentiator tags are randomly associated with members of said plurality of target nucleic acid molecules, such that any given tag is uniquely associated with said locus of origin;

amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons;

sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and

correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule.

Petitioner Illumina, Inc., EX1004, p. 1077

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANT:	Molecular Loop Biosolutions, LLC	ART UNIT:	1634
SERIAL NUMBER:	16/952,764	CONF. NO.:	7812
FILING DATE:	November 19, 2020	EXAMINER:	Steven C. Pohnert
TITLE:	**METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS**		

FILED ELECTRONICALLY

Commissioner for Patents
P. O. Box 1450
Alexandria, VA 22313-1450

Declaration under 37 C.F.R. §1.132

I, Gregory Porreca, declare as follow:

1. I received a PhD in Genetics from Harvard University and a BS in computer science and biology from The College of New Jersey. I am the Founder, President, and Chief Executive Officer of Molecular Loop Biosolutions, LLC, a company focused on producing rapid production sequencer-ready DNA libraries. I have served as a lecturer at Harvard Medical School as well as the Personal Genome Project Director of Technology. I co-developed high throughout DNA sequencing system technology that is now part of the Applied Biosystems SOLiD nucleic acid sequencing platform. I am the named inventor on more than 30 patents or patent applications world-wide. I have authored or co-authored over 14 reviewed publications and have presented numerous invited talks and posters on aspects of the next-generation-sequencing and medical genetics.

2. I am the inventor of the above captioned patent application. I have reviewed the above-captioned patent application, including the pending claims.

3. I make this declaration in order to present facts known to me and my technical opinion with respect to the above-captioned patent application.

4. The patent application provides preparative and analytical methods for identifying target nucleic acid molecules while correcting for error or bias introduced during the workflow. *As-filed application*, pg. 6 (lines 20-23); pg. 39 (13-16). The patent application discloses tagging individual target nucleic acid molecules with differentiator tags before any amplification or sequencing steps. *Id.*, pg. 6 (21-24); pg. 40 (7-9); pg. 42 (14-27). Each differentiator tag has its own unique sequence so that each individual target nucleic acid molecule is tagged, and therefore associated, with a differentiator tag. *Id.*, pg. 41 (10-22). Accordingly, each target:differentiator tag combination uniquely identifies a single nucleic acid molecule through subsequence amplification and sequencing steps. *Id.*, pg. 39 (21-26); pg. 41 (3-7), (9-10); pg. 42

Petitioner Illumina, Inc., EX1004, p. 1078

(13-20). Because each target:differentiator tag combination can be associated back to its original, individual nucleic molecule based on the unique combination of individual nucleic acid molecules and their respective differentiator tags, the invention corrects for bias in subsequent steps. *Id.*, pg. 41 (7-9), (11-12), (23-24); pg. 42 (12), (17-20). For example, the overrepresentation of sequence reads having a combination of nucleic acid molecule sequence and its differentiator tag sequence is indicative of bias in the protocol at one or more steps between the incorporation of the differentiator tag sequence and the sequencing of the combination. *Id.* Accordingly, the bias may be corrected to provide an accurate representation of the individual nucleic acid molecule in the original pool of nucleic acid molecules. *Id.* For example, it can be recognized that target:differentiator tag combinations having the same sequence are derived from the same original individual nucleic acid molecule and are not representative of two copies of the nucleic acid molecule in the original sample or genome. *Id.*, pg. 7 (11-19), pg. 27 (27-33), pg. 28 (1-2). Target:differentiator tag combinations observed more than once can then be collapsed into a single read to cancel the effect of bias on consensus basecall accuracy. *Id.*, pg. 79, lines 8-15.

5. I have reviewed the Office Action mailed February 10, 2021, from the United States Patent and Trademark Office. I understand that the claims pending in the captioned application are alleged to be obvious by a 2007 Nature Methods publication by myself as first named author, entitled "Multiplex amplification or large sets of human exons." I also understand that the pending claims are alleged to be obvious over my 2007 Nature Methods publication in combination with a 2007 Genome Biology publication by Wang et al. entitled "Analysis of molecular inversion probe performance for allele copy number determination" and a 2007 Journal of Clinical Oncology publication by Schiffman et al. entitled "Adapting molecular inversion probe (MIP) technology for allele quantification in childhood leukemia." I have reviewed each of those cited references.

6. A distinguishing feature of the claims in my patent application is that each individual nucleic acid molecule is tagged with a unique differentiator tag prior to any amplification or sequencing steps. *As filed application*, pg. 40 (lines 7-9); pg. 42 (14-27). My claimed methods overcome systematic and random errors in genome capture, amplification, and sequencing. *Id.*, g. 6 (31-33); pg. 7 (1-4); pg. 28 (1-7); pg. 38 (25-28). In particular, my claimed invention in this patent application allows for a reduction in errors associated with the capture and amplification of nucleic acids by "collapsing" duplicate sequences, thereby providing a significant improvement in the art. *Id.*, pg. 42 (2-3); pg. 45 (2-6); pg. 80 (11-12); and FIG. 5, final "step".

7. As first named author of the 2007 Nature Methods publication, I understand what the publication discloses. The 2007 Nature Methods publication discloses a method of exon capture in which two 20 nucleotide targeting arms of a molecular inversion probe are designed to have unique sequences that hybridize immediately upstream and downstream, respectively, of an exon of interest. *Nature Methods, Figure 1.* Enzymatic gap filling between the two targeting arms results in a circular nucleic acid molecule that comprises the complement to the exon of interest and the two unique targeting arms. *Id.*

8. As first named author of the 2007 Nature Methods publication, I can affirm that the targeting arms of the molecular inversion probes disclosed in the 2007 Nature Methods

Petitioner Illumina, Inc., EX1004, p. 1079

publication are not the same as differentiator tags disclose by the invention claimed in this patent application. This is because the unique targeting arms of the molecular inversion probes disclosed in the 2007 Nature Methods publication must be able to hybridize to nucleic acid sequences flanking exons of interest. The becomes a handicap in comparison to the differentiator tags of the claimed invention, for example when determining gene copy number polymorphisms (CNPs), in which the gene as well as its flanking sequences have become duplicated one or more times in a genome. Because the targeting arms of the molecular inversion probes disclosed in my 2007 Nature Methods publication will hybridize to the flanking sequences of *all* copies of the gene, the unique targeting arms cannot be used to differentiate between the identical copies of the gene as separate nucleic acid molecules.

9. As first named author of the 2007 Nature Methods publication, I can affirm that the concept of tagging nucleic molecules with differentiator tags prior to amplification had not been envisioned at the time of the 2007 Nature Methods publication and that I first conceived the claimed invention in 2009 consonant with the filing of the earliest application to which this patent application claims priority.

10. It is further my expert opinion that the additional references cited in the office action do not remedy the deficiencies in the methods reported in my 2007 Nature Methods publication. For example, Wang teaches only the use of molecular inversion probes for the improved measurement of allele copy number, that is the number of different variants of an allele of a gene at a chromosomal location. *Wang, R246.7, Discussion.* However, because the unique targeting arms of the molecular inversion probes needed to capture different alleles of the gene are actually identical, the different copies of the allele must be identified and differentiated by their own sequence rather than by the sequence of the unique targeting arms.

11. Additionally, it is my expert opinion that the molecular inversion probes of Wang suffer from the same handicap as the molecular inversion probes disclosed in my 2007 Nature Methods publication when dealing with copy number polymorphisms. That is, because the targeting arms of the molecular inversion probes will hybridize to the flanking sequences of *any* copies of the gene, the unique targeting arms cannot be used to differentiate between the identical copies of the gene as separate nucleic acid molecules. Wang makes clear this limitation in stating that copy number polymorphisms are likely attributable to identified false positives in their data that are actually true additional copies of the gene. *Id., 246.3, Systematic false positives.* Wang specifically states that one of the shortcomings of their workflow is the presence of CNPs in the autosomes. *Id., 246.3, Trade off between resolution and performance.* Even with the solutions proposed for "ameliorating the effect of miscalling two adjacent markers in a CNP as a false positive", Wang states that doing so is only valid for the assessment of the technology performance "and cannot generate any meaningful biological findings." *Id.*

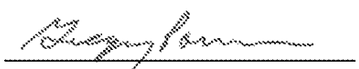
12 It if further my opinion that Schiffman discloses only the use of molecular inversion probes to analyze loss of heterozygosity events, that is allele quantification or allele copy number as disclosed in Wang. Schiffman does not disclose the use of differentiator probes, and therefore cannot remedy the deficiencies of my 2007 Nature Methods publication or Wang.

Petitioner Illumina, Inc., EX1004, p. 1080

13. The assertion that the cited references teach or suggest the presently claimed method is simply unsupported, as none of the references report tagging individual nucleic acid molecules each with a unique barcode sequence.

14. I declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine, imprisonment, or both, under §1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Date: **March 4, 2021**



Gregory Porreca, PhD

Petitioner Illumina, Inc., EX1004, p. 1081

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/952,764	11/19/2020	Gregory Porreca	MLB-002/06US 34468/40	7812

21710 7590 10/18/2021

BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

EXAMINER	
ART UNIT	PAPER NUMBER
MAIL DATE	DELIVERY MODE
10/18/2021	

Please find below and/or attached an Office communication concerning this application or proceeding.

PTOL-90A (Rev. 04/07)

Notification of Non-Compliant Appeal Brief (37 CFR 41.37)	**Application No.** 16/952,764	**Applicant(s)** Porreca et al.	
	Examiner POHNERT, STEVEN C	**Art Unit** 1634	**AIA (FITF) Status** No

--The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

The Appeal Brief filed on 08 October 2021 is defective for failure to comply with one or more provisions of 37 CFR 41.37(c).

To avoid dismissal of the appeal, applicant must file an amended brief or other appropriate correction (see MPEP 1205.03) within **ONE MONTH or THIRTY DAYS** from the mailing date of this Notification, whichever is longer. **EXTENSIONS OF THIS TIME PERIOD MAY BE GRANTED UNDER 37 CFR 1.136.**

1. ☐ The brief does not contain the items required under 37 CFR 41.37(c), or the items are not under the proper heading or in the proper order.

2. ☑ (a) The brief does not contain a concise explanation of the subject matter defined in each of the rejected independent claims, referring to the specification in the Record by page and line number or by paragraph number and to the drawings, if any, by reference characters; and/or (b) the brief fails to identify, for each rejected independent claim and for each dependent claim argued separately that contains a means plus function or step plus function recitation under 35 U.S.C. 112, sixth paragraph, the structure, material, or acts described in the specification as corresponding to each claimed function with reference to the specification in the Record by page and line number or paragraph number, and to the drawings, if any, by reference characters (37 CFR 41.37(c)(1)(iii)).

3. ☐ The brief does not contain a correct copy of the appealed claims as an appendix thereto (37 CFR 41.37(c)(1)(v)).

4. ☑ Other (including any explanation in support of the above items):

The Summary of Claimed Subject Matter is non-compliant, as it does not identify the independent claims on appeal to the Specification by page and line number.(specifically claims 68 and 76) An entirely new Appeal Brief is not necessary, only the corrected section..

/REGINALD TYSON/
(571)272-1634

Petitioner Illumina, Inc., EX1004, p. 1083

Electronic Acknowledgement Receipt	
EFS ID:	44078164
Application Number:	16952764
International Application Number:	
Confirmation Number:	7812
Title of Invention:	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS
First Named Inventor/Applicant Name:	Gregory Porreca
Customer Number:	21710
Filer:	Thomas C. Meyers/Jeanne Waters
Filer Authorized By:	Thomas C. Meyers
Attorney Docket Number:	MLB-002/06US 34468/40
Receipt Date:	20-OCT-2021
Filing Date:	19-NOV-2020
Time Stamp:	16:35:08
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Supplemental Appeal Brief	MLB-002-06US-Summar-subject-matter-102021.pdf	122272 4965a752cca2a04f955a6ba10685bf98b53b3418	no	3

Warnings:

Information:	
Total Files Size (in bytes):	122272

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.
National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.
New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Attorney Docket No.: MLB-002/06US 34468/40

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPELLANT:	Molecular Loop Biosciences, Inc.	ART UNIT:	1634
SERIAL NUMBER:	16/952,764	CONF. NO.:	7812
FILING DATE:	November 19, 2020	EXAMINER:	Steven C. Pohnert
TITLE:	**METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS**		

FILED ELECTRONICALLY

Commissioner for Patents
P. O. Box 1450
Alexandria, VA 22313-1450

RESPONSE

This paper is filed in response to the Notification of Non-Compliant Appeal Brief mailed October 18, 2021 in connection with the above-identified application.

A revised **Summary of Claimed Subject Matter** begins on page 2.

Remarks begin on page 3.

Applicant submits that no fee is deemed necessary in connection with the filing of this paper. Nonetheless, applicant authorizes any required fees to be charged to Deposit Account No. 500369.

Summary of Claimed Subject Matter

The subject matter common to all of the pending claims is a method for correcting errors or bias introduced during nucleic acid sequencing workflows. The claimed methods comprise introducing a set of differentiator tags to nucleic acid molecules from multiple loci, such that nucleic acid molecules from one or more loci of origin are associated with a different tag. After amplifying and sequencing the differentiator tagged nucleic acid molecules, sequencing reads from copies of a nucleic acid molecule with the same differentiator tag are collapsed into a single count.

All examined dependent claims recite the method for correcting errors or bias introduced during sequencing.

Support for independent claim 68 is found in the as-filed specification, *inter alia*, at page 2, lines 18-28, page 3, lines 7-20; page 6, line 21 to page 7, line 19; page 27, line 27 to page 28, line 2; page 58, lines 3-9, and page 79, lines 8-15. Support is also be found in the Drawings, for example in FIG. 5 and FIG. 6.

Support for independent claim 76 is found in the as-filed specification, *inter alia*, at page 2, lines 18-28, page 3, lines 7-20; page 6, line 21 to page 7, line 19; page 27, line 27 to page 28, line 2; page 58, lines 3-9, and page 79, lines 8-15. Support is also be found in the Drawings, for example in FIG. 5 and FIG. 6.

Petitioner Illumina, Inc., EX1004, p. 1087

REMARKS

Applicant submits herewith a revised Summary of Claimed Subject Matter filed in response to the Notification of Non-Compliant Appeal Brief mailed October 18, 2021.

In particular, Applicant respectfully requests that the revised Summary of Claimed Subject Matter section provided on page 2 of this response be entered and replace the Summary of Claimed Subject Matter section provided in the Appeal Brief filed on October 8, 2021. Applicant submits that the revised Summary of Claimed Subject Matter section now identifies the independent claims (claims 68 and 76), and further provides a concise explanation of the subject matter defined in the independent claims, including reference to the specification by page and line number and figures.

Accordingly, Applicant respectfully submits that the revised Summary of Claimed Subject Matter section is now in compliance with 37 CFR 41.37(c)(1)(iii) and, as such, the appeal brief meets all requirements for filing of an appeal brief under 37 CFR 41.37.

Dated: October 20, 2021

Respectfully submitted,

BROWN RUDNICK LLP

BROWN RUDNICK LLP
One Financial Center
Boston, MA 02111
Tel: (617) 856-8483
Fax: (617) 856-8201

/Thomas C. Meyers/
Thomas C. Meyers, Reg. No. 36,989
Attorney for Appellant
Email: tmeyers@brownrudnick.com

63911551 v1

Petitioner Illumina, Inc., EX1004, p. 1088

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/952,764	11/19/2020	Gregory Porreca	MLB-002/06US 34468/40	7812

21710 7590 02/01/2022

BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

EXAMINER	
POHNERT, STEVEN C	

ART UNIT	PAPER NUMBER
1634	

NOTIFICATION DATE	DELIVERY MODE
02/01/2022	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

ip@brownrudnick.com
usactions@brownrudnick.com

UNITED STATES PATENT AND TRADEMARK OFFICE

Commissioner for Patents
United States Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450
www.uspto.gov

BEFORE THE PATENT TRIAL AND APPEAL BOARD

Application Number: 16/952,764
Filing Date: 19 Nov 2020
Appellant(s): Porreca et al.

Thomas Meyers
For Appellant

EXAMINER'S ANSWER

This is in response to the appeal brief filed 4/9/2021.

(1) Grounds of Rejection to be Reviewed on Appeal

Every ground of rejection set forth in the Office action dated 4/9/2021 from which the appeal is taken is being maintained by the examiner except for the grounds of rejection (if any) listed under the subheading "WITHDRAWN REJECTIONS." New grounds of rejection (if any) are provided under the subheading "NEW GROUNDS OF REJECTION."

The following ground(s) of rejection are applicable to the appealed claims.

Claim Rejections - 35 USC § 112

1. The following is a quotation of the first paragraph of 35 U.S.C. 112(a):

> (a) IN GENERAL.—The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor or joint inventor of carrying out the invention.

The following is a quotation of the first paragraph of pre-AIA 35 U.S.C. 112:

> The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor of carrying out his invention.

2. Claims 68-69, 71, 73-74, 76 rejected under 35 U.S.C. 112(a) or 35 U.S.C. 112 (pre-AIA), first paragraph, as failing to comply with the written description requirement. The claim(s) contains subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor or a joint inventor, or for applications subject to pre-AIA 35 U.S.C. 112, the inventor(s), at the time the application was filed, had possession of the claimed invention.

Claim 68 recites, "correcting for error or bias introduced during said workflow by collapsing target: differentiator tag combinations observed more than once into a single count." The specification recites collapsing three times. The specification in figure 5 demonstrates 4 bases with one base difference in sequences. It demonstrates in a very simple example that sequences can be preferentially amplified. The specification does not any guidance on what is required of collapsed, how collapsing is done or how collapsing corrects errors. How to determine to collapse sequences of greater complexity, or what is require of a target sequence to identify it to be collapsed. Thus the claims lack adequate written description.

Further claim 74 recites, "further comprising the step of determining the copy number of a genomic region or transcript in a patient from whom the sample is obtained." The specification and claims provide no guidance on how to determine a copy number of genomic region or transcript based on collapsing reads of target: differentiator tag combinations. Thus the claims lack adequate written description.

Claim 76 recites, "correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule." The specification recites recognize six times. The specification does not recite, in the context of the limitations of claim 76.The specification in figure 5 demonstrates 4 bases with one base difference in sequences. It demonstrates in a very simple example that sequences can be preferentially amplified. The specification does not any guidance on what is required of recognizing, how recognizing is done or how recognizing corrects errors. How to determine to

recognizing sequences of greater complexity, or what is require of a target sequence to identify it to be corrected. Thus the claims lack adequate written description.

3. The following is a quotation of 35 U.S.C. 112(b):

> (b) CONCLUSION.—The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the inventor or a joint inventor regards as the invention.

The following is a quotation of 35 U.S.C. 112 (pre-AIA), second paragraph:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

4. Claims 68-69,71, 73-74, 76 rejected under 35 U.S.C. 112(b) or 35 U.S.C. 112 (pre-AIA), second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which the inventor or a joint inventor (or for applications subject to pre-AIA 35 U.S.C. 112, the applicant), regards as the invention.

Claim 68 recites, "correcting for error or bias introduced during said workflow by collapsing target: differentiator tag combinations observed more than once into a single count." The specification recites collapsing three times. The specification provides no guidance on how to perform collapsing, why collapsing is done or how collapsing corrects errors of bias. Thus the recitation is incomplete and unclear as it provides no guidance on how to perform collapsing, why collapsing is done or how collapsing corrects errors of bias.

Further claim 74 recites, "further comprising the step of determining the copy number of a genomic region or transcript in a patient from whom the sample is obtained." The recitation is incomplete and unclear as the specification and claims provide no guidance on how to determine a copy number of genomic region or transcript based on collapsing reads of target: differentiator tag combinations.

Claim 76 recites, "correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule." The specification recites recognize six times. The specification provides no guidance on how to perform recognizing, why recognizing is done or how recognizing corrects errors of bias. Thus the recitation is incomplete and unclear as it provides no guidance on how to perform recognizing, why recognizing is done or how recognizing corrects errors of bias.

Claim Rejections - 35 USC § 101

5. 35 U.S.C. 101 reads as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

6. Claims 68-69, 71, 73-74, 76 are rejected under 35 U.S.C. 101 because the claimed invention is directed to a mental step or abstract idea without significantly more. The claim(s) recite(s) the abstract idea or mental step of correcting. This judicial exception is not integrated into a practical application because if there is no steps which depend from or otherwise flow from the correcting step. The claim(s) does/do not include additional elements that are sufficient to amount to significantly more than the judicial exception because if there are no additional steps integrating the correcting.

Claim analysis

The instant claim 68 is directed towards a method for correcting for errors or bias introduced during nucleic acid analysis workflow, the method comprising the steps of: obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin; introducing a set of differentiator tags, wherein

members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag; amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons; sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and correcting for error or bias introduced during said workflow. The correcting step is a mental step or abstract idea.

The instant claim 76 is directed towards method for correcting for errors or bias introduced during nucleic acid analysis workflow, the method comprising the steps of: obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin; introducing a set of unique differentiator tags, wherein members of said set of unique differentiator tags are randomly associated with members of said plurality of target nucleic acid molecules, such that any given tag is uniquely associated with said locus of origin; amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons; sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule.. The correcting step is a mental step or abstract idea.

The steps of obtaining, introducing, amplifying and sequencing are active steps.

Dependent claims set forth further limitations to about the reference level, method of detecting expression, subject and source of cancer.

According to the 2019 Patent Eligibility Guidance an initial two step analysis is required for determining statutory eligibility.

Step 1. Is the claim directed to a process, machine, manufacture, or composition of matter? In the instant case the Step 1 requirement is satisfied as the claims are directed towards a process.

Step 2A Prong one. Does the claim recite a law of nature, a natural phenomenon or an abstract idea? Yes, abstract idea and law.

With regards to claim 68, the claim recites, "and correcting for error or bias introduced during said workflow." This is an abstract idea or mental step.

With regards to claim 76, the claim recites, "correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule." This is an abstract idea or mental step.

Step 2A prong two. Does the claim recite additional elements that integrate the judicial exception into a practical application? The answer is no as the claims provide no active step in which the correcting is relied upon or otherwise integrated.

Step 2B. Does the claim recite additional elements that are significantly more than the judicial exceptions? No

Claims 68 and 76 require the steps of obtaining, introducing, amplifying and sequencing are active steps. The specification teaches:

Sample preparative methods may fall into three classes: 1) single- or several-target amplification (e.g., uniplex PCR, 'multiplex' PCR), 2) multi-target hybridization

enrichment (e.g., Agilent SureSelect 'hybrid capture' [Gnirke et al 2009, Nature methods 27:182-9], Roche/Nimblegen 'sequence capture' [Hodges et al 2007, Nature genetics 39:1522-7], and 3) multi-target circularization selection (e.g. molecular inversion probes or padlock probes, [Porreca et al 2007, Nature methods 4:931-6, Turner et al 2009, Nature methods 6:315-6], 'selectors' [Dahl et al 2005, Nucleic acids research 33:e71]). arrays, single nucleotide polymorphism (SNP) arrays (Affymetrix, Santa Clara, Calif.).

Claim Rejections - 35 USC § 103

7. In the event the determination of the status of the application as subject to AIA 35 U.S.C. 102 and 103 (or as subject to pre-AIA 35 U.S.C. 102 and 103) is incorrect, any correction of the statutory basis for the rejection will not be considered a new ground of rejection if the prior art relied upon, and the rationale supporting the rejection, would be the same under either status.

8. The following is a quotation of pre-AIA 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

9. Claim 68-69, 76 is/are rejected under pre-AIA 35 U.S.C. 103(a) as being unpatentable over Porreca (Nature methods (2007) volume 4, pages 931-936).

The specification provides no guidance on how, "correcting for error or bias introduced during said workflow by collapsing target: differentiator tag combinations observed more than once into a single count" is done thus this is considered broad.

Further the specification teaches, "As used herein, a "differentiator tag sequence" is a sequence of a nucleic acid (a preparative nucleic acid), which in the context of a plurality of different isolated nucleic acids, identifies a unique, independently isolated

nucleic acid." Thus this broadly encompasses the targeting arms of a molecular inversion probe (MIP).

Porreca teaches figure 1:

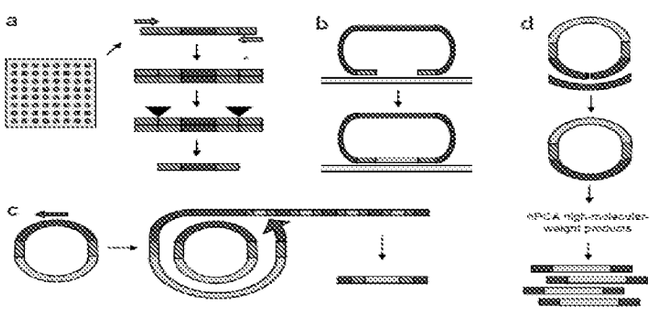


Figure 1 | Schematic of multiplex exon capture. (a) A library of 100-mer targeting oligo precursors is synthesized in parallel on a programmable microarray. Each oligo consists of common 15 nt flanking sequences (purple), unique 20 nt targeting arms (red) and a common 30 nt linker (dark blue). The oligos are released and PCR amplified with a single set of primers directed at the common flanking sequences. Double digestion with nicking restriction endonucleases (black triangles) releases a library of single-stranded 70-mers. (b) The unique targeting arms (red) of individual targeting oligos are designed to hybridize immediately upstream and downstream of each exon of interest. Hybridization to genomic DNA (yellow) is followed by an enzymatic gap-filling and ligation step, such that a copy of the sequence of interest is incorporated into a circle (light blue). The figure is not drawn to scale; the length of the gap-fill ranged from 60 to 191 bp. (c) Enrichment and amplification of exon-capture circles includes exonuclease digestion of linear material (not shown), linear rolling circle amplification and PCR amplification. Primers for the latter two steps are directed at the common linker sequence (dark blue). (d) To generate a shotgun sequencing library, amplicons are reccircularized via their common linker (dark blue) and subjected to hyperbranched RCA with phi29 polymerase and random hexamers. High-molecular-weight products of the hRCA reaction are randomly sheared and appended with universal adaptors to generate the shotgun sequencing library.

On page 935, 2nd column Porreca teaches genomic DNA from lyphoblastoid cell lines were hybridized to targeting oligos (MIPS via unique 20 nucleotide targeting arms), ligated and amplified and then sequenced.

Porreca does not specifically teach, "correcting for error or bias introduced during said workflow by collapsing target: differentiator tag combinations observed more than once into a single count"

However, Porreca teaches, "All evaluated positions at which the genotype is homozygous were dominated by the correct base-call (n ¼ 388). The data on heterozygous positions (n ¼ 90) are more ambiguous. For example, if we were to call positions with allele frequencies between 0.2 and 0.8 as heterozygous, then only 25 of 90 heterozygous positions (31%) would be called correctly. An alternative perspective is that because the goal of variation discovery is often to interrogate for rare variants, the key task is to identify positions that deviate from the reference genome with high sensitivity and specificity. If we call positions with observed allele frequencies of o0.8 as

likely to be mutated relative to the reference genome, we can estimate a sensitivity of 63% for detecting rare mutations while maintaining high specificity (sensitivity, 57 of 90 heterozygous positions had a reference allele frequency of less than 0.8; specificity, 0 of 313 homozygous positions had a reference allele frequency of less than 0.8)." (page 934, 1st column)

Therefore it would have been prima facie obvious to one of ordinary skill in the art at the time the invention was made to sequence the arms of the MIPS and collapse (or identify) sequences including arms with greater than 0.8 allele frequency and identify they them as homozygous. The artisan would have a reasonable expectation of success as the artisan is merely using the guidance of Porreca to identify and collapse sequences that are homozygous. The artisan would have a reasonable expectation of success as the artisan is using techniques demonstrated by the art to work.

With regards to claim 69, Porreca teaches homozygous positions were dominated by the correct base call. Thus it is obvious to collapse all sequence reads before making a base call.

The specification provides no guidance on how, "correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule" is done thus this is considered broad.

Further the specification teaches, "As used herein, a "differentiator tag sequence" is a sequence of a nucleic acid (a preparative nucleic acid), which in the context of a plurality of different isolated nucleic acids, identifies a unique, independently isolated

nucleic acid." Thus this broadly encompasses the targeting arms of a molecular inversion probe (MIP).

However, Porreca teaches, "All evaluated positions at which the genotype is homozygous were dominated by the correct base-call (n ¼ 388). The data on heterozygous positions (n ¼ 90) are more ambiguous. For example, if we were to call positions with allele frequencies between 0.2 and 0.8 as heterozygous, then only 25 of 90 heterozygous positions (31%) would be called correctly. An alternative perspective is that because the goal of variation discovery is often to interrogate for rare variants, the key task is to identify positions that deviate from the reference genome with high sensitivity and specificity. If we call positions with observed allele frequencies of o0.8 as likely to be mutated relative to the reference genome, we can estimate a sensitivity of 63% for detecting rare mutations while maintaining high specificity (sensitivity, 57 of 90 heterozygous positions had a reference allele frequency of less than 0.8; specificity, 0 of 313 homozygous positions had a reference allele frequency of less than 0.8)." (page 934, 1st column)

Therefore it would have been prima facie obvious to one of ordinary skill in the art at the time the invention was made to sequence the arms of the MIPS and recognize (or identify) sequences including arms with greater than 0.8 allele frequency and identify they them as homozygous. The artisan would motivated merely using the guidance of Porreca to recognize sequences with the same target sequences and MIP targeting arms as homozygous. The artisan would have a reasonable expectation of success as the artisan is using techniques demonstrated by the art to work.

10. Claim74 is rejected under pre-AIA 35 U.S.C. 103(a) as being unpatentable over Porreca(Nature methods (2007) volume 4, pages 931-936).as applied to claim 68-69, 76 above, and further in view of Wang (Genome Biology (2007) volume 8, R246.1-246.14).

The teachings of Porreca are set forth above and teach the detection of homozygous alleles by use of unique targeting arms of MIPS and recognizing or collapsing the combination of unique target arms and alleles (or target sequences).

Porreca does not specifically teach determination of copy number.

However, Wang teaches, “One of the techniques that have previously been described for the measurement of ACN is molecular inversion probes (MIPs) [10-12]. Briefly, MIP probes are circularizable oligonucleotides, where the two ends carry two sequences that are complementary to two sequences on the genome separated by one nucleotide (exactly where the variant to be genotyped is). After hybridization to the genomic DNA, the reaction is split into four tubes where a single nucleotide is added to each tube. Upon the addition of the nucleotide, the MIP probe is ligated closed (but this only occurs in the tube with the nucleotide that is complementary to the allele on the genome), turning the probe into a circle. This structure can be selected for by the use of exonucleases, allowing for minimal cross talk' between probes and making it possible to obtain high quality data from highly multiplexed assays (>50,000- plex)” (page 1-2).

Wang teaches, “We describe in this manuscript significant improvements we have made to the MIP-based measurements of ACN. By increasing the proportion of genomic targets that are hybridized to the MIP probes, we have improved the performance while requiring a smaller amount of DNA. Additionally, for copy number

measurements there are substantial advantages in uniformity and robustness when utilizing one-color readouts, especially at high levels of multiplexing" (page 7). Wang teaches, "These observations led us to suspect that it should be possible to genotype mixed DNA populations, such as occurs in tumor samples contaminated with normal tissue."

Therefore it would have been prima facie obvious to one of skill in the art at the time the invention was made to use MIPs to determine copy number. The artisan would be motivated as it would allow detection using less DNA and would allow genotype of mixed populations of cells to be determined. The artisan would have a reasonable expectation of success as the artisan is merely using MIPs as the art suggests.

11. Claims 71, 73 are rejected under pre-AIA 35 U.S.C. 103(a) as being unpatentable over Porreca(Nature methods (2007) volume 4, pages 931-936).as applied to claim 68-69, 76 above, and further in view of Schiffman (Journal of Clinical Oncology (2007) volume 25, p530).

The teachings of Porreca are set forth above and teach the detection of homozygous alleles by use of unique targeting arms of MIPS and recognizing or collapsing the combination of unique target arms and alleles (or target sequences).

Porreca does not specifically teach determination of blood samples or circulating tumor nucleic acids.

However, Schiffman teaches the use of MIPs for analysis of blood samples subjects with leukemia (leukemia is a circulating cancer or tumor).

Therefore it would have been prima facie obvious to one of ordinary skill in the art at the time the invention was made to use blood samples from subjects with

leukemia for examination by MIPs. The artisan would be motivated to examine gene copy number changes and LOH in blood samples to characterize the leukemia. The artisan would have a reasonable expectation of success as the artisan is using known samples in know methods.

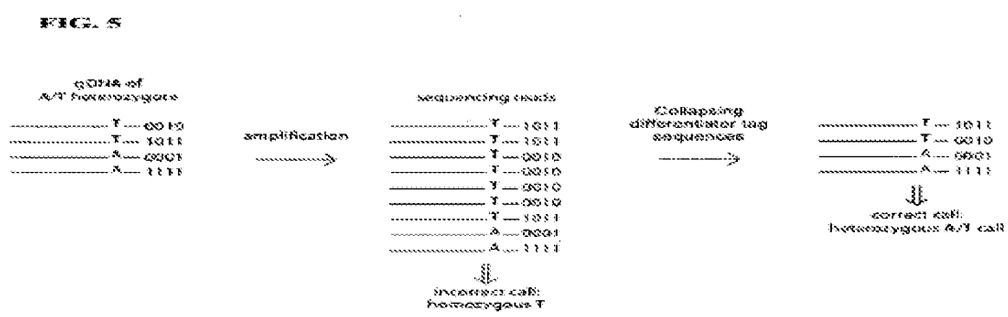
(2) Response to Argument

The brief begins arguments by noting the prosecution history of the parent case 15/231687. The parent application has been abandoned after the applicant attempted to switch invention from a method of nucleic acid identification to a method of correcting errors or bias introduced during nucleic acid analysis workflow in the response of 10/2/2020.

Applicant then filed the instant application as a continuation.

Response to 112(a) Arguments

It is noted the instant rejection is a written description rejection. As noted in the rejection the specification has antecedent basis for the limitations of the claim and a single example (figure 5) how to collapse a simple example in all the sequences with same differentiator tags have the same sequence.



However the invention encompasses a plurality of target molecules from more than one locus, such that one or more loci are associated with one for more

differentiator tag, which are then amplified, sequenced and collapsed for error correction. The specification states, "However, they are not sufficiently robust and cost-effective to be scaled up for reliable high throughput analysis of many genetic loci in large numbers of patients. The frequency of incorrect or ambiguous calls is too high for current technology to manage large numbers of patient samples without involving expensive and time-consuming steps to resolve uncertainties and provide confidence in the information output." (pages 1-2) Thus the context of the invention as provided by the specification is not the analysis of 4 differentiator tags with the sequences that vary by a single nucleotide, but thousands of loci which combine for millions of reads as evidenced by the inventors own work (Porreca (Nature methods (2007) volume 4, pages 931-936)). Thus merely providing a simple example with 4 differentiator tags, in which amplicon has the same sequence does not provide adequate written description for the analysis of thousands or more loci with sequences with which according to the specification on page 2 2nd paragraph:

> According to aspects of the invention, certain types of genetic information can be under-represented or over-represented in a genetic analysis due to a combination of stochastic variation and systematic bias in any of the preparative stages (¢.g. capture, amplification, etc.), determining stages (e.g., allele-specific detection, sequencing, etc.}, data interpretation stages (e.g., determining whether the assay information is sufficient to identify a subject as homozygous or heterozygous), and/or other stages.

Thus the office asserts the specification lacks adequate written description how "correcting for error or bias introduced during said workflow by collapsing target: differentiator tag combinations observed more than once into a single count" is done.

The brief begins asserting the rejection indicates the specification fails to provide any guidance with respect to how to collapsing is done or how collapsing corrects

errors. This argument has been thoroughly reviewed but is not considered persuasive as the brief is taking the argument out of context," The specification in figure 5 demonstrates 4 bases with one base difference in sequences. It demonstrates in a very simple example that sequences can be preferentially amplified. The specification does not any guidance on what is required of collapsed, how collapsing is done or how collapsing corrects errors. How to determine to collapse sequences of greater complexity, or what is require of a target sequence to identify it to be collapsed."

The brief continues by asserting, "Appellant notes that the terms "collapsing reads" to correct errors or bias, "determining the copy number of a genomic region" from collapsed reads, and 'recognizing target/differentiator tags having the same sequence as being derived from the same input molecule" are disclosed and fully-supported in the as-filed specification." This argument has been thoroughly reviewed but is not considered persuasive as while the specification provides antecedent basis, it does not describe how to "correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count."

The brief continues on page 10 to point to support in the parent application on page 79, where it describes figure 5 and indicates that 4 loci with 4 different differentiator tags and apparently 1 of 2 alleles is amplified and then "collapsed" into 4 sequences. This as detailed above does not adequately describe how to deal with more complex sequences with more variations due to starting material, capture, infidelity of the polymerase or tag attachment etc. Further it does not provide any detail on how this corrects errors or even which type of error would be corrected.

The brief continues by jumping to page 6 which states:

In some embodiments, the differentiator tag sequence permit the detection of bias based on the occurrence of combinations of differentiator tag and target sequences observed in a sequencing reaction. In other embodiments, the methods reduce errors caused by bias, or the risk of bias, associated with the capture, amplification and sequencing of genetic loci, e.g., for diagnostic purposes. Pg. 6, lines 21-28

The cited portion discusses the use differentiator tags to reduce errors or correct bias, but does not explain how to correcting sequence reads or bias introduced said workflow is corrected by collapsing target:differentiator tags is done.

The brief continues by reproducing the next portion of page 6, lines 29 to top of page 7. The cited portion states, "This information can be used to help analyze a threshold number of independently isolated target nucleic acids from a biological sample in order to obtain sequence information that is reliable and can be used to make a genotype conclusion (e.g., call) with a desired degree of confidence." However this does not teach how in complex samples examining a plurality of loci each with one or more differentiator tag how correcting for error or bias introduced during said workflow by collapsing target: differentiator tag combinations observed more than once into a single count is done.

The brief continues by jumping from page 6 to page 27 to the top of page 28 which states:

"Aspects of the invention are based, in part, on the identification of different sources of ambiguity and error in genetic analyses, and, in part, on the identification of one or more approaches to avoid, reduce, recognize, and/or resolve these errors and ambiguities at different stages in a genetic analysis. Aspects of the invention relate to methods and compositions for addressing bias and/or stochastic variation associated with one or more preparative and/or analytical steps of a nucleic acid evaluation technology."

This appears to disclose a wish list of what the inventor and/or writer of the application hopes the invention can accomplish, but does not provide adequate written

description to one of ordinary skill in the art at the time the invention was made to correcting for error or bias introduced during said workflow by collapsing target: differentiator tag combinations observed more than once into a single count is done.

Thus the brief, similar to the rejection, indicates the specification has support for the concept of correcting errors or biases, but does not teach one of skill in the how such correction is done. The brief is essentially arguing we have explained what we want the last step of the claim to do, but has not adequately described how the limitation is actually done in fashion consistent with the complexity of data the specification demonstrates is common to the art.

The brief then jumps to determining copy number variation for claim 74. The brief cites page 58, lines 3-9. The cited and bolded portion provides antecedent basis for the limitation of genomic copy numbers. However the cited portion does not indicate how copy number variation would be done using differentiator tags and target sequences.

The brief then jumps to page 7 of the specification which states:

Accordingly, in some aspects, methods of genotyping a subject are provided. In some embodiments, the methods comprise determining the sequence of at least a threshold number of independently isolated nucleic acids, wherein the sequence of each isolated nucleic acid comprises a target nucleic acid sequence and a differentiator tag sequence, wherein the threshold number is a number of unique combinations of target nucleic acid and differentiator tag sequences, wherein the isolated nucleic acids are identified as independently isolated if they comprise unique combinations of target nucleic acid and differentiator tag sequences, and wherein the target nucleic acid sequence is the sequence of a genomic locus of a subject.

However the cite portion of page 7 merely discusses unique combination of target nucleic acids and differentiator tags. Thus the context suggest the brief is arguing that each unique target: differentiator tag could be a copy number variation. However the claims encompass a biological sample, which could be hundreds if not

millions of copies of the genome and each loci has one or more differentiator tags. Thus it is unclear how this implication would work with possibly millions of copies of a loci each with a different differentiator tag, making each a unique target: differentiator tag combination.

The brief continues by asserting there is clear guidance from examples 2 and 3 and figures 5 and 6. The arguments with respect to figure 5 is not persuasive for the reasons of record.

With regards to example 5, the specification teaches it is "**MIP capture of "target locus and "control" loci.**" First it is noted the claims do not recite control loci. Thus the mere title suggests different scope than the claims. Example 5 begins with discussion of target loci length and abundance is digitally measured by sequencing. Again, the claims do not expressly recite limitations consistent with these teachings. Example 5 continues by discussing abundance vs length, calibration and statistical representation. However the example provides no clear nexus with respect to this teachings and correcting for error or bias introduced during said workflow by collapsing target: differentiator tag combinations observed more than once into a single count. Most of the specification real estate (pages) of example 5 are dedicated to MIP assays. While MIP assays are encompassed by the claim, they are not required.

Likewise the teachings of example 6 are limited to reaction conditions for MIPS and do not provide how correcting for error or bias introduced during said workflow by collapsing target: differentiator tag combinations observed more than once into a single count is done.

The brief continues by asserting the Office erred in refusing to give weight to the technical opinions and factual findings of the Porreca declaration. The brief continues by specifically pointing to paragraph 4 and 6 of the declaration.

In paragraph 4 inventor Porreca states, "The patent application discloses tagging individual target nucleic acid molecules with differentiator tags before any amplification or sequencing steps. Id., pg. 6 (21-24); pg. 40 (7-9); pg. 42 (14-27). Each differentiator tag has its own unique sequence so that each individual target nucleic acid molecule is tagged, and therefore associated, with a differentiator tag. Id., pg. 41 (10-22). Accordingly, each target: differentiator tag combination uniquely identifies a single nucleic acid molecule through subsequence amplification and sequencing steps. Id., pg. 39 (21-26); pg. 41 (3-7), (9-10); pg. 42." This argument has been thoroughly reviewed but is not considered persuasive as the pending claims recite, "obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin; introducing a set of differentiator tags, wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag." Thus the claims encompass a loci associated with **one** or more differentiator tag. The specification provides no limiting definition of loci and thus the broadest reasonable interpretation encompasses a region and thus is not necessarily limited to a single nucleic acid sequence. Thus the assertions are not consistent with the claims.

Paragraph 4 of continues by providing the teachings of simplified example of figure 5 of the instant application into the inventor's terms.

The declaration in paragraph again asserts, " A distinguishing feature of the claims in my patent application is that each individual nucleic acid molecule is tagged with a unique differentiator tag prior to any amplification or sequencing steps." his argument has been thoroughly reviewed but is not considered persuasive as the pending claims recite, "obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin; introducing a set of differentiator tags, wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag." Thus the claims encompass a loci associated with **one** or more differentiator tag. The specification provides no limiting definition of loci and thus the broadest reasonable interpretation encompasses a region and thus is not necessarily limited to a single nucleic acid sequence. Thus the assertions are not consistent with the claims.

The brief continues by asserting the Office dismisses the "expert conclusions" (opinions) of the inventor. This argument has been thoroughly reviewed but is not considered persuasive as the Final action stated,

> In paragraph five the declaration begins by providing the inventors opinion on the teachings of Porreca(Nature methods (2007) volume 4, pages 931-936)by providing applicant's opinion that the differentiator tags of the instant claims are different than the MIPS of Porreca (Nature methods (2007) volume 4, pages 931-936). This argument has been thoroughly reviewed but is not considered persuasive as the specification teaches, "As used herein, a "differentiator tag sequence" is a sequence of a nucleic acid (a preparative nucleic acid), which in the context of a plurality of different isolated nucleic acids, identifies a unique, independently isolated nucleic acid."
>
> The declaration continues by asserting each individual nucleic acid is tagged with a unique differentiator tag. This argument has been thoroughly reviewed but is not considered persuasive as this is not a limitation of the claim. Although the claims are interpreted in light of the specification, limitations from the specification are not read into the claims. See *In re Van Geuns*, 988 F.2d 1181, 26 USPQ2d 1057 (Fed. Cir. 1993).

The declaration continues by providing the inventors opinion that the teachings of Porreca (Nature methods) does not envision the differentiator tags of the claims. This argument has been thoroughly reviewed but considered persuasive as the specification teaches, "As used herein, a "differentiator tag sequence" is a sequence of a nucleic acid (a preparative nucleic acid), which in the context of a plurality of different isolated nucleic acids, identifies a unique, independently isolated nucleic acid." Thus this broadly encompasses the targeting arms of a molecular inversion probe (MIP).

The declaration continues by asserting, "Because the targeting arms of the molecular inversion probes disclosed in my 2007 Nature Methods publication will hybridize to the flanking sequences of all copies of the gene, the unique targeting arms cannot be used to differentiate between the identical copies of the gene as separate nucleic acid molecules." This argument has been thoroughly reviewed but is not considered persuasive as the claims do not specifically require the differentiator tag be used to differentiate between the identical copies of the gene as separate nucleic acid molecules.

Thus the final rejection (Response to Amendment) does not merely disregard the inventor's interpretation of the claim or art, but distinctly explains the differences between inventor's opinion and the claims and specification.

Further MPEP 716.01(c) states:

III. OPINION EVIDENCE

Although factual evidence is preferable to opinion testimony, such testimony is entitled to consideration and some weight so long as the opinion is not on the ultimate legal conclusion at issue. While an opinion as to a legal conclusion is not entitled to any weight, the underlying basis for the opinion may be persuasive. *In re Chilowsky,* 306 F.2d 908, 134 USPQ 515 (CCPA 1962) (expert opinion that an application meets the requirements of 35 U.S.C. 112 is not entitled to any weight; however, facts supporting a basis for deciding that the specification complies with 35 U.S.C. 112 are entitled to some weight); *In re Lindell,* 385 F.2d 453, 155 USPQ 521 (CCPA 1967) (Although an affiant's or declarant's opinion on the ultimate legal issue is not evidence in the case, "some weight ought to be given to a persuasively supported statement of one skilled in the art on what was not obvious to him." 385 F.2d at 456, 155 USPQ at 524 (emphasis in original)).

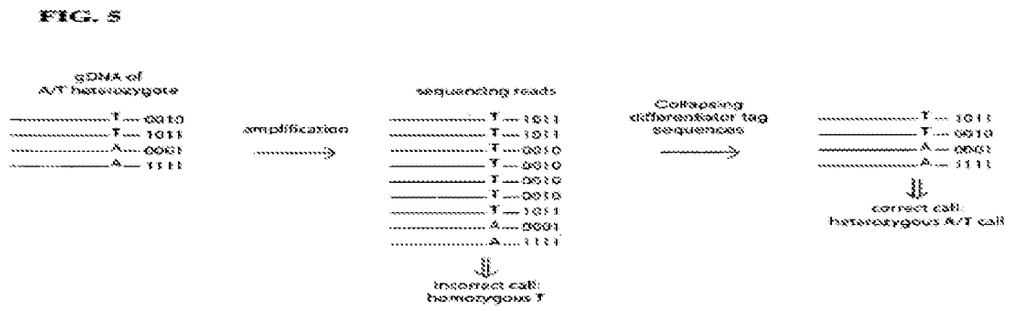
In assessing the probative value of an expert opinion, the examiner must consider the nature of the matter sought to be established, the strength of any opposing evidence, the interest of the expert in the outcome of the case, and the presence or absence of factual support for the expert's opinion. *Ashland Oil, Inc. v. Delta Resins & Refractories, Inc.,* 776 F.2d 281, 227 USPQ 657 (Fed. Cir. 1985), *cert. denied,* 475 U.S. 1017 (1986). See also *In re Oelrich,* 579 F.2d 86, 198 USPQ 210 (CCPA 1978) (factually based expert opinions on the level of ordinary skill in the art were sufficient to rebut the *prima facie* case of obviousness); *Ex parte Gray,* 10 USPQ2d 1922 (Bd. Pat. App. & Inter. 1989) (statement in publication dismissing the "preliminary identification of a human b-NGF-like molecule" in the prior art, even if considered to be an expert opinion, was inadequate to overcome the rejection based on that prior art because there was no factual evidence supporting the statement); *In re Carroll,* 601 F.2d 1184, 202 USPQ 571 (CCPA 1979) (expert opinion on what the prior art taught, supported by documentary evidence and formulated prior to the making of the claimed invention, received considerable deference); *In re Beattie,* 974 F.2d 1309, 24 USPQ2d 1040 (Fed. Cir. 1992) (declarations of seven persons skilled in the art offering opinion evidence praising the merits of the claimed invention were found to have little value because of a lack of factual support); *Ex parte George*, 21 USPQ2d 1058 (Bd. Pat. App. & Inter.

1991) (conclusory statements that results were "unexpected," unsupported by objective factual evidence, were considered but were not found to be of substantial evidentiary value).

In the instant case the declaration is the opinion of the inventor as it provides nothing other than the inventors opinion, but provides no evidence other than relying on parts of the specification which are not required of the claim as detailed previous. .

Response to 112(b) Arguments

It is noted the instant rejection is a 112(b) rejection. As noted in the rejection the specification has antecedent basis for the limitations of the claim and a single example (figure 5) how to collapse a simple example in all the sequences with same differentiator tags have the same sequence.



However the invention encompasses a plurality of target molecules from more than one locus, such that one or more loci are associated with one for more differentiator tag, which are then amplified, sequenced and collapsed for error correction. The specification states, “However, they are not sufficiently robust and cost-effective to be scaled up for reliable high throughput analysis of many genetic loci in large numbers of patients. The frequency of incorrect or ambiguous calls is too high for current technology to manage large numbers of patient samples without involving expensive and time-consuming steps to resolve uncertainties and provide confidence in

the information output." (pages 1-2) Thus the context of the invention as provided by the specification is not the analysis of 4 differentiator tags with the sequences that vary by a single nucleotide, but thousands of loci which combine for millions of reads as evidenced by the inventors own work (Porreca(Nature methods (2007) volume 4, pages 931-936)). Thus merely providing a simple example with 4 differentiator tags, in which amplicon has the same sequence does not provide adequate written description for the analysis of thousands or more loci with sequences with which according to the specification on page 2 2nd paragraph:

> According to aspects of the invention, certain types of genetic information can be under-represented or over-represented in a genetic analysis due to a combination of stochastic variation and systematic bias in any of the preparative stages (¢.g. capture, amplification, etc.), determining stages (e.g., allele-specific detection, sequencing, etc.}, data interpretation stages (e.g., determining whether the assay information is sufficient to identify a subject as homozygous or heterozygous), and/or other stages.

Thus the office asserts the metes and bounds of how "correcting for error or bias introduced during said workflow by collapsing target: differentiator tag combinations observed more than once into a single count" is done.

The brief begins asserting the rejection indicates the specification fails to provide any guidance with respect to how to collapsing is done or how collapsing corrects errors. This argument has been thoroughly reviewed but is not considered persuasive as the brief is taking the argument out of context," The specification in figure 5 demonstrates 4 bases with one base difference in sequences. It demonstrates in a very simple example that sequences can be preferentially amplified. The specification does not any guidance on what is required of collapsed, how collapsing is done or how

collapsing corrects errors. How to determine to collapse sequences of greater complexity, or what is require of a target sequence to identify it to be collapsed."

The brief continues by asserting, "Appellant notes that the terms "collapsing reads" to correct errors or bias, "determining the copy number of a genomic region" from collapsed reads, and 'recognizing target/differentiator tags having the same sequence as being derived from the same input molecule" are disclosed and fully-supported in the as-filed specification." This argument has been thoroughly reviewed but is not considered persuasive as while the specification provides antecedent basis, it does not describe how to "correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count."

The brief continues on page 12 to point to support in the parent application on page 79, where it describes figure 5 and indicates that 4 loci with 4 different differentiator tags and apparently 1 of 2 alleles is amplified and then "collapsed" into 4 sequences. This as detailed above does not adequately perform the collapsing step with more complex sequences with more variations due to starting material, capture, infidelity of the polymerase or tag attachment etc. Further it does not provide any detail on how this corrects errors or even which type of error would be corrected.

The brief (page 12-13) continues by jumping to page 6 which states:

In some embodiments, the differentiator tag sequence permit the detection of bias based on the occurrence of combinations of differentiator tag and target sequences observed in a sequencing reaction. In other embodiments, the methods reduce errors caused by bias, or the risk of bias, associated with the capture, amplification and sequencing of genetic loci, e.g., for diagnostic purposes. Pg. 6, lines 21-28

The cited portion discusses the use differentiator tags to reduce errors or correct bias, but does not explain how to correcting sequence reads or bias introduced said workflow is corrected by collapsing target:differentiator tags is done.

The brief (page 13) continues by reproducing the next portion of page 6, lines 29 to top of page 7. The cited portion states, "This information can be used to help analyze a threshold number of independently isolated target nucleic acids from a biological sample in order to obtain sequence information that is reliable and can be used to make a genotype conclusion (e.g., call) with a desired degree of confidence." However this does not teach how in complex samples examining a plurality of loci each with one or more differentiator tag how correcting for error or bias introduced during said workflow by collapsing target: differentiator tag combinations observed more than once into a single count is done.

The brief (page 13) continues by jumping from page 6 to page 27 to the top of page 28 which states:

"Aspects of the invention are based, in part, on the identification of different sources of ambiguity and error in genetic analyses, and, in part, on the identification of one or more approaches to avoid, reduce, recognize, and/or resolve these errors and ambiguities at different stages in a genetic analysis. Aspects of the invention relate to methods and compositions for addressing bias and/or stochastic variation associated with one or more preparative and/or analytical steps of a nucleic acid evaluation technology."

This appears to disclose a wish list of what the inventor and/or writer of the application hopes the invention can accomplish, but does not how to accomplish these lofty goals to one of ordinary skill in the art at the time the invention was made to correcting for error or bias introduced during said workflow by collapsing target: differentiator tag combinations observed more than once into a single count is done.

Thus the brief, similar to the rejection, indicates the specification has support for the concept of correcting errors or biases, but does not teach one of skill in the how such correction is done. The brief is essentially arguing we have explained what we want the last step of the claim to do, but has not adequately described how the limitation is actually done in fashion consistent with the complexity of data the specification demonstrates is common to the art.

The brief (page 13, middle) then jumps to determining copy number variation for claim 74. The brief cites page 58, lines 3-9. The cited and bolded portion provides antecedent basis for the limitation of genomic copy numbers. However the cited portion does not indicate how copy number variation would be done using differentiator tags and target sequences.

The brief then jumps to page 7 of the specification which states:

Accordingly, in some aspects, methods of genotyping a subject are provided. In some embodiments, the methods comprise determining the sequence of at least a threshold number of independently isolated nucleic acids, wherein the sequence of each isolated nucleic acid comprises a target nucleic acid sequence and a differentiator tag sequence, wherein the threshold number is a number of unique combinations of target nucleic acid and differentiator tag sequences, wherein the isolated nucleic acids are identified as independently isolated if they comprise unique combinations of target nucleic acid and differentiator tag sequences, and wherein the target nucleic acid sequence is the sequence of a genomic locus of a subject.

However the cite portion of page 7 merely discusses unique combination of target nucleic acids and differentiator tags. Thus the context suggest the brief is arguing that each unique target: differentiator tag could be a copy number variation. However the claims encompass a biological sample, which could be hundreds if not millions of copies of the genome and each loci (as broad as that is) has one or more differentiator tags. Thus it is unclear how this implication would work with possibly millions of copies of a loci each with a different differentiator tag, making each a unique

target: differentiator tag combination. Thus it would appear the copy number would be limited to only the number of unique differentiator tag:target sequence combination, not the number in the cell.

The brief continues by reproducing 2173.02 (III)(B), which the brief mislabeled 2173(II)(B) and asserting the rejection has dismissed the language of the specification and deprived the applicant of an opportunity to respond to the rejection. This argument has been thoroughly reviewed but is not considered persuasive as the rejection was first presented in the non-final of 2/10/2021. The response of 3/5/2021 provided arguments that were labeled 112(b), but equally appeared to be directed to the 112(a). Thus the examiner refuted the arguments in detail in response to arguments of the written description 112(a) pages 8-10 of the final mailed 4/9/2021.

The brief continues by asserting the Office erred in refusing to give weight to the technical opinions and factual findings of the Porreca declaration. The brief continues by specifically pointing to paragraph 4 and 6 of the declaration.

In paragraph 4 inventor Porreca states, “The patent application discloses tagging individual target nucleic acid molecules with differentiator tags before any amplification or sequencing steps. Id., pg. 6 (21-24); pg. 40 (7-9); pg. 42 (14-27). Each differentiator tag has its own unique sequence so that each individual target nucleic acid molecule is tagged, and therefore associated, with a differentiator tag. Id., pg. 41 (10-22). Accordingly, each target:differentiator tag combination uniquely identifies a single nucleic acid molecule through subsequence amplification and sequencing steps. Id., pg. 39 (21-26); pg. 41 (3-7), (9-10); pg. 42.” This argument has been thoroughly reviewed but is not considered persuasive as the pending claims recite, “obtaining a biological

sample comprising a plurality of target nucleic acid molecules from more than one locus of origin; introducing a set of differentiator tags, wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag." Thus the claims encompass a loci associated with **one** or more differentiator tag. The specification provides no limiting definition of loci and thus the broadest reasonable interpretation encompasses a region and thus is not necessarily limited to a single nucleic acid sequence. Thus the assertions are not consistent with the claims.

Paragraph 4 of continues by providing the teachings of simplified example of figure 5 of the instant application into the inventor's terms.

The declaration in paragraph again asserts, " A distinguishing feature of the claims in my patent application is that each individual nucleic acid molecule is tagged with a unique differentiator tag prior to any amplification or sequencing steps." his argument has been thoroughly reviewed but is not considered persuasive as the pending claims recite, "obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin; introducing a set of differentiator tags, wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag." Thus the claims encompass a loci associated with **one** or more differentiator tag. The specification provides no limiting definition of loci and thus the broadest reasonable interpretation encompasses a region and thus is not necessarily limited to a single nucleic acid sequence. Thus the assertions are not consistent with the claims.

The brief continues by asserting the Office dismisses the " expert conclusions" (opinions) of the inventor. This argument has been thoroughly reviewed but is not considered persuasive as the Final action stated,:

In paragraph five the declaration begins by providing the inventors opinion on the teachings of Porreca(Nature methods (2007) volume 4, pages 931-936)by providing applicant's opinion that the differentiator tags of the instant claims are different than the MIPS of Porreca (Nature methods (2007) volume 4, pages 931-936). This argument has been thoroughly reviewed but is not considered persuasive as the specification teaches, "As used herein, a "differentiator tag sequence" is a sequence of a nucleic acid (a preparative nucleic acid), which in the context of a plurality of different isolated nucleic acids, identifies a unique, independently isolated nucleic acid."

The declaration continues by asserting each individual nucleic acid is tagged with a unique differentiator tag. This argument has been thoroughly reviewed but is not considered persuasive as this is not a limitation of the claim. Although the claims are interpreted in light of the specification, limitations from the specification are not read into the claims. See *In re Van Geuns*, 988 F.2d 1181, 26 USPQ2d 1057 (Fed. Cir. 1993).

The declaration continues by providing the inventors opinion that the teachings of Porreca (Nature methods) does not envision the differentiator tags of the claims. This argument has been thoroughly reviewed but considered persuasive as the specification teaches, "As used herein, a "differentiator tag sequence" is a sequence of a nucleic acid (a preparative nucleic acid), which in the context of a plurality of different isolated nucleic acids, identifies a unique, independently isolated nucleic acid." Thus this broadly encompasses the targeting arms of a molecular inversion probe (MIP).

The declaration continues by asserting, "Because the targeting arms of the molecular inversion probes disclosed in my 2007 Nature Methods publication will hybridize to the flanking sequences of all copies of the gene, the unique targeting arms cannot be used to differentiate between the identical copies of the gene as separate nucleic acid molecules." This argument has been thoroughly reviewed but is not considered persuasive as the claims do not specifically require the differentiator tag be used to differentiate between the identical copies of the gene as separate nucleic acid molecules.

Thus the final rejection (Response to Amendment) does not merely disregard the inventor's interpretation of the claim or art, but distinctly explains the differences between inventor's opinion and the claims and specification.

Further MPEP 716.01(c) states:

III. OPINION EVIDENCE

Although factual evidence is preferable to opinion testimony, such testimony is entitled to consideration and some weight so long as the opinion is not on the ultimate legal conclusion at issue. While an opinion as to a legal conclusion is not entitled to any weight, the underlying basis for the opinion may be

persuasive. *In re Chilowsky,* 306 F.2d 908, 134 USPQ 515 (CCPA 1962) (expert opinion that an application meets the requirements of **35 U.S.C. 112** is not entitled to any weight; however, facts supporting a basis for deciding that the specification complies with **35 U.S.C. 112** are entitled to some weight); *In re Lindell,* 385 F.2d 453, 155 USPQ 521 (CCPA 1967) (Although an affiant's or declarant's opinion on the ultimate legal issue is not evidence in the case, "some weight ought to be given to a persuasively supported statement of one skilled in the art on what was not obvious to him." 385 F.2d at 456, 155 USPQ at 524 (emphasis in original)).

In assessing the probative value of an expert opinion, the examiner must consider the nature of the matter sought to be established, the strength of any opposing evidence, the interest of the expert in the outcome of the case, and the presence or absence of factual support for the expert's opinion. *Ashland Oil, Inc. v. Delta Resins & Refractories, Inc.,* 776 F.2d 281, 227 USPQ 657 (Fed. Cir. 1985), *cert. denied,* 475 U.S. 1017 (1986). See also *In re Oelrich,* 579 F.2d 86, 198 USPQ 210 (CCPA 1978) (factually based expert opinions on the level of ordinary skill in the art were sufficient to rebut the *prima facie* case of obviousness); *Ex parte Gray,* 10 USPQ2d 1922 (Bd. Pat. App. & Inter. 1989) (statement in publication dismissing the "preliminary identification of a human b-NGF-like molecule" in the prior art, even if considered to be an expert opinion, was inadequate to overcome the rejection based on that prior art because there was no factual evidence supporting the statement); *In re Carroll,* 601 F.2d 1184, 202 USPQ 571 (CCPA 1979) (expert opinion on what the prior art taught, supported by documentary evidence and formulated prior to the making of the claimed invention, received considerable deference); *In re Beattie,* 974 F.2d 1309, 24 USPQ2d 1040 (Fed. Cir. 1992) (declarations of seven persons skilled in the art offering opinion evidence praising the merits of the claimed invention were found to have little value because of a lack of factual support); *Ex parte George*, 21 USPQ2d 1058 (Bd. Pat. App. & Inter. 1991) (conclusory statements that results were "unexpected," unsupported by objective factual evidence, were considered but were not found to be of substantial evidentiary value).

In the instant case the declaration is the opinion of the inventor as it provides nothing other than the inventors opinion, but provides no evidence other than relying on parts of the specification which are not required of the claim as detailed previous. .

Response to 101 Argument's

The brief (page 15) (4.4) begins by asserting the Office erred in the 101 rejection.

The brief continues by citing Diamond v. Chakrabarty, 447 U.S. 303, 308-309 (1980), Mayo Collaborative Servs. v. Prometheus Labs., Inc., 556 U.S. 66 (2012); Alice Corp. v. CLS Bank Int'l, 573 U.S. 208 (2014). These are noted.

The brief continues by noting, "The USPTO's 2019 Revised Patent Subject Matter Eligibility Guidance (the "Revised Guidance") explains that the judicial category of mental concepts seeks to exclude from patentability any process that would include an individual performing the claimed steps mentally or with a pencil and paper.." It is

noted that in response to the 112 (a) and 112(b) the brief arguments the mental step of the claim can be done as depicted in figure 5. Thus the mental step of the claim can clearly be done in the mind or with pencil and paper.

The response continues by asserting the active steps of claims cannot be done in the mind. This argument has been thoroughly reviewed but is not considered persuasive as specification teaches:

> Sample preparative methods may fall into three classes: 1) single- or several-target amplification (e.g., uniplex PCR, 'multiplex' PCR), 2) multi-target hybridization enrichment (e.g., Agilent SureSelect 'hybrid capture' [Gnirke et al 2009, Nature methods 27:182-9], Roche/Nimblegen 'sequence capture' [Hodges et al 2007, Nature genetics 39:1522-7], and 3) multi-target circularization selection (e.g. molecular inversion probes or padlock probes, [Porreca et al 2007, Nature methods 4:931-6, Turner et al 2009, Nature methods 6:315-6], 'selectors' [Dahl et al 2005, Nucleic acids research 33:e71]). arrays, single nucleotide polymorphism (SNP) arrays (Affymetrix, Santa Clara, Calif.).

Further the prior art of record in the case demonstrates the active steps of the claim are routine and conventional necessary to provide the data for the mental step of the claim. The record and provided no evidence to the contrary.

The brief continues by noting the "USPTO's October 2019 Subject Matter Eligibility Update ("October Update"') explains that a claim is not "directed to" a judicial exception, and is thus patent eligible, if the judicial exception is integrated into a practical application of the judicial exception." This argument has been thoroughly reviewed but is not considered persuasive as the preamble of the claim recites, "method for correcting for errors or bias introduced during nucleic acid analysis workflow" and the last active step of the claim recites, "correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count." Thus the claims are directed to a mental step and provide

additional routine and conventional steps. The claims have no further limitation which integrate the correcting errors and there is no evidence the active steps are anything more than routine and conventional.

The response continues providing arguments that were not previously presented to the 101 rejection with respect to claim 85 of May 2016 USPTO guidelines (it is noted the representative again misidentified the citation). AS indicated claim 85 provides a single active step of amplifying, which is a different fact pattern of the instant claims which provide the mental step or abstract idea of "correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count." Thus the fact pattern is different and instant claims clearly provide a judicial exception in the correcting step.

The brief again tries to bring up the prosecution of the parent application. AS the Board knows each and every application is examined on its own merits. Further the claims of the parent application merely required identifying a target sequence from an amplified nucleic acid that had been sequenced with differentiator tags (claims 2/18/2020). Thus the last step merely restated what appears to be an inherent property of the sequencing of amplicons.

Response to 103 Argument's

The briefs continues by traversing the art rejection by providing the representatives opinion of the rejection and asserting that the art does not teach every limitation of the claim. This argument has been thoroughly reviewed but is not considered persuasive as the limitations at issue are, ""differentiator tag sequence" and "correcting for error or bias introduced during said workflow by collapsing

target:differentiator tag combinations observed more than once into a single count." The record clearly demonstrates the correcting limitations lacks adequate written description and the metes and bounds are unclear. Further, the specification teaches, "As used herein, a "differentiator tag sequence" is a sequence of a nucleic acid (a preparative nucleic acid), which in the context of a plurality of different isolated nucleic acids, identifies a unique, independently isolated nucleic acid." Thus this broadly encompasses the targeting arms of a molecular inversion probe (MIP).

The brief continues by asserting the rejection, "concedes that Porreca 2007 does not teach "correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count." "(The brief asserts this is on page 12 of the final rejection , which is confusing as page 12 is part of the 101 rejection).

However, the rejection states, "Porreca does not **specifically** teach, "correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count"" (page 16 final rejection).

The rejection continues by stating, "Porreca teaches, "All evaluated positions at which the genotype is homozygous were dominated by the correct base-call (n = 388). The data on heterozygous positions (n = 90) are more ambiguous. For example, if we were to call positions with allele frequencies between 0.2 and 0.8 as heterozygous, then only 25 of 90 heterozygous positions (31%) would be called correctly. An alternative perspective is that because the goal of variation discovery is often to interrogate for rare variants, the key task is to identify positions that deviate from the reference genome with high sensitivity and specificity. If we call positions with observed allele frequencies

of o0.8 as likely to be mutated relative to the reference genome, we can estimate a sensitivity of 63% for detecting rare mutations while maintaining high specificity (sensitivity, 57 of 90 heterozygous positions had a reference allele frequency of less than 0.8; specificity, 0 of 313 homozygous positions had a reference allele frequency of less than 0.8)." (page 934, 1st column)." The teachings of Porreca precedes this with," With an Illumina Genome Analyzer, we obtained 125 Mb of sequence that could be confidently aligned to targeted exon sequences. There were ~ 500 kb of genomic positions for which we obtained tenfold or greater coverage. To evaluate the accuracy of these resequencing data, we focused on 478 validated HapMap SNPs, genotyped in the individual from whose DNA the sequencing library was generated (GM12248) that overlapped with positions for which we had generated 10 times or greater coverage (Fig. 4)." Thus one of skill in the art recognizes that multiple reads of MIPS which encompass differentiator tags and target sequences are being analyzed by sequencing and sequences with the same sequence and same allele at the position of interest are condensed (or collapsed) to a single read (homozygous), while those reads with 0.2 and 0.8 frequencies were collapsed to heterozygous.

The response further concedes that Porreca recognizes that issues with PCR amplification on a nonuniformity error on "pge4 discussion." The teachings of Porreca are on pages 931-934, thus the comment is a little confusing. However, on page 934, 2nd column, Porreca teaches issues with PCR amplification uniformity which are corrected or at least partially addressed by, " evaluated positions at which the genotype is homozygous were dominated by the correct base-call (n = 388). The data on heterozygous positions (n = 90) are more ambiguous. For example, if we were to call

positions with allele frequencies between 0.2 and 0.8 as heterozygous, then only 25 of 90 heterozygous positions (31%) would be called correctly" (Porreca (2007, page 934, 1st column). .

The brief continues by referencing the declaration of Porreca. The brief argues by providing the opinion of inventor Porreca by stating, "As Dr. Porreca points out, this is because each molecular inversion probe is designed to uniquely capture an exon of interest with targeting arms that hybridize to sequences flanking the exon. /d., para. 7. The targeting arms are unique only in the sense that PCR primers are unique; unique designs each synthesized in the millions (this point is settled where Supplementary Figure 1 in Porreca 2007 states that the 480 probes were used at a "concentration of ~167 pM for each species"). These arguments have been thoroughly reviewed but is not considered persuasive as brief and declaration concedes the each molecular probe is unique. The brief and declaration continues by trying to differentiate differentiator proves from the unique sequences of the molecular inversion probe are unique but do not differentiate separate single molecules (top of page 20). This argument has been thoroughly reviewed but is not considered persuasive as the claims do not require a different tag on every single molecule as asserted. The claims recite, "wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag." Thus one or more loci (not every single molecule) are associated with more than one differentiator tag. Thus the brief and declaration are arguing limitations not in the claim.

The brief continues by asserting Porreca (inventor) in the declaration provides his expert opinion his prior art does not render obvious his narrow interpretation of the claims. First, MPEP 716.01(c) states:

III. OPINION EVIDENCE

Although factual evidence is preferable to opinion testimony, such testimony is entitled to consideration and some weight so long as the opinion is not on the ultimate legal conclusion at issue. While an opinion as to a legal conclusion is not entitled to any weight, the underlying basis for the opinion may be persuasive. *In re Chilowsky,* 306 F.2d 908, 134 USPQ 515 (CCPA 1962) (expert opinion that an application meets the requirements of **35 U.S.C. 112** is not entitled to any weight; however, facts supporting a basis for deciding that the specification complies with **35 U.S.C. 112** are entitled to some weight); *In re Lindell,* 385 F.2d 453, 155 USPQ 521 (CCPA 1967) (Although an affiant's or declarant's opinion on the ultimate legal issue is not evidence in the case, "some weight ought to be given to a persuasively supported statement of one skilled in the art on what was not obvious to him." 385 F.2d at 456, 155 USPQ at 524 (emphasis in original)).

In assessing the probative value of an expert opinion, the examiner must consider the nature of the matter sought to be established, the strength of any opposing evidence, the interest of the expert in the outcome of the case, and the presence or absence of factual support for the expert's opinion. *Ashland Oil, Inc. v. Delta Resins & Refractories, Inc.,* 776 F.2d 281, 227 USPQ 657 (Fed. Cir. 1985), *cert. denied,* 475 U.S. 1017 (1986). See also *In re Oelrich,* 579 F.2d 86, 198 USPQ 210 (CCPA 1978) (factually based expert opinions on the level of ordinary skill in the art were sufficient to rebut the *prima facie* case of obviousness); *Ex parte Gray,* 10 USPQ2d 1922 (Bd. Pat. App. & Inter. 1989) (statement in publication dismissing the "preliminary identification of a human b-NGF-like molecule" in the prior art, even if considered to be an expert opinion, was inadequate to overcome the rejection based on that prior art because there was no factual evidence supporting the statement); *In re Carroll,* 601 F.2d 1184, 202 USPQ 571 (CCPA 1979) (expert opinion on what the prior art taught, supported by documentary evidence and formulated prior to the making of the claimed invention, received considerable deference); *In re Beattie,* 974 F.2d 1309, 24 USPQ2d 1040 (Fed. Cir. 1992) (declarations of seven persons skilled in the art offering opinion evidence praising the merits of the claimed invention were found to have little value because of a lack of factual support); *Ex parte George*, 21 USPQ2d 1058 (Bd. Pat. App. & Inter. 1991) (conclusory statements that results were "unexpected," unsupported by objective factual evidence, were considered but were not found to be of substantial evidentiary value).

Further, although the claims are interpreted in light of the specification, limitations from the specification are not read into the claims. See *In re Van Geuns*, 988 F.2d 1181, 26 USPQ2d 1057 (Fed. Cir. 1993).

The brief continues by arguing the Porreca states his opinion Schiffman does not teach differentiator tags. However, the opinion of Porreca does not address the specification teaches, "As used herein, a "differentiator tag sequence" is a sequence of a nucleic acid (a preparative nucleic acid), which in the context of a plurality of

different isolated nucleic acids, identifies a unique, independently isolated nucleic acid." Thus this broadly encompasses the targeting arms of a molecular inversion probe (MIP). Thus the inventor is attempting to assert he does not like the broadest reasonable interpretation of the claims in view of the specification.

For the above reasons, it is believed that the rejections should be sustained.

Respectfully submitted,

/Steven Pohnert/
Primary Examiner, Art Unit 1634

Conferees:

/Ashwin Mehta/

Primary Examiner

/DAVE T NGUYEN/
Supervisory Patent Examiner, Art Unit 1634

Requirement to pay appeal forwarding fee. In order to avoid dismissal of the instant appeal in any application or ex parte reexamination proceeding, 37 CFR 41.45 requires payment of an appeal forwarding fee within the time permitted by 37 CFR 41.45(a), unless appellant had timely paid the fee for filing a brief required by 37 CFR 41.20(b) in effect on March 18, 2013.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/952,764	11/19/2020	Gregory Porreca	MLB-002/06US 34468/40	7812

21710 7590 02/02/2022

BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

EXAMINER	
ART UNIT	PAPER NUMBER
MAIL DATE	DELIVERY MODE
02/02/2022	

Please find below and/or attached an Office communication concerning this application or proceeding.

PTOL-90A (Rev. 04/07)

UNITED STATES DEPARTMENT OF COMMERCE
U.S. Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450

APPLICATION NO./ CONTROL NO.	FILING DATE	FIRST NAMED INVENTOR/ PATENT IN REEXAMINATION	ATTORNEY DOCKET NO.
16/952,764	11/19/2020	Porreca et al.	MLB-002/06US 3446

	EXAMINER	
BROWN RUDNICK LLP **ONE FINANCIAL CENTER** **BOSTON, MA 02111**	STEVEN C POHNERT PhD	
	ART UNIT	**PAPER**
	1634	20220128a

DATE MAILED: _____

Please find below and/or attached an Office communication concerning this application or proceeding.

Commissioner for Patents

This is to note the the examiners answer should indicate on the front page the Examiner's Answer is in response to the appeal brief filed 10/8/2021 and 20/20/2021.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to STEVEN C POHNERT PhD whose telephone number is (571)272-3803. The examiner can normally be reached on Monday- Friday about 6:00 AM-5:30 PM, every second Friday off.
Examiner interviews are available via telephone, in-person, and video conferencing using a USPTO supplied web-based collaboration tool. To schedule an interview, applicant is encouraged to use the USPTO Automated Interview Request (AIR) at http://www.uspto.gov/interviewpractice.
If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Dave Nguyen can be reached on 571-272-0731. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.
Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see https://ppair-my.uspto.gov/pair/PrivatePair. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USAA

/Steven Pohnert/
Primary Examiner, Art Unit 1634

PTO-90C (Rev.04-03)

Electronic Acknowledgement Receipt	
EFS ID:	45368098
Application Number:	16952764
International Application Number:	
Confirmation Number:	7812
Title of Invention:	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS
First Named Inventor/Applicant Name:	Gregory Porreca
Customer Number:	21710
Filer:	Thomas C. Meyers/Sandra Richards
Filer Authorized By:	Thomas C. Meyers
Attorney Docket Number:	MLB-002/06US 34468/40
Receipt Date:	31-MAR-2022
Filing Date:	19-NOV-2020
Time Stamp:	17:27:33
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	DA
Payment was successfully received in RAM	$1180
RAM confirmation Number	E20223UH27472110
Deposit Account	
Authorized User	
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Reply Brief Filed	MLB-002-06ReplyBriefFinal.pdf	248455 ee104e3ae7a3dd302ca11191f65b284accc895d4	no	19
Warnings:					
Information:					
2	Fee Worksheet (SB06)	fee-info.pdf	38431 da3a79cb07011891d53421dfa802059215828579	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	286886		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.
National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.
New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Attorney Docket No.: MLB-002/06US 34468/40

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPELLANT:	Molecular Loop Biosciences, Inc.	ART UNIT:	1634
SERIAL NUMBER:	16/952,764	CONF. NO.:	7812
FILING DATE:	November 19, 2020	EXAMINER:	Steven C. Pohnert
TITLE:	**METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS**		

FILED ELECTRONICALLY

Commissioner for Patents
P. O. Box 1450
Alexandria, VA 22313-1450

REPLY BRIEF

This paper is in reply to the Examiner's Answer to Appeal Brief, mailed February 1, 2022 ("Examiner's Answer"), in response to Appellant's Appeal Brief filed October 8, 2021, and Supplemental Appeal Brief filed October 20, 2021, appealing a final rejection of claims 68-69, 71, 73-74, and 76 in the Final Office Action mailed on April 9, 2021, from the United States Patent and Trademark Office.

On July 16, 2021, Appellant filed a Pre-Appeal Brief Request for Review. A Notice of Panel Decision from Pre-Appeal Brief Review was mailed on September 10, 2021, from the Office, indicating that the application remains under appeal because there is at least one actual issue for appeal.

Appellant herewith submits the fees for small entity, pursuant to 37 CFR 41.20(b)(4), for forwarding an appeal to the Board.

I. Real Party in Interest

The real party in interest is Molecular Loop Biosciences, Inc., formerly Molecular Loop Biosolutions, LLC, having a place of business at 57 Reservoir Street #1, Cambridge, Massachusetts 02138.

Molecular Loop Biosciences, Inc. is the owner of the subject application by virtue of an assignment from inventors Gregory Porreca and Uri Laserson to Good Start Genetics, Inc., recorded on December 7, 2020 with the U.S. Patent and Trademark Office at Reel/Frame 054561/0315, and an assignment from Good Start Genetics, Inc. to Molecular Loop Biosolutions, LLC, recorded on December 7, 2020 with the U.S. Patent and Trademark Office at Reel/Frame 054561/0341 and a corporate conversion of Molecular Loop Biosolutions LLC to Molecular Loop Biosciences, Inc., recorded on September 28, 2021 at the United States Patent and Trademark Office at Reel/Frame 057634/0714.

Petitioner Illumina, Inc., EX1004, p. 1133

II. Summary of Claimed Subject Matter

The subject matter common to all of the pending claims is a method for correcting errors or bias introduced during nucleic acid sequencing workflows. The claimed methods comprise introducing a set of differentiator tags to nucleic acid molecules from multiple loci, such that nucleic acid molecules from one or more loci of origin are associated with a different tag. After amplifying and sequencing the differentiator tagged nucleic acid molecules, sequencing reads from copies of a nucleic acid molecule with the same differentiator tag are collapsed into a single count.

All examined dependent claims recite the method for correcting errors or bias introduced during sequencing.

Support for independent claim 68 is found in the as-filed specification, *inter alia*, at page 2, lines 18-28, page 3, lines 7-20; page 6, line 21 to page 7, line 19; page 27, line 27 to page 28, line 2; page 58, lines 3-9, and page 79, lines 8-15. Support is also be found in the Drawings, for example in FIG. 5 and FIG. 6.

Support for independent claim 76 is found in the as-filed specification, *inter alia*, at page 2, lines 18-28, page 3, lines 7-20; page 6, line 21 to page 7, line 19; page 27, line 27 to page 28, line 2; page 58, lines 3-9, and page 79, lines 8-15. Support is also be found in the Drawings, for example in FIG. 5 and FIG. 6.

III. Argument

Rejections under 35 U.S.C. §112, first paragraph

Claims 68-69, 71, 73-74, and 76 had been rejected under 35 U.S.C. §112, first paragraph, as lacking written description. The Examiner's Answer, at page 15, now acknowledges that "the specification has antecedent basis for the limitations of the claim" and that the specification includes an example (FIG. 5) showing how to collapse all sequences with the same differentiator tags that have the same sequence. However, the Examiner's Answer, at page 9, now, for the first time, states that the specification does "not adequately describe how to deal with more complex sequences with more variations due to starting material, capture, infidelity of the polymerase or tag attachment etc. Further it does not provide any detail on how this corrects errors or even which type of error would be corrected."

Applicant submits that, while the Examiner concedes that the specification contains a written description of the invention, the Examiner's Answer now asserts a legal standard for written description that is unsupported by case law and has only first appeared in Examiner's Answer.

The law requires that a patent specification must describe the claimed invention in sufficient detail that one skilled in the art can reasonably conclude that the inventor had possession of the claimed invention. *Moba, B.V. v. Diamond Automation, Inc.*, 325 F.3d 1306, 1319, 66 USPQ2d 1429, 1438 (Fed. Cir. 2003).

As conceded in the Examiner's Answer, the specification includes the elements of the claims and shows how the claimed method may be performed. Because performance of the claimed method is shown in the Examples, and because the specification recites and addresses the elements of the claims, one skilled in the art would reasonably conclude that the inventor had possession of the claimed invention. While the Examiner appears to agree with that, he has now generated his own standard against which the claims should be examined. That new standard is contrary to the language of 35 U.S.C. §112, first paragraph, and the relevant caselaw.

The examiner ignores the legal requirement that written description be evaluated from the perspective of what "one skilled in the art" would reasonably conclude. A skilled artisan would understand that the specification gives an experimental example showing the use of differentiator tag sequences to detect and correct bias in a MIP-capture reaction. In the example, 1000 genomic targets corresponding to 1000 exons were captured and tagged with differentiator tag sequences. *As-filed application at page 74.* Following amplification, sequencing is performed according to manufacturer's directions, reading both the target sequence and the differentiator tag sequence. *Id. at page 79.* The data are then used to correct any biases detected by quantification of differentiator tag sequence abundance. For example, the specification at page 79 describes the construction of a table for each target and differentiator tag ID, as follows:

Target ID	Differentiator Tag sequence ID	Count
1	3547	1
2	4762	1
1	9637	1
1	1078	5
3	4762	1
1	2984	1

The specification states that "all count entries should be '1', since any particular target:differentiator tag mapping will not occur more than once by chance, and therefore will only be observed if bias were present in the sample preparation process. For any target:differentiator tag combination observed more than once, all such reads are 'collapsed' into a single read before consensus basecalls are determined." *Id. at page 80.* That is to say, in the table above, the Count of sequence reads for Target ID 1 paired with Differential Tag sequence ID "1078" (row 4), would be "collapsed" from 5 sequence reads to 1 sequence read.

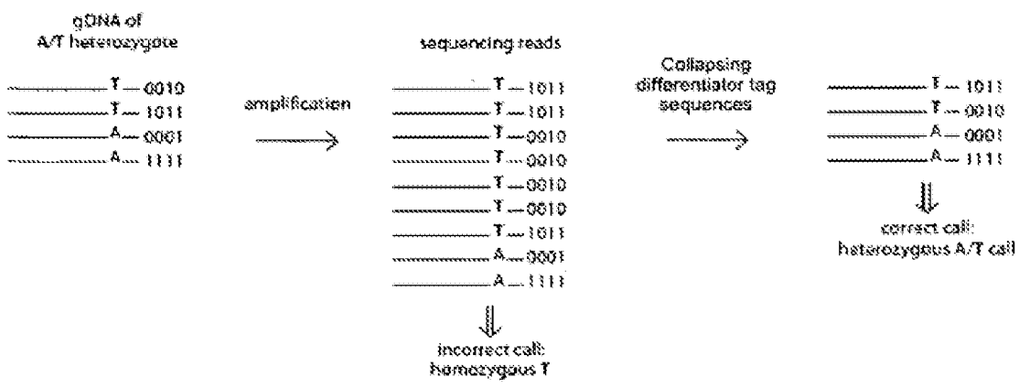
The specification gives no limitation on the number of rows that can be analyzed and count number that can be collapsed. The skilled artisan would conclude that this example may be

used for any number of genomic loci and can be used to analyze and collapse any number of reads. Accordingly, the specification very clearly describes how collapsing tags is actually done in a fashion consistent with complex sequencing data as understood by an actual person having ordinary skill in the art.

The specification also describes, in detail, the errors introduced during sample preparation and the manner in which the errors are corrected by the method of the claimed invention. For example, in the table above from Example 3, without first collapsing reads, the sequence read count for target 1, would be 8 (1 read with tag 3547, 1 with tag 9637, 5 with tag 1078, and 1 with 2984). However, the specification recognizes that count entries above 1 can only be observed if bias was present somewhere in the sample preparation process. Therefore, read counts above 1 are collapsed down to one, and the total count for target 1 is then calculated at 4 – the accurate number of copies of target 1 in the sequencing data prior to sample-introduced errors.

Error correction is further described in FIG. 5, reproduced below. with regard to consensus genotyping.

FIG. 5



As shown, the original sample at a target location is a heterozygote A/T, with two copies of each allele. However, as a result of bias introduced during amplification, the T allele is represented in the sequence reads 7 times and the A allele is represented only twice. The

consensus base call would erroneously identify the target as homozygous T. By collapsing reads with the same tag, the correct number of copies of the A alleles and T alleles are identified, and the correct call of heterozygous A/T can be made by a skilled artisan.

As shown above, the specification describes in specific detail potential errors introduced during sample preparation, how reads are collapsed following sequencing, and the manner in which errors are corrected by collapsing reads. A skilled artisan would understand the potential errors that exist during sample preparation for sequencing and that the inventors were in possession of a novel method for correcting these errors.

Clearly, Appellant's specification shows how to correct errors for any sequence or groups of sequences. There is no additional experimentation left to be done. Moreover, the Examiner's inappropriate and unsubstantiated standard that imposes a requirement that an applicant ".....adequately describe how to deal with more complex sequences with more variations due to starting material, capture, infidelity of the polymerase or tag attachment etc." and to ".... provide (any) detail on how this corrects errors or even which type of error would be corrected", while based on an inaccurate reading of the Specification, is also NOT the legal standard. Nonetheless, if the Examiner's standard were the appropriate legal standard, it would clearly be met by the instant application as detailed above.

In view of the above remarks, the Examiner's use of an inappropriate legal standard, and the Examiner's lack of anything other than speculation to support his position, Appellant requests that the Patent Trial and Appeal Board (PTAB) reverse the rejection of claims 68-69, 71, 73-74, and 76 under 35 U.S.C. §112, first paragraph.

Rejections under 35 U.S.C. §112, second paragraph

Pages 5-6 of the Examiner's Answer state that claims 68-69, 71, 73-74, and 76 remain rejected under 35 U.S.C. §112, second paragraph, as being indefinite.

In response to Appellant's brief, the Examiner's Answer, at page 25, appears to argue that the example provided in the instant Specification is insufficient as written description

support for "the analysis of thousands or more loci", Examiner's Answer at P. 25. The Examiner then states, at P. 25, that "(T)he specification in figure 5 demonstrates 4 bases with one base difference in sequences. It demonstrates in a very simple example that sequences can be preferentially amplified. The specification does not (sic) any guidance on what is required of collapsed (sic), how collapsing is done or how collapsing corrects errors."

The Examiner's Answer, at pages 24-31, is largely a verbatim reiteration of the comments made by the Examiner in the rejection under 35 U.S.C. §112, first paragraph – arguing again that the invention is not "sufficiently robust" to handle the "complexity of data." The Examiner's comments again conclude, at page 28, that Appellant's Brief "is essentially arguing we have explained what we want the last step of the claim to do, but has not adequately described how the limitation is actually done in a fashion consistent with the complexity of data the specification demonstrates in common to the art."

The law requires that "[a] decision on whether a claim is indefinite under 35 U.S.C. 112(b) or pre-AIA 35 U.S.C. 112, second paragraph requires a determination of whether those skilled in the art would understand what is claimed when the claim is read in light of the specification." *Power-One, Inc. v. Artesyn Techs., Inc.,* 599 F.3d 1343, 1350 (Fed. Cir. 2010); *Orthokinetics, Inc. v. Safety Travel Chairs, Inc.*, 806 F.2d 1565 (Fed. Cir. 1986)).

As conceded in the Examiner's Answer, the specification includes the elements of the claims and shows how the claimed method may be performed. Because performance of the claimed method is shown in the Examples, and the specification recites and addresses the elements of the claims, one skilled in the art would understand exactly what is claimed.

The Examiner has essentially imposed a new legal standard and has engaged in technically inaccurate speculation. Whether the specification would support using the invention on more complex samples or with additional loci is not the legal standard. There is no legal basis for the Examiner's position. Nevertheless, the Examiner's own comments indicate that a skilled artisan *would*, in fact, understand what is claimed.

For example, at page 24 of the Examiner's Answer, the Examiner states that "the invention encompasses a plurality of target molecules from more than one locus, such that one or more loci are associated with one for [sic] more differentiator tag [sic], which are then amplified and collapsed for error correction." The Examiner's Answer, at page 25, states that "the context of the invention as provided by the specification is not the analysis of 4 differentiator tags with sequences that vary by a single nucleotide, but thousands of loci which combine for millions of reads as evidenced by the inventors own work." The Examiner appears to have no problem understanding the scope of the invention and has not provided any technical arguments to indicate otherwise.

The Specification provides more than sufficient technical basis to support that those of skill in the art would understand what is claimed when the claim is read in light of the specification– even under the Examiner's novel and unsubstantiated legal standard. The Examiner has not provided a reasoned technical analysis or any expert evaluation that supports that the claims are indefinite under Section 112.

In view of the above remarks, Appellant requests that the Patent Trial and Appeal Board (PTAB) reverse the rejection of claims 68-69, 71, 73-74, and 76 under 35 U.S.C. §112, second paragraph

Rejections under 35 U.S.C. §101

Pages 6-9 of the Examiner's Answer state that claims 68-69, 71, 73-74, and 76 remain rejected under 35 U.S.C. §101 as being directed to a mental step.

The Supreme Court has held that everything under the sun made by a human is patentable under §101, subject to three narrow judicial exceptions: natural phenomena, laws of nature, and abstract ideas. *Diamond v. Chakrabarty*, 447 U.S. 303, 308-309 (1980). Even if the claim is directed to one of the judicial exceptions, then a second part of the analysis provides that the claim is still patent eligible if it contains "significantly more" than the judicial exception. *Mayo*

Collaborative Servs. v. Prometheus Labs., Inc., 556 U.S. 66 (2012); *Alice Corp. v. CLS Bank Int'l*, 573 U.S. 208 (2014).

The USPTO's 2019 Revised Patent Subject Matter Eligibility Guidance (the "*Revised Guidance*") explains that the judicial category of mental concepts seeks to exclude from patentability any process that would include an individual performing the claimed steps mentally or with a pencil and paper. *Revised Guidance on third page.* As a corollary, this judicial category does not include inventions that "could not, as a practical matter, be performed entirely in a human's mind." *CyberSource Corp. v. Retail Decisions, Inc.*, 654 F.3d 1366, 1375 (Fed. Cir. 2015)(emphasis added).

The claimed invention is directed to a laboratory workflow for sequencing nucleic acids. The claimed invention is not directed to a mental step.

For example, the claims recite tagging individual nucleic acid molecules with a differentiator tag prior to amplifying and sequencing steps. Each of the claims **requires multiple physical laboratory steps** to tag nucleic acid molecules with differentiator tags, amplify the tagged molecules, and to sequence the amplicons. It is impossible for those steps to be carried out as a mental process. Moreover, the steps of collapsing sequence reads and/or correcting for errors requires, first, the use of sequencing technology as claimed to generate sequence reads that include the differentiator tags, followed by the use of a computer to identify the tags. These steps could not possibly be performed mentally; but more to the point, the claims **REQUIRE PHYSICAL STEPS**.

Any unbiased reading of the claims makes clear the physical manipulation steps required therein. The claimed methods could not be carried out mentally or even by a computer without physical manipulation. The rejection is a gross distortion of the law around Section 101 and finds no basis in the law. In view of the above remarks, Appellant requests that the Patent Trial and Appeal Board (PTAB) reverse the rejection of claims 68-69, 71, 73-74, and 76 under 35 U.S.C. 101.

Rejections under 35 U.S.C. §103

Pages 9-12 of the Examiner's Answer state that claims 68-69 and 76 remain rejected under 35 U.S.C. §103 in view of Porreca 2007. Pages 13-14 of the Examiner's Answer state that claims 68-69 and 76 remain rejected under 35 U.S.C. §103 in view of Porreca 2007 in light of Wang. Pages 14-15 of the Examiner's Answer state that claims 71 and 73 remain rejected under 35 U.S.C. §103 in view of Porreca 2007 in light of Schiffman.

None of Porreca 2007, Wang, or Shiffman disclose the use of differentiator tags and therefore none of the cited references disclose collapsing target:differentiator tag combinations. The elements missing from the prior art preclude a rejection based thereon.

Appellant again notes that the initial rejection of claims 68-69 and 76 under §103 is based on a single reference that the Examiner's Reply indicates, at page 35, does not specifically teach "correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count." The subsequent rejection under §103 adds additional references that, like the primary reference, fail to disclose elements of the claims.

In order to establish a *prima facie* case of obviousness, the prior art must provide motivation to modify a reference or to combine multiple references in a manner that would lead one of ordinary skill in the art to an Appellant's claimed invention with a reasonable expectation of success. *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398 (2007).

Porreca 2007 does not, and cannot, provide any motivation that would lead one of ordinary skill in the art to modify its own teachings to lead to the claimed invention. Rather, Porreca 2007 itself recognized that PCR amplification steps were a source of nonuniformity error but **did not suggest any solution**, let alone the differentiator tags of the claimed invention. *Porreca 2007, pg. 4, Discussion.* A skilled artisan could not read Porreca 2007, which points out an existing issue causing errors in its methods, and come to the claimed invention, without the benefit of hindsight that the Examiner, who is not a skilled artisan, is relying on.

The Examiner's Reply, at page 35, states that differentiator tags are being broadly considered to encompass the targeting arms of a molecular inversion probe (MIP). The Examiner's Reply, at page 36, states that Porreca 2007, in genotyping homozygosity and heterozygosity, reports collapsing multiple MIPS which encompass the same sequence and same allele at a position into one read if an allele has greater than 0.8 specificity at a location, thereby identifying the location as homozygous. That statement reveals that in Porreca 2007 there are always multiple MIPS of the same sequence. The MIPS are not themselves unique molecules. This distinction reveals the fundamental misconception that underlies the Examiner's position.

As distinct from Porreca 2007, it is the differentiator tags of the present claims that are unique molecules (which makes them useful as "differentiator" tags) and not the MIPS. By collapsing MIPS that encompass the same sequence and same allele at a target position into one read, one can only obtain a maximum of 2 reads, one for each allele at a target position. In contrast, by tagging each molecule from a nucleic acid sample with one differentiator tag of the invention, each individual *copy* of the target in a sample may be identified as a different target:differentiator tag combination. This can be any number of different target:differentiator tag combinations depending on the number of copies of the nucleic acid in the original sample.

For example, page 79 of the specification, details the use of a table to identify and collapse reads. The table details that for a "Target ID 1", four separate sequence reads can be identified after collapsing target:differentiator tag combinations into a single read (each read with separate differentiator tags 3547, 96377, 1078, and 2984). Collapsing all of the reads of Target ID 1 into a single count, as the Examiner's Reply suggests Porreca 2007 teaches, would result in the incorrect identification of only a single copy of Target 1 being present in the original sample.

This difference between differentiator tags and MIPs is further highlighted in Fig. 5 of the as-filed specification. As shown in Fig. 5, the original sample comprises two copies of each allele, T and A. As a result of amplification bias, 7 sequence reads of the T allele are generated and 2 sequence reads of the A allele are generated. As a consequence, a skilled artisan would

understand that when calculating allele specificity as the Examiner's Reply suggests Porreca 2007 teaches, the T allele would have a 0.8 specificity, and the target location would be misidentified as homozygous T. In contrast, by using differentiator tags of the present invention, target:differentiator tag combinations are collapsed and not only is the correct genotype call identified, but the correct number of each copy of a starting allele in the sample is identified.

Accordingly, the MIPs of Porreca 2007, unlike the present invention, cannot correct for error or bias introduced during the workflow by being collapsed, but would rather introduce new errors.

Neither Wang nor Shiffman make any mention of differentiator tags and therefore cannot remedy the deficiencies of Porreca 2007.

The difference between the cited art and the claimed invention was clearly evidenced by the March 5, 2021, expert declaration of Dr. Gregory Porreca, which provides record evidence supporting the non-obviousness of the claims. Dr. Porreca is a skilled artisan and his expert declaration is the only evidence of the opinion of a skilled artisan that has been provided on the record. In his expert declaration, Dr. Porreca—inventor of the instant claims and first author of Porreca 2007—shows that none of the cited references, including his 2007 paper, disclose a method comprising tagging individual nucleic acid molecules with differentiator tags prior to amplification or sequencing steps. *Porreca dec., par. 6-12.*

In assessing the probative value of an expert opinion, the examiner must consider the nature of the matter sought to be established, the strength of any opposing evidence, the interest of the expert in the outcome of the case, and the presence or absence of factual support for the expert's opinion. *Ashland Oil, Inc. v. Delta Resins & Refractories, Inc.*, 776 F.2d 281, *cert. denied*, 475 U.S. 1017 (1986).

Dr. Porreca is the first author of Porreca 2007 and is a primary source on the nature of what is taught or suggested by that reference. His statements on what Porreca 2007 teach are first-hand facts about the scope of that reference. Dr. Porreca shows that the MIPs of Porreca

2007 could not correct for errors or bias introduced during the nucleic acid analysis workflow.

Specifically, Dr. Porreca states that at the time of his 2007 publication, he had not yet envisioned tagging nucleic acid molecules with differentiator tags. *Id., par. 9*. Moreover, Dr. Porreca reports that the unique targeting arms of the molecular inversion probes disclosed in Porreca 2007 are not differentiator probes at all. *Id. pars. 8, 11, 12.* As Dr. Porreca points out, this is because each molecular inversion probe is designed to uniquely capture an exon of interest with targeting arms that hybridize to sequences flanking the exon. *Id., para. 7.* The targeting arms are unique only in the sense that PCR primers are unique; unique designs each synthesized in the millions (this point is settled where Supplementary Figure 1 in Porreca 2007 states that the 480 probes were used at a "concentration of ~167 pM for each species"). If a probe is used at picomolar concentration, that means that a very large number of clonal copies of the molecule is present. Because the probes disclosed in Porreca 2007 are only "unique" among each other by virtue of being specific to their target exons while still each being used at ~167 pM concentration, the expert declaration concludes that their targeting arms cannot differentiate separate single molecules. *Id.*

Dr. Porreca further explained that Wang teaches only the use of molecular inversion probes for the improved measurement of allele copy number, that is, the number of different variants of an allele of a gene at a chromosomal location. *Id., pars 10-11.* However, because the unique targeting arms of the molecular inversion probes needed to capture different alleles of the gene are actually identical, the different copies of the allele must be identified and differentiated by their own sequence rather than by the sequence of the unique targeting arms. As a result, Wang suffers from the same handicap as the molecular inversion probes disclosed in Porreca 2007 when dealing with copy number polymorphisms. *Id.* That is, because the targeting arms of the molecular inversion probes will hybridize to the flanking sequences of any copies of the gene, the unique targeting arms cannot be used to differentiate between the identical copies of the gene as separate nucleic acid molecules. *Id.* Wang makes clear this limitation in stating that

copy number polymorphisms are likely attributable to identified false positives in their data that are actually true additional copies of the gene. *Id.* Wang specifically states that one of the shortcomings of their workflow is the presence of copy number polymorphisms (CNPs) in the autosomes. *Id.* Even with the solutions proposed for "ameliorating the effect of miscalling two adjacent markers in a CNP as a false positive", Wang states that doing so is only valid for the assessment of the technology performance "and cannot generate any meaningful biological findings." *Id.*

Dr. Porreca also explained that Schiffman does not disclose differentiator tags. *Id., para. 12.* Rather, Schiffman discloses only the use of molecular inversion probes to analyze loss of heterozygosity events, which are the same allele quantification or allele copy number events as disclosed in Wang. *Id.* As a result, Schiffman does not disclose the use of differentiator probes, and therefore cannot remedy the deficiencies of Porreca 2007 or Wang. *Id.*

The Examiner's Answer, at page 37, states that the claims do not require a different tag on every single molecule and therefore the Appellant's brief and declaration of Dr. Porreca are not arguing limitations of the claim. The Examiner is simply wrong and is not reading either the claim or the Porreca declaration correctly. As discussed above and in Dr. Porreca's declaration, the MIPs of Porreca 2007 **cannot** correct for errors or bias introduced during the nucleic acid analysis workflow. This is a clearly recited element of the claims that is not taught in Porreca 2007. The Examiner has failed to properly understand the Porreca declaration and has substituted his own technical opinion for that of an expert. Even without the declaration, the Examiner has failed to provide a proper basis for rejection because **he has failed to show the presence of claim elements in the cited art**.

In view of the above remarks, Appellant requests that the Patent Trial and Appeal Board (PTAB) reverse the rejection of claims 68-69, 71, 73-74, and 76 under 35 U.S.C. §103.

Summary

Appellant has met the legal standards for patentability and the Examiner has failed to

meet his burden of putting forth rejections consistent with the legal standards. In view of the numerous errors, speculation, lack of understanding of the technical aspects of the application and failure to adhere to legal standards, Appellant requests that the Patent Trial and Appeal Board (PTAB) reverse the rejections of claims 68-69, 71, 73-74, and 76.

Dated: March 31, 2022

Respectfully submitted,

BROWN RUDNICK LLP

BROWN RUDNICK LLP
One Financial Center
Boston, MA 02111
Tel: (617) 856-8483
Fax: (617) 856-8201

/Thomas C. Meyers/
Thomas C. Meyers, Reg. No. 36,989
Attorney for Appellant
Email: tmeyers@brownrudnick.com

I. Appendix A: Claims Appendix

1-67. (Canceled)

68. (Previously presented) A method for correcting for errors or bias introduced during nucleic acid analysis workflow, the method comprising the steps of:

obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin;

introducing a set of differentiator tags, wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag;

amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons;

sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and

correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count.

69. (Previously presented) The method of claim 68, wherein said correcting step further comprises collapsing all sequence reads comprising a same differentiator tag into a single read before determining a consensus base call for the target nucleic acid molecule sequences.

70. (Previously presented) The method of claim 69, further comprising the step of determining the presence of a sequencing or amplification error based upon a number of said differentiator tag sequences.

71. (Previously presented) The method of claim 68, wherein said biological sample is selected from the group consisting of blood and biopsy tissue.

72. (Previously presented) The method of claim 68, wherein said plurality of target nucleic acid molecules comprises bacterial or viral nucleic acids isolated from said biological sample.

73. (Previously presented) The method of claim 68, wherein said plurality of target nucleic acid molecules are circulating tumor nucleic acid molecules.

74. (Previously presented) The method of claim 68, further comprising the step of determining the copy number of a genomic region or transcript in a patient from whom the sample is obtained.

75. (Previously presented) The method of claim 68, wherein said plurality of target nucleic acid molecules are maternally-circulating fetal or placental nucleic acid molecules.

76. (Previously presented) A method for correcting for errors or bias introduced during nucleic acid analysis workflow, the method comprising the steps of:

obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin;

introducing a set of unique differentiator tags, wherein members of said set of unique differentiator tags are randomly associated with members of said plurality of target nucleic acid molecules, such that any given tag is uniquely associated with said locus of origin;

amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons;

sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and

correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule.

Petitioner Illumina, Inc., EX1004, p. 1150

Electronic Patent Application Fee Transmittal

Application Number:	16952764
Filing Date:	19-Nov-2020
Title of Invention:	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS
First Named Inventor/Applicant Name:	Gregory Porreca
Filer:	Thomas C. Meyers/Matthew York
Attorney Docket Number:	MLB-002/06US 34468/40

Filed as Small Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
APPEAL FORWARDING FEE	2413	1	1180	1180
Post-Allowance-and-Post-Issuance:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Extension-of-Time:				
Miscellaneous:				
		Total in USD ($)		**1180**

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/952,764	11/19/2020	Gregory Porreca	MLB-002/06US 34468/40	7812

21710 7590 04/06/2022

BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

EXAMINER	
POHNERT, STEVEN C	

ART UNIT	PAPER NUMBER
1634	

NOTIFICATION DATE	DELIVERY MODE
04/06/2022	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

ip@brownrudnick.com
usactions@brownrudnick.com

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

Ex parte GREGORY PORRECA et al.

Appeal 2022-002597
Application 16/952,764
Technology Center 1600

APPEAL DOCKETING NOTICE

The Patent Trial and Appeal Board received the appeal in the above-identified application from the Technology Center on April 04, 2022, and has assigned it the appeal number indicated above.

The recipient of this notice is reminded of its ongoing duty, within 20 days of any change during the proceeding, to update its mandatory notices to identify the real party-in-interest and each judicial or administrative proceeding that could affect, or be affected by, the Board proceeding. 37 C.F.R. § 41.8. For example, another docketed appeal in a related application may have the potential to affect, or be affected by, the Board proceeding.

In all future communications regarding this appeal, please include both the application number and the appeal number. Telephone inquiries can be made by calling 571-272-9797 and referencing the appeal number listed above. The mailing address for the Board is:

PATENT TRIAL and APPEAL BOARD
UNITED STATES PATENT AND TRADEMARK OFFICE
P.O. BOX 1450
ALEXANDRIA, VIRGINIA 22313-1450

If you are interested in seeking expedited treatment for this appeal via the Board's Fast-Track Appeals Pilot Program, please consult our website at https://www.uspto.gov/patents/ptab/fast-track-appeals-pilot-program or the Federal Register Notice at https://www.federalregister.gov/documents/2020/07/02/2020-14244/fast-track-appeals-pilot-program for more information on how to participate and which appeals qualify.

Electronic Acknowledgement Receipt

EFS ID:	47737088
Application Number:	16952764
International Application Number:	
Confirmation Number:	7812
Title of Invention:	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS
First Named Inventor/Applicant Name:	Gregory Porreca
Customer Number:	21710
Filer:	Thomas Meyers/Sandra Richards
Filer Authorized By:	Thomas Meyers
Attorney Docket Number:	MLB-002/06US 34468/40
Receipt Date:	24-MAR-2023
Filing Date:	19-NOV-2020
Time Stamp:	09:19:15
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Assignee showing of ownership per 37 CFR 3.73	MLB-002-06US-373c-statement.pdf	193148 0fa52a6a72cbd7ea483cafaa3ca9d8ef1e720c1c	no	3

Warnings:

Information:					
2	Power of Attorney	Molecular-Loop-POA-010923.pdf	168654 5ec7fee26f3770694f6de78b08c4e8ff766f2139	no	2
Warnings:					
Information:					
			Total Files Size (in bytes):	361802	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

PTO/AIA/96 (08-12)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

STATEMENT UNDER 37 CFR 3.73(c)

Applicant/Patent Owner: Molecular Loop Biosciences, Inc.

Application No./Patent No.: 16/952,764 Filed/Issue Date: November 19, 2020

Titled: METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS

_____, a _____

(Name of Assignee) (Type of Assignee, e.g., corporation, partnership, university, government agency, etc.)

states that, for the patent application/patent identified above, it is (choose **one** of options 1, 2, 3 or 4 below):

1. ☑ The assignee of the entire right, title, and interest.

2. ☐ An assignee of less than the entire right, title, and interest (check applicable box):

   ☐ The extent (by percentage) of its ownership interest is _____%. Additional Statement(s) by the owners holding the balance of the interest must be submitted to account for 100% of the ownership interest.

   ☐ There are unspecified percentages of ownership. The other parties, including inventors, who together own the entire right, title and interest are:

   Additional Statement(s) by the owner(s) holding the balance of the interest must be submitted to account for the entire right, title, and interest.

3. ☐ The assignee of an undivided interest in the entirety (a complete assignment from one of the joint inventors was made). The other parties, including inventors, who together own the entire right, title, and interest are:

   Additional Statement(s) by the owner(s) holding the balance of the interest must be submitted to account for the entire right, title, and interest.

4. ☐ The recipient, via a court proceeding or the like (*e.g.*, bankruptcy, probate), of an undivided interest in the entirety (a complete transfer of ownership interest was made). The certified document(s) showing the transfer is attached.

The interest identified in option 1, 2 or 3 above (not option 4) is evidenced by either (choose **one** of options A or B below):

A. ☐ An assignment from the inventor(s) of the patent application/patent identified above. The assignment was recorded in the United States Patent and Trademark Office at Reel _____, Frame _____, or for which a copy thereof is attached.

B. ☑ A chain of title from the inventor(s), of the patent application/patent identified above, to the current assignee as follows:

1. From: Inventors To: Good Start Genetics, Inc.

   The document was recorded in the United States Patent and Trademark Office at

   Reel 054561, Frame 0315, or for which a copy thereof is attached.

2. From: Good Start Genetics, Inc. To: Molecular Loop Biosolutions, LLC

   The document was recorded in the United States Patent and Trademark Office at

   Reel 054561, Frame 0341, or for which a copy thereof is attached.

This collection of information is required by 37 CFR 3.73(b). The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450**.

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

PTO/AIA/96 (08-12)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

STATEMENT UNDER 37 CFR 3.73(c)

3. From: Molecular Loop Biosolutions, LLC To: Molecular Loop Biosciences, Inc.

The document was recorded in the United States Patent and Trademark Office at
Reel 057634, Frame 0714, or for which a copy thereof is attached.

4. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at
Reel _____, Frame _____, or for which a copy thereof is attached.

5. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at
Reel _____, Frame _____, or for which a copy thereof is attached.

6. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at
Reel _____, Frame _____, or for which a copy thereof is attached.

☐ Additional documents in the chain of title are listed on a supplemental sheet(s).

☑ As required by 37 CFR 3.73(c)(1)(i), the documentary evidence of the chain of title from the original owner to the assignee was, or concurrently is being, submitted for recordation pursuant to 37 CFR 3.11.

[NOTE: A separate copy (i.e., a true copy of the original assignment document(s)) must be submitted to Assignment Division in accordance with 37 CFR Part 3, to record the assignment in the records of the USPTO. See MPEP 302.08]

The undersigned (whose title is supplied below) is authorized to act on behalf of the assignee.

/Thomas C. Meyers/	March 24, 2023
Signature	Date
Thomas C. Meyers	36,989
Printed or Typed Name	Title or Registration Number

Privacy Act Statement

The **Privacy Act of 1974 (P.L. 93-579)** requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether disclosure of these records is required by the Freedom of Information Act.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (*i.e.*, GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspection or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

PTO/AIA/80 (07-17)
Approved for use through 03/31/2021. OMB 0651-0035
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no person is required to respond to a collection of information unless it displays a valid OMB control number

POWER OF ATTORNEY TO PROSECUTE APPLICATIONS BEFORE THE USPTO

I hereby revoke all previous powers of attorney given in the application identified in the attached statement under 37 CFR 3.73(c).

I hereby appoint:

[X] Practitioners associated with Customer Number: 179511

OR

[ ] Practitioner(s) named below (if more than ten patent practitioners are to be named, then a customer number must be used):

Name	Registration Number	Name	Registration Number

As attorney(s) or agent(s) to represent the undersigned before the United States Patent and Trademark Office (USPTO) in connection with any and all patent applications assigned only to the undersigned according to the USPTO assignment records or assignment documents attached to this form in accordance with 37 CFR 3.73(c).

Please change the correspondence address for the application identified in the attached statement under 37 CFR 3.73(c) to:

[X] The address associated with Customer Number: 179511

OR

[ ]

Firm or individual name		
Address		
City	State	Zip
Country		
Telephone	Email	

Assignee name and address: Molecular Loop Biosciences, Inc.
300 Tradecenter, Suite 5400, Woburn, MA 01801

A copy of this form, together with a statement under 37 CFR 3.73(c) (Form PTO/AIA/96 or equivalent) is required to be filed in each application in which this form is used. The statement under 37 CFR 3.73(c) may be completed by one of the practitioners appointed in this form, and must identify the application in which this Power of Attorney is to be filed.

SIGNATURE of Assignee of Record

The individual whose signature and title is supplied below is authorized to act on behalf of the assignee.

Signature	Date09-Jan-2023
NameGregory Porreca	Telephone978-254-7799
TitlePresident and CEO	

This collection of information is required by 37 CFR 1.31, 1.32, and 1.33. The information is required to obtain or retain a benefit by the public, which is to update (and by the USPTO to process) the file of a patent or reexamination proceeding. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 18 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

Privacy Act Statement

The **Privacy Act of 1974 (P.L. 93-579)** requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether disclosure of these records is required by the Freedom of Information Act.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (*i.e.*, GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspection or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

Electronic Acknowledgement Receipt

EFS ID:	47737296
Application Number:	16952764
International Application Number:	
Confirmation Number:	7812
Title of Invention:	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS
First Named Inventor/Applicant Name:	Gregory Porreca
Customer Number:	21710
Filer:	Thomas Meyers/Sandra Richards
Filer Authorized By:	Thomas Meyers
Attorney Docket Number:	MLB-002/06US 34468/40
Receipt Date:	24-MAR-2023
Filing Date:	19-NOV-2020
Time Stamp:	10:00:25
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Assignee showing of ownership per 37 CFR 3.73	MLB-002-06US-373c-statement.pdf	198358 143b8fb6867a2e0ef599894629b820e5c2d3b7f4	no	3

Warnings:

Information:					
2	Power of Attorney	Molecular-Loop-POA-010923.pdf	168654 5ec7fee26f3770694f6de78b08c4e8ff766f2139	no	2
Warnings:					
Information:					
			Total Files Size (in bytes):	367012	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

PTO/AIA/96 (08-12)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

STATEMENT UNDER 37 CFR 3.73(c)

Applicant/Patent Owner: Molecular Loop Biosciences, Inc.

Application No./Patent No.: 16/952,764 Filed/Issue Date: November 19, 2020

Titled: METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS

Molecular Loop Biosciences, Inc., a corporation

(Name of Assignee) (Type of Assignee, e.g., corporation, partnership, university, government agency, etc.)

states that, for the patent application/patent identified above, it is (choose **one** of options 1, 2, 3 or 4 below):

1. ☑ The assignee of the entire right, title, and interest.

2. ☐ An assignee of less than the entire right, title, and interest (check applicable box):

   ☐ The extent (by percentage) of its ownership interest is _____%. Additional Statement(s) by the owners holding the balance of the interest must be submitted to account for 100% of the ownership interest.

   ☐ There are unspecified percentages of ownership. The other parties, including inventors, who together own the entire right, title and interest are:

   Additional Statement(s) by the owner(s) holding the balance of the interest must be submitted to account for the entire right, title, and interest.

3. ☐ The assignee of an undivided interest in the entirety (a complete assignment from one of the joint inventors was made). The other parties, including inventors, who together own the entire right, title, and interest are:

   Additional Statement(s) by the owner(s) holding the balance of the interest must be submitted to account for the entire right, title, and interest.

4. ☐ The recipient, via a court proceeding or the like (*e.g.*, bankruptcy, probate), of an undivided interest in the entirety (a complete transfer of ownership interest was made). The certified document(s) showing the transfer is attached.

The interest identified in option 1, 2 or 3 above (not option 4) is evidenced by either (choose **one** of options A or B below):

A. ☐ An assignment from the inventor(s) of the patent application/patent identified above. The assignment was recorded in the United States Patent and Trademark Office at Reel _____, Frame _____, or for which a copy thereof is attached.

B. ☑ A chain of title from the inventor(s), of the patent application/patent identified above, to the current assignee as follows:

1. From: Inventors To: Good Start Genetics, Inc.

   The document was recorded in the United States Patent and Trademark Office at

   Reel 054561, Frame 0315, or for which a copy thereof is attached.

2. From: Good Start Genetics, Inc. To: Molecular Loop Biosolutions, LLC

   The document was recorded in the United States Patent and Trademark Office at

   Reel 054561, Frame 0341, or for which a copy thereof is attached.

This collection of information is required by 37 CFR 3.73(b). The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450**.

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

PTO/AIA/96 (08-12)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

STATEMENT UNDER 37 CFR 3.73(c)

3. From: Molecular Loop Biosolutions, LLC To: Molecular Loop Biosciences, Inc.

The document was recorded in the United States Patent and Trademark Office at
Reel 057634, Frame 0714, or for which a copy thereof is attached.

4. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at
Reel _____, Frame _____, or for which a copy thereof is attached.

5. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at
Reel _____, Frame _____, or for which a copy thereof is attached.

6. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at
Reel _____, Frame _____, or for which a copy thereof is attached.

☐ Additional documents in the chain of title are listed on a supplemental sheet(s).

☑ As required by 37 CFR 3.73(c)(1)(i), the documentary evidence of the chain of title from the original owner to the assignee was, or concurrently is being, submitted for recordation pursuant to 37 CFR 3.11.

[NOTE: A separate copy (i.e., a true copy of the original assignment document(s)) must be submitted to Assignment Division in accordance with 37 CFR Part 3, to record the assignment in the records of the USPTO. See MPEP 302.08]

The undersigned (whose title is supplied below) is authorized to act on behalf of the assignee.

/Thomas C. Meyers/	March 24, 2023
Signature	Date
Thomas C. Meyers	36,989
Printed or Typed Name	Title or Registration Number

Privacy Act Statement

The **Privacy Act of 1974 (P.L. 93-579)** requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether disclosure of these records is required by the Freedom of Information Act.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (*i.e.*, GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspection or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

PTO/AIA/80 (07-17)
Approved for use through 03/31/2021. OMB 0651-0035
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no person is required to respond to a collection of information unless it displays a valid OMB control number

POWER OF ATTORNEY TO PROSECUTE APPLICATIONS BEFORE THE USPTO

I hereby revoke all previous powers of attorney given in the application identified in the attached statement under 37 CFR 3.73(c).

I hereby appoint:

[X] Practitioners associated with Customer Number: 179511

OR

[ ] Practitioner(s) named below (if more than ten patent practitioners are to be named, then a customer number must be used):

Name	Registration Number	Name	Registration Number

As attorney(s) or agent(s) to represent the undersigned before the United States Patent and Trademark Office (USPTO) in connection with any and all patent applications assigned only to the undersigned according to the USPTO assignment records or assignment documents attached to this form in accordance with 37 CFR 3.73(c).

Please change the correspondence address for the application identified in the attached statement under 37 CFR 3.73(c) to:

[X] The address associated with Customer Number: 179511

OR

[ ]

Firm or individual name		
Address		
City	State	Zip
Country		
Telephone	Email	

Assignee name and address: Molecular Loop Biosciences, Inc.
300 Tradecenter, Suite 5400, Woburn, MA 01801

A copy of this form, together with a statement under 37 CFR 3.73(c) (Form PTO/AIA/96 or equivalent) is required to be filed in each application in which this form is used. The statement under 37 CFR 3.73(c) may be completed by one of the practitioners appointed in this form, and must identify the application in which this Power of Attorney is to be filed.

SIGNATURE of Assignee of Record

The individual whose signature and title is supplied below is authorized to act on behalf of the assignee.

Signature	Date09-Jan-2023
NameGregory Porreca	Telephone978-254-7799
TitlePresident and CEO	

This collection of information is required by 37 CFR 1.31, 1.32, and 1.33. The information is required to obtain or retain a benefit by the public, which is to update (and by the USPTO to process) the file of a patent or reexamination proceeding. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 18 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

Privacy Act Statement

The **Privacy Act of 1974 (P.L. 93-579)** requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether disclosure of these records is required by the Freedom of Information Act.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (*i.e.*, GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspection or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
16/952,764	11/19/2020	Gregory Porreca	MLB-002/06US 34468/40

CONFIRMATION NO. 7812

179511
Withers Bergman LLP
101 Federal Street, Suite 1900
Boston, MA 02110

POA ACCEPTANCE LETTER

OC000000137694929

Date Mailed: 04/10/2023

NOTICE OF ACCEPTANCE OF POWER OF ATTORNEY

This is in response to the Power of Attorney filed 03/24/2023.

The Power of Attorney in this application is accepted. Correspondence in this application will be mailed to the above address as provided by 37 CFR 1.33.

Questions about the contents of this notice and the requirements it sets forth should be directed to the Office of Data Management, Application Assistance Unit, at **(571) 272-4000** or **(571) 272-4200** or **1-888-786-0101**.

/ggasgedom/

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
16/952,764	11/19/2020	Gregory Porreca	MLB-002/06US 34468/40

CONFIRMATION NO. 7812

21710
BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

POWER OF ATTORNEY NOTICE

OC000000137694919

Date Mailed: 04/10/2023

NOTICE REGARDING CHANGE OF POWER OF ATTORNEY

This is in response to the Power of Attorney filed 03/24/2023.

- The Power of Attorney to you in this application has been revoked by the applicant. Future correspondence will be mailed to the new address of record(37 CFR 1.33).

Questions about the contents of this notice and the requirements it sets forth should be directed to the Office of Data Management, Application Assistance Unit, at **(571) 272-4000** or **(571) 272-4200** or **1-888-786-0101**.

/ggasgedom/

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/952,764	11/19/2020	Gregory Porreca	MLB-002/06US 34468/40	7812

179511 7590 09/06/2023

Withers Bergman LLP
101 Federal Street, Suite 1900
Boston, MA 02110

EXAMINER	
POHNERT, STEVEN C	

ART UNIT	PAPER NUMBER
1634	

NOTIFICATION DATE	DELIVERY MODE
09/06/2023	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

Samantha.Metcalf@withersworldwide.com
US.Docketing@withersworldwide.com
kristina.scandore@withersworldwide.com

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

Ex parte GREGORY PORRECA and URI LASERSON

Appeal 2022-002597
Application 16/952,764
Technology Center 1600

Before DONALD E. ADAMS, ERIC B. GRIMES, and
JEFFREY N. FREDMAN, *Administrative Patent Judges.*

ADAMS, *Administrative Patent Judge.*

DECISION ON APPEAL

Pursuant to 35 U.S.C. § 134(a), Appellant[1] appeals from the Examiner's decision to reject claims 68, 69, 71, 73, 74, and 76 (*see* Appeal Br. 1). We have jurisdiction under 35 U.S.C. § 6(b).

We REVERSE.

---

[1] "Appellant" refers to "applicant" as defined in 37 C.F.R. § 1.42. Appellant identifies the real party in interest as "Molecular Loop Biosciences, Inc., formerly Molecular Loop Biosolutions, LLC" (Appellant's October 8, 2021, Appeal Brief ("Appeal Br.") 3).

STATEMENT OF THE CASE

Appellant's disclosure "relates to methods and compositions for determining genotypes in patient samples" (Spec.[2] 1). Claim 68, 71, 74, and 76 are reproduced below:

> 68. A method for correcting for errors or bias introduced during nucleic acid analysis workflow, the method comprising the steps of:
>
> obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin;
>
> introducing a set of differentiator tags, wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag;
>
> amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons;
>
> sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and
>
> correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count.

(Appeal Br. 23.)

> 71. The method of claim 68, wherein said biological sample is selected from the group consisting of blood and biopsy tissue.

(*Id.* at 24.)

> 74. The method of claim 68, further comprising the step of determining the copy number of a genomic region or transcript in a patient from whom the sample is obtained.

(*Id.*)

[2] Appellant's December 11, 2020, Specification.

Petitioner Illumina, Inc., EX1004, p. 1173

> 76. A method for correcting for errors or bias introduced during nucleic acid analysis workflow, the method comprising the steps of:
>
> obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin;
>
> introducing a set of unique differentiator tags, wherein members of said set of unique differentiator tags are randomly associated with members of said plurality of target nucleic acid molecules, such that any given tag is uniquely associated with said locus of origin;
>
> amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons;
>
> sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and
>
> correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule.

(*Id.* at 24–25.)

Grounds of rejection before this Panel for review:

I. Claims 68, 69, 71, 73, 74, and 76 stand rejected under 35 U.S.C. § 112, second paragraph.

II. Claims 68, 69, 71, 73, 74, and 76 stand rejected under the written description provision of 35 U.S.C. § 112, first paragraph.

III. Claims 68, 69, 71, 73, 74, and 76 stand rejected under 35 U.S.C. § 101.

Petitioner Illumina, Inc., EX1004, p. 1174

IV. Claims 68, 69, and 76 stand rejected under 35 U.S.C. § 103(a) as unpatentable over Porreca.[3]

V. Claim 74 stands rejected under 35 U.S.C. § 103(a) as unpatentable over the combination of Porreca and Wang.[4]

VI. Claims 71 and 73 stand rejected under 35 U.S.C. § 103(a) as unpatentable over the combination of Porreca and Schiffman.[5]

I–II

Definiteness and Written Description:

ISSUE

Rejection I: Does the preponderance of evidence support Examiner's conclusion that Appellant's claims are indefinite?

Rejection II: Does the preponderance of evidence on this record support Examiner's finding that Appellant's Specification fails to provide written descriptive support for the claimed invention?

ANALYSIS

We find substantial overlap between Examiner's indefiniteness and written description rejections and, therefore, consider the rejections together.

Appellant's claimed methods require, *inter alia*, the amplification of specific, i.e., target, nucleic acid molecules to generate amplicons.

---

[3] Porreca et al., *Multiplex amplification of large sets of human exons*, 4 Nature Methods 931–36 (2007).
[4] Wang et al., *Analysis of molecular inversion probe performance for allele copy number determination*, 8:R246 Genome Biology R246.1–R246.14 (2007).
[5] Schiffman et al., *Adapting molecular inversion probe (MIP) technology for allele quantification in childhood leukemia*, Journal of Clinical Oncology 25:18_suppl. Journal of Clinical Oncology 9530-953025 (2007).

Appellant's claimed methods amplify the target nucleic acid molecules using a nucleic acid that (a) binds the target, i.e., targeting arms,[6] and (b) contains a unique sequence, i.e., a differentiator tag sequence,[7] that is "useful to distinguish independent target molecules from each other thereby allowing an analysis to be based on a known number of individual target molecules" (Spec. 38; *see also id.* at 40 (Appellant discloses that "each nucleic acid molecule may be tagged with a unique differentiator tag sequence in a configuration that permits the differentiator tag sequence to be sequenced along with the target nucleic acid sequence of interest" and, thus, "[t]he incorporation of the nucleic acid comprising a differentiator tag sequence at a particular step allows the detection and correction of biases in subsequent steps of the protocol."); *id.* at 39 (Appellant discloses that the "combinations of target nucleic acid and differentiator tag sequence (target:differentiator

---

[6] Appellant discloses that "a target nucleic acid region that is being evaluated may be determined by the identity of the targeting arms of a probe[, i.e., a molecular inversion probe (MIP),] that is designed to capture the target region (or sequence thereof)" (Spec. 12; *see also id.* at 14). Appellant discloses that a MIP

> is a single linear strand of nucleic acid that comprises a first targeting arm at its 5' end and a second targeting arm at its 3' end, wherein the first targeting arm is capable of specifically hybridizing to a first region flanking one end of the sub-target nucleic acid, and wherein the second targeting arm is capable of specifically hybridizing to a second region flanking the other end of the sub target nucleic acid on the same strand of the target nucleic acid.

(*id.* at 14–15.)

[7] Appellant discloses MIP probes that contain a differentiator tag sequence (*see* Spec. 42).

Petitioner Illumina, Inc., EX1004, p. 1176

tag sequences) of an isolated nucleic acid of a preparative method provide a basis for identifying unique, independently isolated target nucleic acids.")).

> For example, if each of a plurality of target molecule sequences obtained in an assay is associated with a different differentiator tag, then the target sequences can be considered to be independent of each other and a genotype likelihood can be determined based on this information. In contrast, if each of the plurality of target molecule sequences obtained in the assay is associated with the same differentiator tag, then they probably all originated from the same target molecule due to over representation (e.g., due to biased amplification) of this target molecule in the assay.

(*Id.* at 38–39.) As Declarant explains:

> Each differentiator tag has its own unique sequence so that each individual target nucleic acid molecule is tagged, and therefore associated, with a differentiator tag. . . . Accordingly, each target:differentiator tag combination uniquely identifies a single nucleic acid molecule through subsequence amplification and sequencing steps. . . . Because each target:differentiator tag combination can be associated back to its original, individual nucleic molecule based on the unique combination of individual nucleic acid molecules and their respective differentiator tags, the invention corrects for bias in subsequent steps. . . . For example, the overrepresentation of sequence reads having a combination of nucleic acid molecule sequence and its differentiator tag sequence is indicative of bias in the protocol at one or more steps between the incorporation of the differentiator tag sequence and the sequencing of the combination. . . . Accordingly, the bias may be corrected to provide an accurate representation of the individual nucleic acid molecule in the original pool of nucleic acid molecules. . . . For example, it can be recognized that target:differentiator tag combinations having the same sequence are derived from the same original individual nucleic acid molecule and are not representative of two copies of the nucleic acid molecule in the original sample or genome. . . . Target:differentiator tag combinations observed more than once can then be *collapsed*

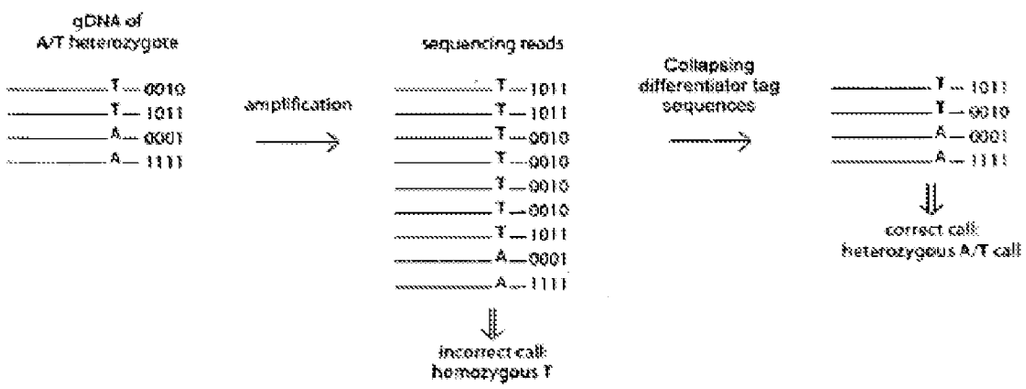
Petitioner Illumina, Inc., EX1004, p. 1177

> into a single read to cancel the effect of bias on consensus basecall accuracy.

(Porreca Decl.[8] ¶ 4 (emphasis added) (citing Spec. 7, 27, 28, 39, 41, 42, 79).)

Appellant's Figure 5, reproduced below, illustrates how target:differentiator tag combinations are collapsed to correct for errors or bias and, thus, provide a correct determination of a heterozygous sequence of genomic DNA (gDNA).

FIG. 5



Appellant's "FIG. 5 depicts a non-limiting method for genotyping based on target and differentiator tag sequences. Sequencing reads of target and differentiator tags sequences are collapsed to make diploid genotype calls" (Spec. 42). More specifically, Appellant's Figure 5 illustrates gDNA that is heterozygous for A and T, wherein the gDNA is amplified using the following four target:differentiator tag combinations, T:0010, T:1011, A:0001, and A:1111 (*see* Fig. 5). Sequencing reads after amplification result in 3 amplicons corresponding to T:1011, 4 amplicons corresponding to T:0010, 1 amplicon corresponding to A:0001, and 1 amplicon corresponding to A:1111 (*see id.*). According to Figure 5, if the analysis stopped at this

[8] Declaration of Gregory Porreca, Ph.D., signed March 4, 2021.

point with 7 amplicons corresponding to target T and 2 amplicons corresponding to target A, error and bias may be introduced wherein an inference is made that the two amplicons corresponding to A are erroneous and, therefore, lead to an *incorrect* conclusion that the gDNA was homozygous for T (*see* Reply Br.[9] 5 (citing Spec. 80) (Appellant explains: "The [S]pecification states that all count entries should be 1, since any particular target:differentiator tag mapping will not occur more than once by chance, and therefore will only be observed if bias were present in the sample preparation process" (internal quotations omitted)).

Appellant's claimed methodology requires the step of *recognizing* that target:differentiator tags having the same sequence are derived from the same input molecule. Stated differently, Appellant's claimed methodology requires that multiple amplicons having the same differentiator tag sequence are *collapsed* into a single count (i.e., *collapsing* the 3 amplicons corresponding to T:1011 into a single count and the 4 amplicons corresponding to T:0010 into a single count) (*see* Reply Br. 5 (citing Spec. 80) (Appellant explains that its Specification discloses that "[f]or any target:differentiator tag combination observed more than once, all such reads are collapsed into a single read before consensus basecalls are determined" (internal quotations omitted)). Applying Appellant's methodology to the sequencing reads illustrated in Figure 5 results in 2 counts for target T and 2 counts for target A (*e.g.*, 1 count for T:1011; 1 count for T:0010, 1 count for A:0001, and 1 count for A:1111) and, thereby, leads to the *correct* conclusion that the gDNA was heterozygous for A and T (*see, e.g.*, Porreca Decl. ¶ 6 (Declarant explains the claimed methodology "allows for a

[9] Appellant's March 31, 2022, Reply Brief.

Petitioner Illumina, Inc., EX1004, p. 1179

reduction in errors associated with the capture and amplification of nucleic acids by 'collapsing' duplicate sequences, thereby providing a significant improvement in the art."); *see generally* Spec. 77–80 (Exemplifying the "[u]se of differentiator tag sequences to detect and correct bias in a MIP-capture reaction of a set of exon targets."); Reply Br. 5).

Claims 68 and 76:

For the foregoing reasons, we are not persuaded by Examiner's finding that Appellant's "[S]pecification provides no guidance on how to perform collapsing [or recognizing], why collapsing [or recognizing] is done or how collapsing [or recognizing] corrects errors of bias" (Ans. 5–6; *see also id.* at 4 (Examiner finds that Appellant's "[S]pecification does not [provide] any guidance on what is required of collapsed, how collapsing is done or how collapsing corrects errors . . . or what is require[d] of a target sequence to identify it [is] to be collapsed.")).

As Appellant explains, its "[S]pecification gives no limitation on the number of . . . [target:differentiator tag combinations] that can be analyzed [or] . . . count number that can be collapsed" and, thus, a "skilled artisan would conclude that th[e] example [provided in Appellant's Specification] may be used for any number of genomic loci and can be used to analyze and collapse any number of [sequence] reads" (Reply Br. 5–6). We agree and, therefore, are not persuaded by Examiner's finding that Appellant's Figure 5 "demonstrates in a very simple example that sequences can be preferentially amplified" and does not provide guidance on "[h]ow to determine to collapse sequences of greater complexity" (Ans. 4–5).

For the foregoing reasons, we are not persuaded by Examiner's finding that the phrase "correcting for error or bias introduced during said

workflow by collapsing target:differentiator tag combinations observed more than once into a single count," as recited in Appellant's claim 68, is indefinite and lacks written descriptive support in Appellant's Specification (Ans. 4, 5). For the same reasons, we are not persuaded by Examiner's finding that the phrase "correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule," as recited in Appellant's claim 76, is indefinite and lacks written descriptive support in Appellant's Specification (Ans. 4, 6).

Claim 74:

Example 2 of Appellant's Specification exemplifies the detection of three targets (1, 2, and 3) (*see* Spec. 79; *see also* Reply Br. 5). Focusing on target 1, Appellant exemplifies the use of four different differentiator tags for target 1, specifically differentiator tag IDs 3547, 9637, 1078, and 2984 (*see* Spec. 79; *see also* Reply Br. 5). Appellant further exemplifies results, i.e., counts, for each of the foregoing differentiator tag IDs as follows: 1 count for 3547, 1 count for 9637, *5 counts* for 1078, and 1 count for 2984 (*see* Spec. 79; *see also* Reply Br. 5). As Appellant explains:

> [W]ithout first collapsing reads, the sequence read count for target 1, would be 8 (1 read with tag 3547, 1 with tag 9637, 5 with tag 1078, and 1 with 2984). However, the [S]pecification recognizes that count entries above 1 can only be observed if bias was present somewhere in the sample preparation process. Therefore, read counts above 1 are collapsed down to one, and the total count for target 1 is then calculated at 4–the accurate

Petitioner Illumina, Inc., EX1004, p. 1181

> *number of copies of target 1* in the sequencing data prior to sample-introduced errors.

(Reply Br. 6 (emphasis added); *see id.* (Appellant explains that "the [c]ount of sequence reads for [t]arget ID 1 paired with [differentiator] [t]ag sequence ID '1078' . . . would be 'collapsed' from 5 sequence reads[, i.e., counts,] to 1 sequence read," i.e., count. Thus, the copy number of a genomic region or transcript in a patient from whom the sample is obtained for target 1 is 4).)

For the foregoing reasons, we are not persuaded by Examiner's finding that Appellant's "specification and claims provide no guidance on how to determine a copy number of genomic region or transcript based on collapsing reads of target:differentiator tag combinations" (Ans. 4; *see also id.* at 5). For the same reasons, we are not persuaded by Examiner's finding that the phrase "further comprising the step of determining the copy number of a genomic region or transcript in a patient from whom the sample is obtained," as recited in Appellant's claim 74 is indefinite and lacks written descriptive support in Appellant's Specification (Ans. 4, 5).

CONCLUSION

Rejection I: The preponderance of evidence fails to support Examiner's conclusion that Appellant's claims are indefinite. The rejection of claims 68, 69, 71, 73, 74, and 76 under 35 U.S.C. § 112, second paragraph, is reversed.

Rejection II: The preponderance of evidence on this record fails to support Examiner's finding that Appellant's Specification fails to provide written descriptive support for the claimed invention. The rejection of claims 68, 69, 71, 73, 74, and 76 under the written description provision of 35 U.S.C. § 112, first paragraph, is reversed.

Petitioner Illumina, Inc., EX1004, p. 1182

III

Subject Matter Eligibility:

ISSUE

Does the preponderance of evidence of record support Examiner's finding that Appellant's claimed invention is directed to patent-ineligible subject matter?

PRINCIPLES OF LAW

A. Section 101

An invention is patent-eligible if it claims a "new and useful process, machine, manufacture, or composition of matter." 35 U.S.C. § 101. However, the U.S. Supreme Court has long interpreted 35 U.S.C. § 101 to include implicit exceptions: "Laws of nature, natural phenomena, and abstract ideas" are not patentable. *E.g.*, *Alice Corp. v. CLS Bank Int'l*, 573 U.S. 208, 216 (2014).

In determining whether a claim falls within an excluded category, we are guided by the Court's two-part framework, described in *Mayo* and *Alice*. *Alice*, 573 U.S. at 217–18 (citing *Mayo Collaborative Servs. v. Prometheus Labs., Inc.*, 566 U.S. 66, 75–77 (2012)). In accordance with that framework, we first determine what concept the claim is "directed to." *See Alice*, 573 U.S. at 219 ("On their face, the claims before us are drawn to the concept of intermediated settlement, *i.e.*, the use of a third party to mitigate settlement risk."); *see also Bilski v. Kappos*, 561 U.S. 593, 611 (2010) ("Claims 1 and 4 in petitioners' application explain the basic concept of hedging, or protecting against risk.").

Concepts determined to be abstract ideas, and thus patent ineligible, include certain methods of organizing human activity, such as fundamental

Petitioner Illumina, Inc., EX1004, p. 1183

economic practices (*Alice*, 573 U.S. at 219–20; *Bilski*, 561 U.S. at 611); mathematical formulas (*Parker v. Flook*, 437 U.S. 584, 594–95 (1978)); and mental processes (*Gottschalk v. Benson*, 409 U.S. 63, 67 (1972)). Concepts determined to be patent eligible include physical and chemical processes, such as "molding rubber products" (*Diamond v. Diehr*, 450 U.S. 175, 191 (1981)); "tanning, dyeing, making water-proof cloth, vulcanizing India rubber, smelting ores" (*id.* at 182 n.7 (quoting *Corning v. Burden*, 56 U.S. 252, 267–68 (1854))); and manufacturing flour (*Benson*, 409 U.S. at 69 (citing *Cochrane v. Deener*, 94 U.S. 780, 785 (1876))).

In *Diehr*, the claim at issue recited a mathematical formula, but the Court held that "a claim drawn to subject matter otherwise statutory does not become nonstatutory simply because it uses a mathematical formula." *Diehr*, 450 U.S. at 187; *see also id.* at 191 ("We view respondents' claims as nothing more than a process for molding rubber products and not as an attempt to patent a mathematical formula."). Having said that, the Court also indicated that a claim "seeking patent protection for that formula in the abstract . . . is not accorded the protection of our patent laws, and this principle cannot be circumvented by attempting to limit the use of the formula to a particular technological environment." *Id.* (citation omitted) (citing *Benson* and *Flook*); *see, e.g.*, *id.* at 187 ("It is now commonplace that an *application* of a law of nature or mathematical formula to a known structure or process may well be deserving of patent protection.").

If the claim is "directed to" an abstract idea, we turn to the second step of the *Alice* and *Mayo* framework, where "we must examine the elements of the claim to determine whether it contains an 'inventive concept' sufficient to 'transform' the claimed abstract idea into a patent-

Petitioner Illumina, Inc., EX1004, p. 1184

eligible application." *Alice*, 573 U.S. at 221 (internal quotation marks omitted). "A claim that recites an abstract idea must include 'additional features' to ensure 'that the [claim] is more than a drafting effort designed to monopolize the [abstract idea].'" *Id.* (alterations in original) (quoting *Mayo*, 566 U.S. at 77). "[M]erely requir[ing] generic computer implementation[] fail[s] to transform that abstract idea into a patent-eligible invention." *Id.*

B. USPTO Section 101 Guidance

In January 2019, the U.S. Patent and Trademark Office (USPTO) published revised guidance on the application of § 101. 2019 Revised Patent Subject Matter Eligibility Guidance, 84 Fed. Reg. 50 (Jan. 7, 2019) ("Guidance").[10] "All USPTO personnel are, as a matter of internal agency management, expected to follow the guidance." *Id.* at 51; *see also* October 2019 Update at 1.

Under the Guidance and the October 2019 Update, we first look to whether the claim recites:

> (1) any judicial exceptions, including certain groupings of abstract ideas (i.e., mathematical concepts, certain methods of organizing human activity such as a fundamental economic practice, or mental processes) ("Step 2A, Prong One"); and

[10] In response to received public comments, the Office issued further guidance on October 17, 2019, clarifying the 2019 Revised Guidance. USPTO, October 2019 Update: Subject Matter Eligibility (the "October 2019 Update") (available at https://www.uspto.gov/sites/default/files/documents/peg_oct_2019_update.pdf).

> (2) additional elements that integrate the judicial exception into a practical application (*see* MPEP § 2106.05(a)–(c), (e)–(h) (9th ed. Rev. 08.2017, Jan. 2018)) ("Step 2A, Prong Two").[11]

Guidance, 84 Fed. Reg. at 52–55.

Only if a claim (1) recites a judicial exception and (2) does not integrate that exception into a practical application, do we then look, under Step 2B, to whether the claim:

> (3) adds a specific limitation beyond the judicial exception that is not "well-understood, routine, conventional" in the field (*see* MPEP § 2106.05(d)); or
>
> (4) simply appends well-understood, routine, conventional activities previously known to the industry, specified at a high level of generality, to the judicial exception.

Guidance, 84 Fed. Reg. at 56.

ANALYSIS

Appellant's independent claims 68 and 76, reproduced above, are representative.

(Step 1)

The methods of Appellant's claims 68 and 76 are a statutory category of invention (*see* Ans. 8). Therefore, we proceed to the next step of the analysis.

(Step 2A, Prong 1)

The Guidance identifies three judicially-excepted groupings identified by the courts as abstract ideas: (1) mathematical concepts, (2) certain

---

[11] This evaluation is performed by (a) identifying whether there are any additional elements recited in the claim beyond the judicial exception, and (b) evaluating those additional elements individually and in combination to determine whether the claim as a whole integrates the exception into a practical application. *See* Guidance — Section III(A)(2), 84 Fed. Reg. at 54–55.

methods of organizing human behavior such as fundamental economic principles or practices, and (3) mental processes. *See* Guidance, 84 Fed. Reg. at 52.

Examiner finds that the correcting step in the methods of Appellant's claims 68 and 76 recite a mental process, which is an abstract idea (*see* Ans. 8). We agree. Correcting for error or bias introduced during nucleic acid analysis workflow by: (a) collapsing target:differentiator tag combinations observed more than once into a single count, as recited in Appellant's claim 68, or (b) recognizing target/differentiator tags having the same sequence as being derived from the same input molecule, as recited in Appellant's claim 76, are concepts that can be performed in the human mind and are, therefore, mental steps, i.e., a judicial exception. *See* Guidance, 84 Fed. Reg. at 52.

(Step 2A, Prong 2)

Having determined that Appellant's claims recite a judicial exception, the Guidance requires an evaluation as to whether the claim as a whole integrates the recited judicial exception into a practical application of the exception. *See* Guidance, 84 Fed. Reg. at 54–55.

Methodology involved in nucleic acid, i.e., genetic, analysis intrinsically includes a certain amount of error and bias (*see generally* Spec. 1–3). For example, Appellant explains that

> certain types of genetic information can be under-represented or over-represented in a genetic analysis due to a combination of stochastic variation and systematic bias in any of the preparative stages (e.g., capture, amplification, etc.), determining stages (e.g., allele-specific detection, sequencing, etc.), data interpretation stages (e.g., determining whether the assay information is sufficient to identify a subject as homozygous or heterozygous), and/or other stages.

Petitioner Illumina, Inc., EX1004, p. 1187

(Spec. 2.) The methods of Appellant's claimed invention seek to correct this error and/or bias and, thereby, improve nucleic acid analysis technology. Specifically, Appellant explains:

> Each of the claims requires multiple physical laboratory steps, including tagging nucleic acid molecules with differentiator tags, amplifying the tagged nucleic acid molecules to generate amplicons, and sequencing the amplicons. The further steps of correcting for error or bias provide the benefits of the improved sequencing workflow provided by the claimed invention.

(Appeal Br. 17.) We agree.

> A claim that integrates a judicial exception into a practical application will apply, rely on, or use the judicial exception in a manner that imposes a meaningful limit on the judicial exception, such that the claim is more than a drafting effort designed to monopolize the judicial exception. When the exception is so integrated, then the claim is not directed to a judicial exception . . . and is eligible.

Guidance, 84 Fed. Reg. at 54; *see generally* Appeal Br. 16.

For the foregoing reasons, we find that Appellant's claims integrate the judicial exception into a practical application that improves nucleic acid analysis technology. This concludes the eligibility analysis on this record. Therefore, we need not address Step 2B of the guidance concerning what was, or was not, "well-understood, routine, conventional" in the field. *See* Guidance, 84 Fed. Reg. at 54 ("If the recited exception is integrated into a practical application of the exception, then the claim is eligible at Prong Two of revised Step 2A. This concludes the eligibility analysis.").

CONCLUSION

Rejection III: The preponderance of evidence of record fails to support Examiner's finding that Appellant's claimed invention is directed to

Petitioner Illumina, Inc., EX1004, p. 1188

patent-ineligible subject matter. The rejection of claims 68, 69, 71, 73, 74, and 76 under 35 U.S.C. § 101 is reversed.

IV–VI

Obviousness:

ISSUE

Does the preponderance of evidence relied upon by Examiner support a conclusion of obviousness?

IV

The rejection over Porreca:

ANALYSIS

Examiner finds that Appellant's differentiator tag sequence "broadly encompasses the targeting arms of a molecular inversion probe (MIP)" (Ans. 9–10). We are not persuaded.

Appellant's claimed methods amplify the target nucleic acid molecules using a nucleic acid that (a) binds the target, i.e., targeting arms, (*see* Spec. 12, 14–15, 42) and (b) contains a unique sequence, i.e., a differentiator tag sequence that is "useful to distinguish independent target molecules from each other thereby allowing an analysis to be based on a known number of individual target molecules" (Spec. 38). Thus, contrary to Examiner's assertion, Appellant's differentiator tag sequence is distinct from the targeting arms of a MIP (*see* Ans. 9–10 (Examiner finds that Porreca's differentiator tag sequence broadly encompasses the targeting arms of a MIP.)); *cf.* Porreca Decl. ¶ 8 (Declarant explains "that the targeting arms of the molecular inversion probes disclosed in . . . [Porreca] are not the same as differentiator tags" required by Appellant's claimed invention.)).

Petitioner Illumina, Inc., EX1004, p. 1189

Examiner recognizes that "Porreca does not specifically teach 'correcting for error or bias introduced during . . . workflow by collapsing target:differentiator tag combinations observed more than once into a single count'" (Ans. 10). To make up for this deficiency, Examiner finds that

> Porreca teaches,"All evaluated positions at which the genotype is homozygous were dominated by the correct base-call (n [=] 388). The data on heterozygous positions (n [=] 90) are more ambiguous. For example, if we were to call positions with allele frequencies between 0.2 and 0.8 as heterozygous, then only 25 of 90 heterozygous positions (31%) would be called correctly. An alternative perspective is that because the goal of variation discovery is often to interrogate for rare variants, the key task is to identify positions that deviate from the reference genome with high sensitivity and specificity. If we call positions with observed allele frequencies of [<]0.8 as likely to be mutated relative to the reference genome, we can estimate a sensitivity of 63% for detecting rare mutations while maintaining high specificity (sensitivity, 57 of 90 heterozygous positions had a reference allele frequency of less than 0.8; specificity, 0 of 313 homozygous positions had a reference allele frequency of less than 0.8)."

(Ans. 10–11 (citing Porreca 934).) Examiner, therefore, concludes that, before the effective filing date of Appellant's claimed invention, it would have been prima facie obvious, in view of Porreca, "to sequence the arms of the MIPS and collapse (or identify) sequences including arms with greater than 0.8 allele frequency and identify . . . them as homozygous" (Ans. 11). According to Examiner, a person of ordinary skill in this art "would have [had] a reasonable expectation of success as the artisan is merely using the guidance of Porreca to identify and collapse sequences that are homozygous . . . using techniques demonstrated by the art to work" (*id.*). We are not persuaded.

Petitioner Illumina, Inc., EX1004, p. 1190

Although Porreca may propose a method for reducing error or bias in nucleic acid analysis workflow, the methods set forth in Appellant's claims 68 and 76 require the correction for error or bias introduced during nucleic acid analysis workflow by collapsing, or recognizing, *target:differentiator tag combinations* observed more than once into a single count, or having the same sequence as being derived from the same input molecule, respectively. Examiner's conclusion of obviousness is based on the incorrect premise that Appellant's differentiator tag is the targeting arm of Porreca's MIP; it is not. Thus, Examiner's conclusion of obviousness fails from the outset, because there is no combination of target and differentiator tag in Porreca and, therefore, Porreca could not collapse or recognize *target:differentiator tag combinations* observed more than once. In sum, Examiner failed to establish an evidentiary basis on this record to support a conclusion that Porreca discloses every element of Appellant's claimed methods.

V–VI

The rejection over Porreca in combination with Wang or Schiffman:

ANALYSIS

Examiner relies on Porreca as discussed above (*see* Ans. 13, 14). In particular, Examiner finds that Porreca discloses "the detection of homozygous alleles by use of unique targeting arms of MIP[s] and recognizing or collapsing the combination of unique target arms and alleles (or target sequences)" (Ans. 13, 14). Examiner recognizes, however, that Porreca does not disclose "determination of copy number" or "determination of blood samples or circulating tumor nucleic acids" (Ans. 13, 14). Examiner, therefore, relies on Wang to disclose the use of MIPs to determine allele copy number (*see* Ans. 13–14) and Schiffman to disclose

Petitioner Illumina, Inc., EX1004, p. 1191

"the use of MIPs for analysis of blood samples [in] subject with leukemia[, which is] . . . a circulating cancer or tumor" (Ans. 14).

Based on Porreca in combination with Wang or Schiffman, Examiner concludes that, before the effective filing date of Appellant's claimed invention, it would have been prima facie obvious "to use MIPs to determine copy number," as disclosed by Wang (Ans. 14) and "to use blood samples from subjects with leukemia for examination by MIPs," as disclosed by Schiffman (Ans. 14–15). Examiner failed, however, to establish that either of Wang or Schiffman make up for the deficiencies in Porreca discussed in section IV above. Thus, the preponderance of evidence on this record fails to support Examiner's conclusion of obviousness.

CONCLUSION

The preponderance of evidence relied upon by Examiner fails to support a conclusion of obviousness.

Rejection IV: The rejection of claims 68, 69, and 76 under 35 U.S.C. § 103(a) as unpatentable over Porreca is reversed.

Rejection V: The rejection of claim 74 under 35 U.S.C. § 103(a) as unpatentable over the combination of Porreca and Wang is reversed.

Rejection VI: The rejection of claims 71 and 73 under 35 U.S.C. § 103(a) as unpatentable over the combination of Porreca and Schiffman is reversed.

Petitioner Illumina, Inc., EX1004, p. 1192

DECISION SUMMARY

The following table summarizes our decision:

Claim(s) Rejected	35 U.S.C. §	Reference(s)/ Basis	Affirmed	Reversed
68, 69, 71, 73, 74, 76	112	Indefiniteness		68, 69, 71, 73, 74, 76
68, 69, 71, 73, 74, 76	112	Written Description		68, 69, 71, 73, 74, 76
68, 69, 71, 73, 74, 76	101	Eligibility		68, 69, 71, 73, 74, 76
68, 69, 76	103(a)	Porreca		68, 69, 76
74	103(a)	Porreca, Wang		74
71, 73	103(a)	Porreca, Schiffman		71, 73
Overall Outcome				68, 69, 71, 73, 74, 76

REVERSED

Petitioner Illumina, Inc., EX1004, p. 1193

Notice of Allowability	Application No. 16/952,764	Applicant(s) Porreca et al.	
	Examiner STEVEN C POHNERT PhD	Art Unit 1634	AIA (FITF) Status No

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTOL-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1. ☐ This communication is responsive to _____.
   ☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2. ☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

3. ☑ The allowed claim(s) is/are 68-76 . As a result of the allowed claim(s), you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see **http://www.uspto.gov/patents/init_events/pph/index.jsp** or send an inquiry to **PPHfeedback@uspto.gov.**

4. ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

   Certified copies:

   a) ☐ All b) ☐ Some* c) ☐ None of the:

   1. ☐ Certified copies of the priority documents have been received.
   2. ☐ Certified copies of the priority documents have been received in Application No. _____ .
   3. ☐ Copies of the certified copies of the priority documents have been received in this national stage application from the International Bureau (PCT Rule 17.2(a)).

   * Certified copies not received: _____ .

Applicant has THREE MONTHS FROM THE "MAILING DATE" of this communication to file a reply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application.
THIS THREE-MONTH PERIOD IS NOT EXTENDABLE.

5. ☐ CORRECTED DRAWINGS (as "replacement sheets") must be submitted.
   ☐ including changes required by the attached Examiner's Amendment / Comment or in the Office action of Paper No./Mail Date _____ .

   Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the front (not the back) of each sheet. Replacement sheet(s) should be labeled as such in the header according to 37 CFR 1.121(d).

6. ☐ DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

Attachment(s)

1. ☐ Notice of References Cited (PTO-892)
2. ☐ Information Disclosure Statements (PTO/SB/08), Paper No./Mail Date _____.
3. ☐ Examiner's Comment Regarding Requirement for Deposit of Biological Material _____.
4. ☐ Interview Summary (PTO-413), Paper No./Mail Date. _____.
5. ☑ Examiner's Amendment/Comment
6. ☐ Examiner's Statement of Reasons for Allowance
7. ☐ Other _____.

/Steven Pohnert/
Primary Examiner, Art Unit 1634

Petitioner Illumina, Inc., EX1004, p. 1194

Notice of Pre-AIA or AIA Status

1. The present application is being examined under the pre-AIA first to invent provisions.

Reasons for Allowance

1. The following is an examiner's statement of reasons for allowance:

2. The application is allowed in view of the PTAB decision of 9/6/2023.

Any comments considered necessary by applicant must be submitted no later than the payment of the issue fee and, to avoid processing delays, should preferably accompany the issue fee. Such submissions should be clearly labeled "Comments on Statement of Reasons for Allowance."

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to STEVEN C POHNERT PhD whose telephone number is (571)272-3803. The examiner can normally be reached Monday- Friday about 6:00 AM-5:00 PM, every second Friday off.

Examiner interviews are available via telephone, in-person, and video conferencing using a USPTO supplied web-based collaboration tool. To schedule an interview, applicant is encouraged to use the USPTO Automated Interview Request (AIR) at http://www.uspto.gov/interviewpractice.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Winston Shen can be reached on 571-272-3157. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of published or unpublished applications may be obtained from Patent Center. Unpublished application information in Patent Center is available to registered users. To file and manage patent submissions in Patent Center, visit: https://patentcenter.uspto.gov. Visit https://www.uspto.gov/patents/apply/patent-center for more information about Patent Center and https://www.uspto.gov/patents/docx for information about filing in DOCX format. For additional questions, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Steven Pohnert/
Primary Examiner, Art Unit 1634

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

NOTICE OF ALLOWANCE AND FEE(S) DUE

179511 7590 09/28/2023

Withers Bergman LLP
101 Federal Street, Suite 1900
Boston, MA 02110

EXAMINER	
POHNERT, STEVEN C	
ART UNIT	PAPER NUMBER
1634	

DATE MAILED: 09/28/2023

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/952,764	11/19/2020	Gregory Porreca	MLB-002/06US 34468/40	7812

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	SMALL	$480	$0.00	$0.00	$480	12/28/2023

THE APPLICATION IDENTIFIED ABOVE HAS BEEN EXAMINED AND IS ALLOWED FOR ISSUANCE AS A PATENT. PROSECUTION ON THE MERITS IS CLOSED. THIS NOTICE OF ALLOWANCE IS NOT A GRANT OF PATENT RIGHTS. THIS APPLICATION IS SUBJECT TO WITHDRAWAL FROM ISSUE AT THE INITIATIVE OF THE OFFICE OR UPON PETITION BY THE APPLICANT. SEE 37 CFR 1.313 AND MPEP 1308.

THE ISSUE FEE AND PUBLICATION FEE (IF REQUIRED) MUST BE PAID WITHIN THREE MONTHS FROM THE MAILING DATE OF THIS NOTICE OR THIS APPLICATION SHALL BE REGARDED AS ABANDONED. THIS STATUTORY PERIOD CANNOT BE EXTENDED. SEE 35 U.S.C. 151. THE ISSUE FEE DUE INDICATED ABOVE DOES NOT REFLECT A CREDIT FOR ANY PREVIOUSLY PAID ISSUE FEE IN THIS APPLICATION. IF AN ISSUE FEE HAS PREVIOUSLY BEEN PAID IN THIS APPLICATION (AS SHOWN ABOVE), THE RETURN OF PART B OF THIS FORM WILL BE CONSIDERED A REQUEST TO REAPPLY THE PREVIOUSLY PAID ISSUE FEE TOWARD THE ISSUE FEE NOW DUE.

HOW TO REPLY TO THIS NOTICE:

I. Review the ENTITY STATUS shown above. If the ENTITY STATUS is shown as SMALL or MICRO, verify whether entitlement to that entity status still applies.

If the ENTITY STATUS is the same as shown above, pay the TOTAL FEE(S) DUE shown above.

If the ENTITY STATUS is changed from that shown above, on PART B - FEE(S) TRANSMITTAL, complete section number 5 titled "Change in Entity Status (from status indicated above)".

For purposes of this notice, small entity fees are 40% the amount of undiscounted fees, and micro entity fees are 20% the amount of undiscounted fees.

II. PART B - FEE(S) TRANSMITTAL, or its equivalent, must be completed and returned to the United States Patent and Trademark Office (USPTO) with your ISSUE FEE and PUBLICATION FEE (if required). If you are charging the fee(s) to your deposit account, section "4b" of Part B - Fee(s) Transmittal should be completed. If an equivalent of Part B is filed, a request to reapply a previously paid issue fee must be clearly made, and delays in processing may occur due to the difficulty in recognizing the paper as an equivalent of Part B.

III. All communications regarding this application must give the application number. Please direct all communications prior to issuance to Mail Stop ISSUE FEE unless advised to the contrary.

IMPORTANT REMINDER: Maintenance fees are due in utility patents issuing on applications filed on or after Dec. 12, 1980. It is patentee's responsibility to ensure timely payment of maintenance fees when due. More information is available at www.uspto.gov/PatentMaintenanceFees.

PTOL-85 (Rev. 02/11)

PART B - FEE(S) TRANSMITTAL

Complete and send this form, together with applicable fee(s), by mail or fax, or via EFS-Web.

By mail, send to: Mail Stop ISSUE FEE
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

By fax, send to: (571)-273-2885

INSTRUCTIONS: This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 5 should be completed where appropriate. All further correspondence will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications. **Because electronic patent issuance may occur shortly after issue fee payment, any desired continuing application should preferably be filed prior to payment of this issue fee in order not to jeopardize copendency.**

CURRENT CORRESPONDENCE ADDRESS (Note: Use Block 1 for any change of address)

179511 7590 09/28/2023
Withers Bergman LLP
101 Federal Street, Suite 1900
Boston, MA 02110

Note: A certificate of mailing can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing or transmission.

Certificate of Mailing or Transmission

I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Mail Stop ISSUE FEE address above, or being transmitted to the USPTO via EFS-Web or by facsimile to (571) 273-2885, on the date below.

(Typed or printed name)
(Signature)
(Date)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/952,764	11/19/2020	Gregory Porreca	MLB-002/06US 34468/40	7812

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	SMALL	$480	$0.00	$0.00	$480	12/28/2023

EXAMINER	ART UNIT	CLASS-SUBCLASS
POHNERT, STEVEN C	1634	506-002000

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.363).

❑ Change of correspondence address (or Change of Correspondence Address form PTO/AIA/122 or PTO/SB/122) attached.

❑ "Fee Address" indication (or "Fee Address" Indication form PTO/AIA/47 or PTO/SB/47; Rev 03-02 or more recent) attached. **Use of a Customer Number is required.**

2. For printing on the patent front page, list
(1) The names of up to 3 registered patent attorneys or agents OR, alternatively,
(2) The name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed.

1 _____
2 _____
3 _____

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)

PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. If an assignee is identified below, the document must have been previously recorded, or filed for recordation, as set forth in 37 CFR 3.11 and 37 CFR 3.81(a). Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE (B) RESIDENCE: (CITY and STATE OR COUNTRY)

Please check the appropriate assignee category or categories (will not be printed on the patent) : ❑ Individual ❑ Corporation or other private group entity ❑ Government

4a. Fees submitted: ❑ Issue Fee ❑ Publication Fee (if required)

4b. Method of Payment: *(Please first reapply any previously paid fee shown above)*

❑ Electronic Payment via Patent Center or EFS-Web ❑ Enclosed check ❑ Non-electronic payment by credit card (Attach form PTO-2038)

❑ The Director is hereby authorized to charge the required fee(s), any deficiency, or credit any overpayment to Deposit Account No. _____

5. **Change in Entity Status** (from status indicated above)

❑ Applicant certifying micro entity status. See 37 CFR 1.29

NOTE: Absent a valid certification of Micro Entity Status (see forms PTO/SB/15A and 15B), issue fee payment in the micro entity amount will not be accepted at the risk of application abandonment.

❑ Applicant asserting small entity status. See 37 CFR 1.27

NOTE: If the application was previously under micro entity status, checking this box will be taken to be a notification of loss of entitlement to micro entity status.

❑ Applicant changing to regular undiscounted fee status.

NOTE: Checking this box will be taken to be a notification of loss of entitlement to small or micro entity status, as applicable.

NOTE: This form must be signed in accordance with 37 CFR 1.31 and 1.33. See 37 CFR 1.4 for signature requirements and certifications.

Authorized Signature _____ Date _____

Typed or printed name _____ Registration No. _____

Petitioner Illumina, Inc., EX1004, p. 1198

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/952,764	11/19/2020	Gregory Porreca	MLB-002/06US 34468/40	7812

179511 7590 09/28/2023

Withers Bergman LLP
101 Federal Street, Suite 1900
Boston, MA 02110

EXAMINER	
POHNERT, STEVEN C	
ART UNIT	**PAPER NUMBER**
1634	

DATE MAILED: 09/28/2023

Determination of Patent Term Adjustment under 35 U.S.C. 154 (b)
(Applications filed on or after May 29, 2000)

The Office has discontinued providing a Patent Term Adjustment (PTA) calculation with the Notice of Allowance.

Section 1(h)(2) of the AIA Technical Corrections Act amended 35 U.S.C. 154(b)(3)(B)(i) to eliminate the requirement that the Office provide a patent term adjustment determination with the notice of allowance. See Revisions to Patent Term Adjustment, 78 Fed. Reg. 19416, 19417 (Apr. 1, 2013). Therefore, the Office is no longer providing an initial patent term adjustment determination with the notice of allowance. The Office will continue to provide a patent term adjustment determination with the Issue Notification Letter that is mailed to applicant approximately three weeks prior to the issue date of the patent, and will include the patent term adjustment on the patent. Any request for reconsideration of the patent term adjustment determination (or reinstatement of patent term adjustment) should follow the process outlined in 37 CFR 1.705.

Any questions regarding the Patent Term Extension or Adjustment determination should be directed to the Office of Patent Legal Administration at (571)-272-7702. Questions relating to issue and publication fee payments should be directed to the Customer Service Center of the Office of Patent Publication at 1-(888)-786-0101 or (571)-272-4200.

PTOL-85 (Rev. 02/11)

Petitioner Illumina, Inc., EX1004, p. 1199

OMB Clearance and PRA Burden Statement for PTOL-85 Part B

The Paperwork Reduction Act (PRA) of 1995 requires Federal agencies to obtain Office of Management and Budget approval before requesting most types of information from the public. When OMB approves an agency request to collect information from the public, OMB (i) provides a valid OMB Control Number and expiration date for the agency to display on the instrument that will be used to collect the information and (ii) requires the agency to inform the public about the OMB Control Number's legal significance in accordance with 5 CFR 1320.5(b).

The information collected by PTOL-85 Part B is required by 37 CFR 1.311. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 30 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, Virginia 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, Virginia 22313-1450. Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether disclosure of these records is required by the Freedom of Information Act.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspection or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

Electronic Acknowledgement Receipt	
EFS ID:	48677173
Application Number:	16952764
International Application Number:	
Confirmation Number:	7812
Title of Invention:	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS
First Named Inventor/Applicant Name:	Gregory Porreca
Customer Number:	179511
Filer:	Thomas Meyers/Sandra Richards
Filer Authorized By:	Thomas Meyers
Attorney Docket Number:	MLB-002/06US 34468/40
Receipt Date:	04-OCT-2023
Filing Date:	19-NOV-2020
Time Stamp:	09:20:39
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	DA
Payment was successfully received in RAM	$480
RAM confirmation Number	E202304025479501
Deposit Account	
Authorized User	
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Issue Fee Payment (PTO-85B)	MLB-002-06US-Issue-Fee.pdf	943664 89f2485bf227dcca1e1f303d352e7d4beeb256cb	no	1
Warnings:					
Information:					
2	Fee Worksheet (SB06)	fee-info.pdf	38195 378ff04497bcd2f2ff4bdb30c41843a0155749cf	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	981859		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

PART B - FEE(S) TRANSMITTAL

Complete and send this form, together with applicable fee(s), by mail or fax, or via EFS-Web.

By mail, send to: Mail Stop ISSUE FEE
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

By fax, send to: (571)-273-2885

INSTRUCTIONS: This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 5 should be completed where appropriate. All further correspondence will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications. **Because electronic patent issuance may occur shortly after issue fee payment, any desired continuing application should preferably be filed prior to payment of this issue fee in order not to jeopardize copendency.**

CURRENT CORRESPONDENCE ADDRESS (Note: Use Block 1 for any change of address)

179511 7590 09/28/2023
Withers Bergman LLP
101 Federal Street, Suite 1900
Boston, MA 02110

Note: A certificate of mailing can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing or transmission.

Certificate of Mailing or Transmission

I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Mail Stop ISSUE FEE address above, or being transmitted to the USPTO via EFS-Web or by facsimile to (571) 273-2885, on the date below.

(Typed or printed name)
(Signature)
(Date)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/952,764	11/19/2020	Gregory Porreca	MLB-002/06US 34468/40	7812

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	SMALL	$480	$0.00	$0.00	$480	12/28/2023

EXAMINER	ART UNIT	CLASS-SUBCLASS
POHNERT, STEVEN C	1634	506-002000

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.363).

☐ Change of correspondence address (or Change of Correspondence Address form PTO/AIA/122 or PTO/SB/122) attached.

☐ "Fee Address" indication (or "Fee Address" Indication form PTO/AIA/47 or PTO/SB/47; Rev 03-02 or more recent) attached. **Use of a Customer Number is required.**

2. For printing on the patent front page, list
(1) The names of up to 3 registered patent attorneys or agents OR, alternatively,
(2) The name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed.

1 Withers Bergman LLP
2 Thomas C. Meyers
3

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)

PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. If an assignee is identified below, the document must have been previously recorded, or filed for recordation, as set forth in 37 CFR 3.11 and 37 CFR 3.81(a). Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE: MOLECULAR LOOP BIOSCIENCES, INC.
(B) RESIDENCE: (CITY and STATE OR COUNTRY): CAMBRIDGE, MA (US)

Please check the appropriate assignee category or categories (will not be printed on the patent) : ☐ Individual ☒ Corporation or other private group entity ☐ Government

4a. Fees submitted: ☒ Issue Fee ☐ Publication Fee (if required)

4b. Method of Payment: *(Please first reapply any previously paid fee shown above)*

☒ Electronic Payment via Patent Center or EFS-Web ☐ Enclosed check ☐ Non-electronic payment by credit card (Attach form PTO-2038)

☒ The Director is hereby authorized to charge the required fee(s), any deficiency, or credit any overpayment to Deposit Account No. 604011

5. **Change in Entity Status** (from status indicated above)

☐ Applicant certifying micro entity status. See 37 CFR 1.29
NOTE: Absent a valid certification of Micro Entity Status (see forms PTO/SB/15A and 15B), issue fee payment in the micro entity amount will not be accepted at the risk of application abandonment.

☐ Applicant asserting small entity status. See 37 CFR 1.27
NOTE: If the application was previously under micro entity status, checking this box will be taken to be a notification of loss of entitlement to micro entity status.

☐ Applicant changing to regular undiscounted fee status.
NOTE: Checking this box will be taken to be a notification of loss of entitlement to small or micro entity status, as applicable.

NOTE: This form must be signed in accordance with 37 CFR 1.31 and 1.33. See 37 CFR 1.4 for signature requirements and certifications.

Authorized Signature /Thomas C. Meyers/ Date October 4, 2023

Typed or printed name Thomas C. Meyers Registration No. 36,989

PTOL-85 Part B (08-18) Approved for use through 01/31/2020 OMB 0651-0033 U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Petitioner Illumina, Inc., EX1004, p. 1203

Electronic Patent Application Fee Transmittal

Application Number:	16952764
Filing Date:	19-Nov-2020
Title of Invention:	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS
First Named Inventor/Applicant Name:	Gregory Porreca
Filer:	Thomas Meyers/Sandra Richards
Attorney Docket Number:	MLB-002/06US 34468/40

Filed as Small Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
UTILITY APPL ISSUE FEE	2501	1	480	480

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Extension-of-Time:				
Miscellaneous:				
		Total in USD ($)		**480**

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/952,764	11/19/2020	Gregory Porreca	MLB-002/06US 1006120/13	7812

179511 7590 10/10/2023

Withers Bergman LLP
101 Federal Street, Suite 1900
Boston, MA 02110

EXAMINER	
POHNERT, STEVEN C	

ART UNIT	PAPER NUMBER
1634	

NOTIFICATION DATE	DELIVERY MODE
10/10/2023	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

Samantha.Metcalf@withersworldwide.com
US.Docketing@withersworldwide.com
kristina.scandore@withersworldwide.com

UNITED STATES PATENT AND TRADEMARK OFFICE

Commissioner for Patents
United States Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450
www.uspto.gov

Application No. : 16952764
Applicant : Porreca
Filing Date : 11/19/2020
Date Mailed : 10/10/2023

NOTICE TO FILE CORRECTED APPLICATION PAPERS

Notice of Allowance Mailed

This application has been accorded an Allowance Date and is being prepared for issuance. The application, however, is incomplete for the reasons below.

Applicant is given two (2) months from the mail date of this Notice within which to respond. This time period for reply is extendable under 37 CFR 1.136(a) for only TWO additional MONTHS.

The informalities requiring correction are indicated in the attachment(s). If the informality pertains to the abstract, specification (including claims) or drawings, the informality must be corrected with an amendment in compliance with 37 CFR 1.121 (or, if the application is a reissue application, 37 CFR 1.173). Such an amendment may be filed after payment of the issue fee if limited to correction of informalities noted herein. See Waiver of 37 CFR 1.312 for Documents Required by the Office of Patent Publication, 1280 Off. Gaz. Patent Office 918 (March 23, 2004). In addition, if the informality is not corrected until after payment of the issue fee, for purposes of 35 U.S.C. 154(b)(1)(iv), "all outstanding requirements" will be considered to have been satisfied when the informality has been corrected. A failure to respond within the above-identified time period will result in the application being ABANDONED.

See attachment(s).

A copy of this notice MUST be returned with the reply. Please address response to "Mail Stop Issue Fee, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450".

/Daxa Patel/
Publication Branch
Office of Data Management
(571) 272-4200

Application No. 16952764 **Drawings filed: 11/19/2020**

IDENTIFICATION OF DRAWING DEFICIENCIES

- [ ] There is a hole or the image thereof within the illustration. FIG(s)
- [ ] The illustration is penetrated or traversed by a solid or broken line that is not intended to be part of the drawing, such as a dark line caused by a flaw in the copying process. FIG(s)
- [ ] An ink stamp or the image thereof obscures part of the illustration. FIG(s)
- [ ] The drawing is marred by black smudges, obliterations, or fax/copier marks (for example, speckles or dots in a substantial portion of the drawing). FIG(s)
- [ ] Figure numbers are duplicated or missing. FIG(s)
- [ ] Drawing sheet or figure is missing. FIG(s)
- [ ] Numbers, letters, or reference characters in the drawing have been crossed out or are illegibly handwritten. FIG(s)
- [X] The character of the lines, numbers, and letters is poor. FIG(s) 10
- [ ] The drawing's background shows that the original drawing was made on graph paper or other paper with a pattern or decoration. FIG(s)
- [ ] The FIG. number label is placed in a location that causes the drawing to be read upside down. FIG(s)
- [ ] Data, a reference number, or part of the drawing is truncated or missing, or a lead line has no reference number. FIG(s)
- [ ] The drawing and/or the FIG. label contain(s) foreign language. FIG(s)
- [ ] This utility application contains a photograph of a view that is capable of being illustrated as a line drawing. FIG(s)
- [ ] A petition under 37 CFR 1.84(a)(2) to accept color drawings has been granted, but the brief description of the drawings in the specification does not contain (or has not been amended to contain) the paragraph required by 37 CFR 1.84(a)(2)(iii).
- [ ] This reissue application contains added and/or amended drawings that are not labeled as "New" or "Amended" or "Canceled" as required by 37 CFR 1.173(b)(3). FIG(s)
- [ ] This Design reissue application contains a drawing that is labeled as "Canceled" but is not surrounded by brackets, or a drawing that is surrounded by brackets but is not labeled as "Canceled." See 37 CFR 1.173(b)(3). FIG(s)
- [ ] OTHER:
- [ ] COMMENTS:

UNITED STATES
PATENT AND TRADEMARK OFFICE

P.O. Box 1450
Alexandria, VA 22313 - 1450
www.uspto.gov

ELECTRONIC ACKNOWLEDGEMENT RECEIPT

APPLICATION #	RECEIPT DATE / TIME	ATTORNEY DOCKET #
16/952,764	**10/18/2023 09:56:21 AM ET**	**MLB-002/06US 1006120/13**

Title of Invention

METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS

Application Information

APPLICATION TYPE	Utility - Nonprovisional Application under 35 USC 111(a)	PATENT #	-
CONFIRMATION #	7812	FILED BY	Sandra Richards
PATENT CENTER #	63020729	FILING DATE	11/19/2020
CUSTOMER #	179511	FIRST NAMED INVENTOR	Gregory Porreca
CORRESPONDENCE ADDRESS	-	AUTHORIZED BY	Thomas Meyers

Documents

TOTAL DOCUMENTS: 2

DOCUMENT	PAGES	DESCRIPTION	SIZE (KB)
MLB-002-06US-Response-CAP-101823.pdf	1	Applicant Response to Pre-Exam Formalities Notice	71 KB
MLB-002-06US-Replacement-sheet.pdf	1	Drawings-only black and white line drawings	12 KB

Digest

DOCUMENT	MESSAGE DIGEST(SHA-512)
MLB-002-06US-Response-CAP-101823.pdf	984931F4FFB7D6B70DC7A3B86596A1A41DA0B0C83B786C30B235EAB7E4218A78AAC5B605D2A6BE7DFB03F69BBDD88DF60

33C234396560D97AB938E95C5189648

MLB-002-06US-Replacement-sheet.pdf	C4E6C5606CAF9BCAD18981A2082659B48697FE00CCDC2E7B A2546AF79A38B0F57BF1CE378F193A2164784807F605E403D8 4894367E30B9B162D39636F40865B5

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Attorney Docket No.: MLB-002/06US 1006120/13

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANT:	Molecular Loop Biosciences, Inc.		
APPLICATION NO.:	16/952,764	ART UNIT:	1634
FILING DATE:	November 19, 2020	CONF. NO.:	7812
EXAMINER:	Steven C Pohnert		
TITLE:	METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS		

Commissioner for Patents
P.O. Box. 1450
Alexandria, VA 22313-1450

FILED ELECTRONICALLY

RESPONSE TO NOTICE TO FILE CORRECTED APPLICATION PAPERS

This paper is in response to the Notice to File Corrected Application Papers mailed October 10, 2023, from the United States Patent and Trademark Office. Applicant submits herewith 1 Replacement Sheet of drawings showing FIG. 10 with character of lines, numbers, and letters that is not poor. Applicant requests entry of the Replacement Sheet.

Applicant believes no fee is due but authorizes any required fee to be charged to deposit account 604011.

Dated October 18, 2023

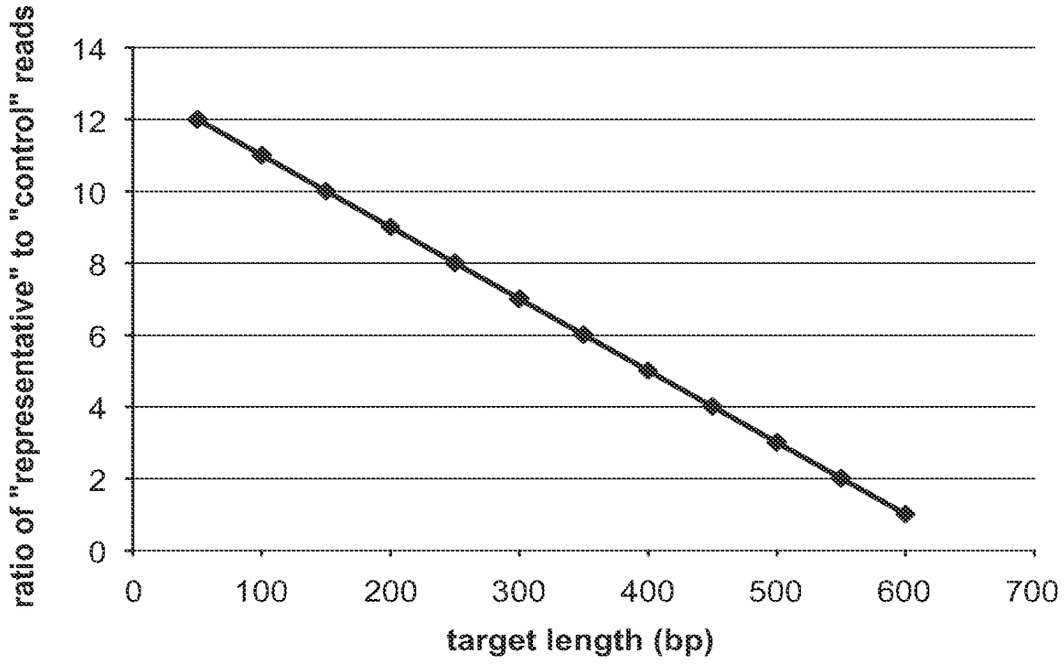
Respectfully submitted,
WITHERS BERGMAN LLP

WITHERS BERGMAN LLP
101 Federal St. Ste. 1900
Boston MA 02110

/Thomas C. Meyers/
Thomas C. Meyers, Reg. No. 36,989
Attorney for Applicant
Email: thomas.meyers@withersworldwide.com

Petitioner Illumina, Inc., EX1004, p. 1211

FIG. 10


ratio of "representative" to "control" reads
14
12
10
8
6
4
2
0
0
100
200
300
400
500
600
700
target length (bp)

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	ISSUE DATE	PATENT NO.	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/952,764	12/12/2023	11840730	MLB-002/06US 1006120/13	7812

179511 7590 11/22/2023

Withers Bergman LLP
101 Federal Street, Suite 1900
Boston, MA 02110

ISSUE NOTIFICATION

The projected patent number and issue date are specified above. The patent will issue electronically. The electronically issued patent is the official patent grant pursuant to 35 U.S.C. § 153. The patent may be accessed on or after the issue date through Patent Center at https://patentcenter.uspto.gov/. The patent will be available in both the public and the private sides of Patent Center. Further assistance in electronically accessing the patent, or about Patent Center, is available by calling the Patent Electronic Business Center at 1-888-217-9197.

The USPTO is implementing electronic patent issuance with a transition period, during which period the USPTO will mail a ceremonial paper copy of the electronic patent grant to the correspondence address of record. Additional copies of the patent (i.e., certified and presentation copies) may be ordered for a fee from the USPTO's Certified Copy Center at https://certifiedcopycenter.uspto.gov/index.html. The Certified Copy Center may be reached at (800)972-6382.

Determination of Patent Term Adjustment under 35 U.S.C. 154 (b)
(application filed on or after May 29, 2000)

The Patent Term Adjustment is 516 day(s). Any patent to issue from the above-identified application will include an indication of the adjustment on the front page.

If a Continued Prosecution Application (CPA) was filed in the above-identified application, the filing date that determines Patent Term Adjustment is the filing date of the most recent CPA.

Applicant will be able to obtain more detailed information by accessing the Patent Center (https://patentcenter.uspto.gov).

Any questions regarding the Patent Term Extension or Adjustment determination should be directed to the Office of Patent Legal Administration at (571)-272-7702. Questions relating to issue and publication fee payments should be directed to the Application Assistance Unit (AAU) of the Office of Patents Stakeholder Experience (OPSE), Stakeholder Support Division (SSD) at (571)-272-4200.

INVENTOR(s) (Please see PATENT CENTER site https://patentcenter.uspto.gov for additional inventors):

Gregory Porreca, Cambridge, MA;
Uri Laserson, Boston, MA;

APPLICANT(s) (Please see PATENT CENTER site https://patentcenter.uspto.gov for additional applicants):

Molecular Loop Biosciences, Inc., Cambridge, MA;

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The USA offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to encourage and facilitate business investment. To learn more about why the USA is the best country in the world to develop technology, manufacture products, and grow your business, visit SelectUSA.gov.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/952,764	11/19/2020	Gregory Porreca	MLB-002/06US 1006120/13	7812

179511 7590 12/12/2023

Withers Bergman LLP
101 Federal Street, Suite 1900
Boston, MA 02110

EXAMINER	
POHNERT, STEVEN C	

ART UNIT	PAPER NUMBER
1634	

NOTIFICATION DATE	DELIVERY MODE
12/12/2023	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

Samantha.Metcalf@withersworldwide.com
US.Docketing@withersworldwide.com
kristina.scandore@withersworldwide.com

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	ISSUE DATE	PATENT NO.
16/952,764	12-Dec-2023	11840730

Withers Bergman LLP
101 Federal Street, Suite 1900
Boston, MA 02110

EGRANT NOTIFICATION

Your electronic patent grant (eGrant) is now available, which can be accessed via Patent Center at https://patentcenter.uspto.gov

The electronic patent grant is the official patent grant under 35 U.S.C. 153. For more information, please visit https://www.uspto.gov/electronicgrants

PATENT APPLICATION FEE DETERMINATION RECORD

Substitute for Form PTO-875

Application or Docket Number: 16/952,764

APPLICATION AS FILED - PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY FEE($)
BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A	N/A	80		N/A	
SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A	N/A	350		N/A	
EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A	400		N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	9 minus 20 =	*	x 50 =	0.00	OR		
INDEPENDENT CLAIMS (37 CFR 1.16(h))	2 minus 3 =	*	x 240 =	0.00			
APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).			0.00			
MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))				0.00			
* If the difference in column 1 is less than zero, enter "0" in column 2.			TOTAL	830		TOTAL	

APPLICATION AS AMENDED - PART II

		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY ADDITIONAL FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY ADDITIONAL FEE($)
AMENDMENT A	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	RATE($)	ADDITIONAL FEE($)		RATE($)	ADDITIONAL FEE($)
AMENDMENT B	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".
*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".
The "Highest Number Previously Paid For" (Total or Independent) is the highest found in the appropriate box in column 1.

PTO/SB/06 (09-11)
Approved for use through 1/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PATENT APPLICATION FEE DETERMINATION RECORD Substitute for Form PTO-875	Application or Docket Number 16/952,764	Filing Date 11/19/2020	☐ To be Mailed

ENTITY: ☐ LARGE ☑ SMALL ☐ MICRO

APPLICATION AS FILED - PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA		RATE ($)	FEE ($)
☐ BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A		N/A	
☐ SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A		N/A	
☐ EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A		N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	minus 20 =	*		x $50 =	
INDEPENDENT CLAIMS (37 CFR 1.16(h))	minus 3 =	*		x $240 =	
☐ APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).				
☐ MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))					
* If the difference in column 1 is less than zero, enter "0" in column 2.				TOTAL	

APPLICATION AS AMENDED - PART II

AMENDMENT	03/05/2021	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	* 9	Minus	** 20	= 0		x $50 =	0
	Independent (37 CFR 1.16(h))	* 2	Minus	*** 3	= 0		x $240 =	0
	☐ Application Size Fee (37 CFR 1.16(s))							
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							
							TOTAL ADD'L FEE	0

AMENDMENT		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	*	Minus	**	=		x $0 =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=		x $0 =	
	☐ Application Size Fee (37 CFR 1.16(s))							
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							
							TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.	SLIE
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".	/JULIET MCMILLAN/
*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".	
The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1.	

This collection of information is required by 37 CFR 1.16. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

PTO/SB/06 (09-11)
Approved for use through 1/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PATENT APPLICATION FEE DETERMINATION RECORD Substitute for Form PTO-875	Application or Docket Number 16/952,764	Filing Date 11/19/2020	☐ To be Mailed

ENTITY: ☑ LARGE ☐ SMALL ☐ MICRO

APPLICATION AS FILED - PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA		RATE ($)	FEE ($)
☐ BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A		N/A	
☐ SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A		N/A	
☐ EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A		N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	minus 20 =	*		x $100 =	
INDEPENDENT CLAIMS (37 CFR 1.16(h))	minus 3 =	*		x $480 =	
☐ APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).				
☐ MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))					
* If the difference in column 1 is less than zero, enter "0" in column 2.				TOTAL	

APPLICATION AS AMENDED - PART II

AMENDMENT	02/01/2021	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	* 9	Minus	** 20	= 0		x $100 =	0
	Independent (37 CFR 1.16(h))	* 2	Minus	*** 3	= 0		x $480 =	0
	☐ Application Size Fee (37 CFR 1.16(s))							
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							
							TOTAL ADD'L FEE	0

AMENDMENT		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	*	Minus	**	=		x $0 =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=		x $0 =	
	☐ Application Size Fee (37 CFR 1.16(s))							
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							
							TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.	SLIE
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".	/SONYA M WILLIAMS/
*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".	
The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1.	

This collection of information is required by 37 CFR 1.16. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

PATENT APPLICATION FEE DETERMINATION RECORD

Substitute for Form PTO-875

Application or Docket Number: 16/952,764

APPLICATION AS FILED - PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY FEE($)
BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A	N/A	80		N/A	
SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A	N/A	350		N/A	
EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A	400		N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	9 minus 20 =	*	x 50 =	0.00	OR		
INDEPENDENT CLAIMS (37 CFR 1.16(h))	2 minus 3 =	*	x 240 =	0.00			
APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).			0.00			
MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))				0.00			
* If the difference in column 1 is less than zero, enter "0" in column 2.			TOTAL	830		TOTAL	

APPLICATION AS AMENDED - PART II

		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY ADDITIONAL FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY ADDITIONAL FEE($)
AMENDMENT A	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	RATE($)	ADDITIONAL FEE($)		RATE($)	ADDITIONAL FEE($)
AMENDMENT B	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".
*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".
The "Highest Number Previously Paid For" (Total or Independent) is the highest found in the appropriate box in column 1.

PATENT APPLICATION FEE DETERMINATION RECORD

Substitute for Form PTO-875

Application or Docket Number: 16/952,764

APPLICATION AS FILED - PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY FEE($)
BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A	N/A	80		N/A	
SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A	N/A	350		N/A	
EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A	400		N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	9 minus 20 =	*	x 50 =	0.00	OR		
INDEPENDENT CLAIMS (37 CFR 1.16(h))	2 minus 3 =	*	x 240 =	0.00			
APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).			0.00			
MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))				0.00			
* If the difference in column 1 is less than zero, enter "0" in column 2.			TOTAL	830		TOTAL	

APPLICATION AS AMENDED - PART II

		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY ADDITIONAL FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY ADDITIONAL FEE($)
AMENDMENT A	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	RATE($)	ADDITIONAL FEE($)		RATE($)	ADDITIONAL FEE($)
AMENDMENT B	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".
*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".
The "Highest Number Previously Paid For" (Total or Independent) is the highest found in the appropriate box in column 1.

Derwent Innovation Patent Export, 2024-07-12 02:38:24 +0000

Search results for: PNS=(US11840730);
Collections searched: DWPI; ARIPO grant; Argentina UM; Argentina app; Armenia grant; Australia UM; Australia app; Australia grant; Austria UM; Austria app; Austria grant; Belarus UM; Belarus grant; Belgium app; Belgium grant; Brazil UM; Brazil app; Brazil grant; Bulgaria UM; Bulgaria app; Bulgaria grant; Canada app; Canada grant; China, Mainland UM; China, Mainland app; China, Mainland grant; Colombia app; Costa Rica UM; Costa Rica app; Croatia UM; Croatia app; Croatia grant; Cuba app; Cuba grant; Czech Republic UM; Czech Republic app; Czech Republic grant; Czechoslovakia grant; Denmark UM; Denmark app; Denmark grant; Estonia UM; Estonia app; Estonia grant; Eurasian (EAPO) grant; Europe (EP) app; Europe (EP) grant; Finland app; Finland grant; France app; France grant; Georgia grant; German Democratic Republic app; German Democratic Republic grant; Germany UM; Germany app; Germany grant; Greece app; Greece grant; Gulf Cooperation Council grant; Hong Kong UM; Hong Kong app; Hong Kong grant; Hungary grant; Iceland grant; India app; India grant; Indonesia UM; Indonesia app; Ireland UM; Ireland app; Ireland grant; Israel grant; Italy grant; Italy grant; Italy grant; Italy grant; Japan UM; Japan app; Japan grant; Korea UM; Korea app; Korea grant; Latvia grant; Lithuania grant; Luxembourgh app; Luxembourgh grant; Malaysia app; Malaysia grant; Mexico app; Mexico grant; Moldova UM; Moldova grant; Monaco grant; Mongolia UM; Mongolia grant; Morocco app; Morocco grant; Netherlands app; Netherlands grant; New Zealand grant; Norway app; Norway grant; OAPI grant; PCT (WO) app; Philippines UM; Philippines app; Philippines grant; Poland UM; Poland grant; Portugal UM; Portugal app; Portugal grant; Romania UM; Romania app; Romania granted; Russia UM; Russia app; Russia grant; Saudi Arabia app; Saudi Arabia grant; Serbia UM; Serbia app; Serbia grant; Singapore app; Singapore grant; Slovakia UM; Slovakia app; Slovakia grant; Slovenia UM; Slovenia grant; South Africa app; Spain UM; Spain app; Spain grant; Sweden app; Sweden grant; Switzerland app; Switzerland grant; Taiwan UM; Taiwan appl; Taiwan grant; Thailand UM; Thailand app; Thailand grant; Tunisia app; Turkey UM; Turkey app; UK app; UK grant; USSR grant; Ukraine UM; Ukraine grant; United States app; United States grant; Uruguay app; Vietnam UM; Vietnam app; Vietnam grant; Other

Table of Contents

1. US11840730B1 Methods and compositions for evaluating genetic markers

Family 1/1

1 record(s) per family

Record 1/1 US11840730B1 Methods and compositions for evaluating genetic markers

Title: Methods and compositions for evaluating genetic markers

Title - DWPI: Analyzing several genetic loci involves contacting each of several target nucleic acids with a probe set and isolating the nucleic acids each having a nucleic acid sequence of a different sub-region for each of the target nucleic acids

Priority Number: US2009174470P | US2009178923P | US2009179358P | US2009182089P | WO2010US1293A | US13266862A | US15231687A

Priority Date: 2009-04-30 | 2009-05-15 | 2009-05-18 | 2009-05-28 | 2010-04-30 | 2010-04-30 | 2016-08-08

Application Number: US16952764A

Application Date: 2020-11-19

Publication Date: 2023-12-12

IPC Class Table:

IPC	Section	Class	Subclass	Class Group	Subgroup
C12Q00016874	C	C12	C12Q	C12Q0001	C12Q00016874
C12Q00016827	C	C12	C12Q	C12Q0001	C12Q00016827
C12Q00016883	C	C12	C12Q	C12Q0001	C12Q00016883

IPC Class Table - DWPI:

IPC - DWPI	Section - DWPI	Class - DWPI	Subclass - DWPI	Class Group - DWPI	Subgroup - DWPI
C12Q000168 (IPC 1-7)	C	C12	C12Q	C12Q0001	C12Q000168 (IPC 1-7)
C12M000100	C	C12	C12M	C12M0001	C12M000100
C12N001509	C	C12	C12N	C12N0015	C12N001509
C12Q000168	C	C12	C12Q	C12Q0001	C12Q000168
C12Q00016827	C	C12	C12Q	C12Q0001	C12Q00016827
C12Q00016874	C	C12	C12Q	C12Q0001	C12Q00016874

Petitioner Illumina, Inc., EX1004, p. 1223

C12Q00016883	C	C12	C12Q	C12Q0001	C12Q00016883
C40B002000	C	C40	C40B	C40B0020	C40B002000
C40B003004	C	C40	C40B	C40B0030	C40B003004
G01N003348	G	G01	G01N	G01N0033	G01N003348
G01N003353	G	G01	G01N	G01N0033	G01N003353

Assignee/Applicant: MOLECULAR LOOP BIOSCIENCES INC.,Cambridge,MA,US

JP F Terms:

JP FI Codes:

Assignee - Original: MOLECULAR LOOP BIOSCIENCES INC.

Any CPC Table:

Type	CPC - Current - Invention	CPC - Current - Additional	Version	Office
Current	**C12Q 1/6827**	-	20130101	EP; US
Current Current	C12Q 1/6874 C12Q 1/6883	-	20130101 20130101	US US
Original Original Original	**C12Q 1/6874** C12Q 1/6827 C12Q 1/6883	-	20130101 20130101 20130101	US US US

ECLA:

Abstract:

Aspects of the invention relates to methods and compositions that are useful to reduce bias and increase the reproducibility of multiplex analysis of genetic loci. In some configurations, predetermined preparative steps and/or nucleic acid sequence analysis techniques are used in multiplex analyses for a plurality of genetic loci in a plurality of samples.

Language of Publication: EN

INPADOC Legal Status Table:

Gazette Date	Code	INPADOC Legal Status Impact
2023-11-22	STCF	-
Description: INFORMATION ON STATUS: PATENT GRANT PATENTED CASE		

Petitioner Illumina, Inc., EX1004, p. 1224

2020-11-30	FEPP	-
Description: FEE PAYMENT PROCEDURE ENTITY STATUS SET TO SMALL (ORIGINAL EVENT CODE: SMAL); ENTITY STATUS OF PATENT OWNER: SMALL ENTITY		
2020-11-19	FEPP	-
Description: FEE PAYMENT PROCEDURE ENTITY STATUS SET TO UNDISCOUNTED (ORIGINAL EVENT CODE: BIG.); ENTITY STATUS OF PATENT OWNER: SMALL ENTITY		

Post-Issuance (US):

Reassignment (US) Table:

Assignee	Assignor	Date Signed	Reel/Frame	Date
MOLECULAR LOOP BIOSCIENCES INC.,WILMINGTON,DE,US	MOLECULAR LOOP BIOSOLUTIONS, LLC	2019-04-11	057634/0714	2021-09-28
Conveyance: CHANGE OF NAME (SEE DOCUMENT FOR DETAILS).				
Corresponent: BROWN RUDNICK LLP ONE FINANCIAL CENTER BOSTON, MA 02111				
MOLECULAR LOOP BIOSOLUTIONS LLC,CAMBRIDGE,MA,US	GOOD START GENETICS, INC.	2018-10-12	054561/0341	2020-12-07
Conveyance: ASSIGNMENT OF ASSIGNORS INTEREST (SEE DOCUMENT FOR DETAILS).				
Corresponent: BROWN RUDNICK, LLP ONE FINANCIAL CENTER BOSTON, MA 02111				
GOOD START GENETICS INC.,CAMBRIDGE,MA,US	PORRECA, GREGORY	2012-03-01	054561/0315	2020-12-07
	LASERSON, URI	2012-02-28	-	-
Conveyance: ASSIGNMENT OF ASSIGNORS INTEREST (SEE DOCUMENT FOR DETAILS).				
Corresponent: BROWN RUDNICK, LLP ONE FINANCIAL CENTER BOSTON, MA 02111				

Maintenance Status (US):

Litigation (US): 2024-06-10 2024 Molecular Loop Biosciences Inc. Illumina Inc. Verinata Health Inc. Delaware District Court 1:24-cv-00680

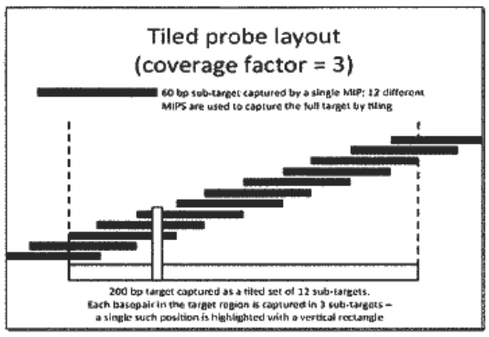
Opposition (EP):

License (EP):

EPO Procedural Status:

Front Page Drawing:

FIG. 1



Priority Number - DWPI: US2009174470P | US2009174923P | US2009178923P | US2009179358P | US2009182089P | WO2010US1293A | US13266862A | US13266862A | US2016231687A | US2020952764A
Priority Date - DWPI: 2009-04-30 | 2009-05-01 | 2009-05-15 | 2009-05-18 | 2009-05-28 | 2010-04-30 | 2012-03-13 | 2012-03-13 | 2016-08-08 | 2020-11-19
Assignee - Current US: MOLECULAR LOOP BIOSCIENCES INC.

Copyright 2007-2024 CLARIVATE

United States Patent and Trademark Office
Office of the Commissioner for Patents

METHODS AND COMPOSITIONS FOR EVALUATING GENETIC MARKERS

PATENT #	APPLICATION #	FILING DATE	ISSUE DATE
11840730	16952764	11/19/2020	12/12/2023

Payment Window Status

WINDOW	STATUS	FEES
3.5 Year	Not Open	Not Due

No maintenance fees are due at this time. 3.5 year window opens on 12/12/2026.

Window	First Day to Pay	Surcharge Starts	Last Day to Pay	Status	Fees
3.5 Year	12/12/2026	06/15/2027	12/13/2027	Not Open	Not Due
7.5 Year	12/12/2030	06/13/2031	12/12/2031	Not Open	Not Due
11.5 Year	12/12/2034	06/13/2035	12/12/2035	Not Open	Not Due

Patent Holder Information

Customer # 179511

Entity Status SMALL

Phone Number 6176139700

Address Withers Bergman LLP
100 Federal Street, Suite 1902
Boston, MA 02110
UNITED STATES