[00217] The amplified cDNA was further quantified by analysis of various transcripts in the amplification products using real time PCR. Aliquots of non-purified amplification products were diluted 1:80 in Tris-EDTA and 2 µl of the diluted amplification products were added to PCR mixture with SYBR Green and the specified PCR primers (sequence information given in Table 3). Relative quantification is given by Ct values for the various measured transcripts (amplified cDNA) in Table 2.

Table 2: Real time PCR quantification of specific amplified cDNA products (non-purified)

Input		PGK-1 (Ct)		GAPE 12 (Ct)		CD44-1 (Ct)		CD44-3 (Ct)	
RNA	DNA	WT-Ovation Pico	Selective amplification	WT-Ovation Pico	Selective amplification	WT-Ovation Pico	Selective amplification	WT-Ovation Pico	Selective amplification
1 ng	No DNA	19.1	18.8	17.2	18.1	23.3	20.4	19.2	21.4
		18.9	19.7	17.9	16.5	22.5	20.7	18.8	22.2
1ng	1ng	19.0	18.9	17.8	17.5	23.4	21.6	18.2	22.0
		19.1	19.2	17.8	17.5	22.8	21.4	18.0	22.0
1 ng	9 ng	19.9	19.3	18.2	18.0	23.3	20.8	20.0	20.3
		19.3	19.5	18.1	18.2	23.1	21.7	19.1	21.8
No RNA	9 ng	33.3	None	29.2	34.3	None	None	35.1	28.4
		23.2	35.9	31.5	33.0	None	None	33.2	34.7
No RNA	No DNA	None	None	39.5	None	None	None	None	None
		None	38.5	35.6	36.4	None	None	None	None

Table 3: PCR primer pairs used for quantification by real time PCR

Assay	PCR Primer sequences	Gene
GAPE-12	tccactggcgtcttcacc	glyceraldehyde-3-phosphate dehydrogenase
	ggcagagatgatgacccttttt	
CD44-3	ttgttactttgacttttcagagcac	CD44 antigen (homing function and Indian blood group system)
	gcaatatacatatcatgctttcctca	
CD44-1	gacaccatggacaagttttgg	CD44 antigen (homing function and Indian blood group system)
	cggcaggttatattcaaatcg	
PGK1	ctgtggcttctggcatacct	phosphoglycerate kinase 1
	cttgctgctttcaggacca	

Example 2: Whole transcriptome analysis of sequences from total RNA in the presence of DNA using actinomycin to inhibit DNA-dependent DNA polymerase activity

[00218] A RNA target is amplified using a method involving a first primer, comprising composite RNA/DNA portions, wherein the 3'-DNA end of the primers comprises either poly-dT sequence or a random sequence. First strand cDNA synthesis is performed with this first primer in the presence of Actinomycin D, an inhibitor of DNA-dependent DNA polymerase. Removal of Actinomycin D and RNA target from the reaction mixture is performed by binding the nucleic acids to magnetic beads. Second strand cDNA synthesis is then carried out and is followed by amplification, which is performed using a composite amplification primer, DNA polymerase with strand displacement activity, and RNase H. The global RNA amplification is carried out using the WT-Ovation Pico amplification system (NuGEN

IDT Exhibit 1020.2286

Technologies Inc. http://www.nugeninc.com/html/03_products5.html) with additional reagents and protocol modifications described below. The amplification products are then analyzed by sequencing with a Genome Sequencer platform (454/Roche Lifesciences). As described herein, the results of the example demonstrate selective amplification and analysis of RNA target from samples comprising mixtures of RNA and DNA.

Materials

[00219] All reagents for the WT-Ovation Pico RNA Amplification System (NuGEN Technologies Inc.) are used in this example. HeLa (S3) total RNA is purchased from Ambion. Human genomic DNA is purchased from Clontech. Random hexamer (N6 DNA oligonucleotide) is purchased from Operon. Actinomycin D is purchased from USB.

General comments

[00220] The WT-Ovation Pico amplification system is used to perform global RNA amplification. For selective amplification of RNA targets from samples comprising genomic DNA, the procedure is modified to include Actinomycin D in the step involving first strand cDNA synthesis. Additional clean up following first strand cDNA synthesis is added to the manufacturer protocol, which results in the removal of Actinomycin D prior to second strand cDNA synthesis. In addition, second strand cDNA synthesis is carried out in the presence of random hexamers (DNA oligonucleotides). All other amplification steps are the same as for the WT-Ovation Pico RNA Amplification System. The sequencing analysis is performed according to the manufacturers instructions.

First strand cDNA synthesis

[00221] First stand cDNA synthesis is initiated randomly across the length of RNA transcripts using composite first primers containing a 3'-random sequence DNA portion (for example, a random hexamer sequence) and a 5'-non-complementary and non-hybridizable RNA portion. In addition to the composite first primer, another composite first primer, comprising a 3'-poly-DT sequence DNA portion and the same 5'-non-hybridizing RNA portion as the previous primer, is used. In both the WT-Ovation Pico RNA amplification protocol and the modified procedure of the current invention, 1 to 2 µl of total RNA sample comprising 0.5 to 50 ng total RNA purified from a sample to be analyzed, with or without DNA, are mixed with 1 to 2 µl of a mixture containing the two first primers. The resulting mixture is transferred to a heating block or a thermal cycler and incubated at 65°C for 2 minutes. The mixture is then removed from the heating block or thermocycler and placed on ice.

[00222] A second mixture containing reverse transcriptase and a buffer is added to the first mixture above to a total volume of 10 µl. In samples subjected to the selective RNA amplification, Actinomycin D is added to the mixture to a final concentration of 50 ng/µl. The reactions are carried out by incubation at the following temperatures and durations: 4°C for 1min, 25°C for 10 min, 42°C for 10 min, 70°C for 15 min., 4°C and hold. The reaction mixture is then placed on ice.

IDT Exhibit 1020.2287

Removal of Actinomycin D

[00223] Prior to second strand cDNA synthesis, removal of the DNA-dependent DNA polymerase and Actinomycin D is performed using magnetic beads under conditions that allow binding of the nucleic acids in the mixture to the beads. This clean up step allows separation of the reaction medium containing Actinomycin D from the nucleic acids and is conducted following the manufacturer instructions. To the reaction mixture above, 16 µl of bead suspensions (RNAClean, Agencourt) are added and the mixture is incubated at room temperature for 10 min. The beads are separated using a magnetic plate and the supernatant is discarded. The beads with the bound nucleic acids from the above reaction mixture are washed three times with 70% ethanol and air dried for 10 min at room temperature. This clean up step is performed for samples subjected to the selective RNA amplification according to the method of the invention and is a modified protocol of the WT-Ovation Pico RNA amplification.

Second strand cDNA synthesis

[00224] A mixture of DNA polymerase and buffer is added to the first strand synthesis reaction mixture for synthesis of the second strand cDNA. In the case of the modified protocol for selective RNA amplification according the method of the current invention, the reaction mixture also contains random hexamers (N6 DNA oligonucleotides) at 100 ng/µl. Second strand cDNA synthesis is carried out in a total volume of 20µl. For selective RNA amplification according to the method of the current invention, the beads used for the removal of Actinomycin D, as described above, are left in the second strand cDNA synthesis reaction mixture. Second strand cDNA synthesis reactions are carried out by incubation at the following temperatures for the following durations: 4°C for 1 min, 25°C for 10 min, 50°C for 30 min, 70°C for 5 min, and cooled to 4°C.

Purification of the double stranded cDNA

[00225] To each second strand cDNA reaction mixture, 32 µl of bead suspension (RNAClean® Agencourt) are added and the mixture is incubated for 10 min at room temperature. The beads are separated on a magnetic plate and further processed as described herein.

Amplification

[00226] Amplification is carried out according to the manufacturer's instruction with modification of the total reaction volume from 160 µl to 80 µl (www.nugeninc.com).

[00227] Beads are suspended in the amplification mixture containing 20 µl of composite amplification primer, 40 µl of reaction buffer, and 20 µl enzyme mixture. The amplification reactions are carried out at following temperatures and durations: 40°C for 1 min., 47°C for 60 min., and 95°C for 5 min. The reactions are cooled to 4°C.

IDT Exhibit 1020.2288

Analysis

The amplified polynucleotides are diluted 1,000-fold, immobilized onto beads, and compartmentalized in a water-in-oil emulsion in the presence of PCR primers and reagents including buffer, enzyme and dNTP's. The immobilized polynucleotides are amplified by thermocycling in the water-in-oil emulsion for 20 cycles. When the emulsion is broken, amplified fragments remain bound to the beads. The beads are enriched to select for optimally sized amplicons and loaded into wells of a fiber optic slide so that there is approximately 1 bead in each well. Nucleotides are flowed across and into the wells in a fixed order in the presence of polymerase, sulfhydrolase, and luciferase. Addition of nucleotides complementary to the target strand results in a chemiluminescent signal that is recorded by a camera. The combination of signal intensity and positional information generated across the plate allows software to determine the DNA sequence.

Results

[00228] The selective amplification of RNA from samples comprising low levels of input total HeLa RNA, with or without excess genomic human DNA, employs a modified protocol of the WT-Ovation Pico RNA amplification system, which includes Actinomycin D in the first strand cDNA synthesis (to inhibit DNA-dependent DNA polymerase) and removal of Actinomycin D prior to second strand cDNA synthesis. The resulting selectively amplified polynucleotide targets are then used to perform whole-transcriptome analysis by massively parallel sequencing. The massively parallel sequencing identifies: (i) transcript abundance at both the exon and transcript level; (ii) transcriptional start and stop sites; (iii) exon-exon linkages in mature transcripts; (iv) single-base changes corresponding to SNPs and mutations; and (v) novel exons belonging to already characterized genes. The massively parallel sequencing also provides numerous examples of novel splicing events when the reads that could not be mapped directly to the genome were searched against a database of all possible exon-exon junctions within and between genes. A large number of reads appear to represent bona fide transcripts of novel expressed elements. The sequence data also allows the identification of single-base variations corresponding to novel SNPs and mutations, with many of these genes either cell cycle–or E2F transcription–related..

[00229] Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be apparent to those skilled in the art that certain changes and modifications may be practiced. Therefore, the descriptions and examples should not be construed as limiting the scope of the invention, which is delineated by the appended claims.

[00230] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will be apparent to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

IDT Exhibit 1020.2289

CLAIMS

WHAT IS CLAIMED IS:

1. A method of selectively generating multiple copies of a polynucleotide sequence of or complementary to target RNA which is in a sample comprising DNA, said method comprising the steps of:

(a) hybridizing to the target RNA a first primer comprising a sequence (A) that is not complementary to the target RNA, and a sequence (B) at the 3'-end which hybridizes to the target RNA;

(b) extending the first primer with at least one enzyme comprising RNA-dependent DNA polymerase activity in the presence of at least one compound comprising DNA-dependent DNA polymerase inhibitor activity, whereby a complex comprising a primer extension product and the target RNA is produced, whereby the first primer extension product comprises a sequence (Y) that is complementary to the target RNA and comprises sequence (A);

(c) disabling or removing at least one compound comprising DNA-dependent DNA polymerase inhibitor activity; and

(d) producing multiple copies of a polynucleotide sequence complementary to the target RNA and/or complementary to sequence (Y) using sequence (A).

2. A method of selectively generating multiple copies of a polynucleotide sequence complementary to a target RNA which is in a sample comprising DNA, said method comprising the steps of:

(a) hybridizing to the target RNA a first primer comprising a sequence (A) that does not hybridize to the target RNA, and a sequence (B) at the 3'-end, which hybridizes to the target RNA;

(b) extending the first primer with at least one enzyme comprising RNA-dependent DNA polymerase activity in the presence of at least one compound comprising DNA-dependent DNA polymerase inhibitor activity, whereby a complex comprising a first primer extension product and the target RNA is produced, whereby the first primer extension product comprises a sequence (Y) that is complementary to the target RNA and comprises sequence (A);

(c) disabling or removing at least one compound comprising DNA-dependent DNA polymerase inhibitor activity;

(d) cleaving the target RNA in the complex of step (b).

(e) extending a second primer along the first primer extension product with at least one enzyme comprising DNA-dependent DNA polymerase activity, wherein the second primer comprises sequence (C) that is complementary to a sequence (C') on the first primer extension product to produce a complex comprising the first primer extension product and a second primer extension product, whereby the second primer extension product comprises the sequence (C), a sequence (Y') complementary to sequence (Y) and sequence (A') complementary to sequence (A).

IDT Exhibit 1020.2290

(f) producing multiple copies of a polynucleotide sequence complementary to the target RNA using a primer with a sequence (A) and/or sequence (A').

3. The method of claims 1 or 2, wherein the first primer comprises a 5' sequence that is not hybridizable to the target RNA.

4. The method of claims 1or 2, wherein the method further comprises clonally producing multiple copies of a polynucleotide sequence complementary to the target RNA using a primer with a sequence (A) and/or sequence (A') on a solid substrate or in an emulsion.

5. The method of claim 3, wherein the sequence (A') is used as a priming site for amplification.

6. The method of claim 3, wherein the producing of multiple copies in step (f) are produced using amplification methods comprising polymerase chain reaction (PCR).

7. The method of claim 6, wherein a PCR primer comprises sequence (A) of the first primer.

8. The method of claim 7, wherein a PCR primer comprises sequence (C) of the second primer.

9. The method of claim 7, wherein the producing of multiple copies in step (f) comprises performing PCR with primer pairs wherein the first PCR primer is complementary to sequence (A') and the second PCR primer is complementary to all or a portion of sequence (C') or all or a portion of sequence (Y).

10. The method of claim 6 wherein the producing of multiple copies in step (f) comprises performing single primer, linear PCR using a PCR primer complementary to sequence (A').

11. The method of claim 3 wherein the producing of multiple copies in step (f) comprises an amplification method comprising the use of an RNA/DNA composite primer, RNase H, and a DNA dependent DNA polymerase with strand displacement activity.

12. The method of claims 1or 2, wherein the first primer comprises an RNA and a DNA sequence wherein a 5'-RNA sequence comprises sequence (A) and a 3'-DNA sequence comprises sequence (B); whereby the complex formed in step (e) comprises an RNA/DNA heteroduplex at one end; and wherein the producing of multiple copies in step (f) comprises the steps of (i) cleaving the RNA from the heteroduplex, (ii) hybridizing to sequence (A') an amplification primer comprising a 5' RNA sequence and a 3' DNA sequence, (iii) extending the amplification primer with at least one enzyme comprising DNA dependent DNA polymerase activity and comprising strand displacement activity, (iv) cleaving the RNA from the extended amplification primer in the heteroduplex with at least one enzyme comprising a specificity for cleaving RNA in a DNA-RNA heteroduplex; and (v) repeating steps (ii) through (iv) to

IDT Exhibit 1020.2291

produce multiple copies of amplification product comprising sequence (Y) that is complementary to a portion of the RNA template.

13. The method of claims 1or 2, wherein the producing of multiple copies in step (f) comprises in-vitro-transcription (IVT).

14. The method of claim 13 wherein sequence (A) comprises a pro-promoter, whereby the complex formed in step (e) comprises a double stranded promoter for a DNA dependent RNA polymerase comprising sequences (A) and (A'), wherein (A) is a DNA sequence.

15. The method of claim 2, wherein the second primer comprises a fragment from the cleaved RNA target.

16. The method of claims 1 or 2, wherein the DNA-dependent DNA polymerase inhibitor activity is derived from a compound comprising at least one of the following: actinomycin, alpha-amanitin, aphidicolin, BP5, novobiocin, rifampicin, rifamycin, sulfoquinovosylmonoacylglycerol, sulfoquinovosyldiacylglycerol, ursane, oleanane triterpenoids, ursolic acid, oleanolic acid, mikanolide, dihydromikanolide, dehydroaltenusin, catapol, taxinine, cephalomanninine, dipeptide alcohols, corylifolin; bakuchiol; resveratrol; Neobavaisoflavone; daidzein; bakuchicin, levodopa, dopamine, anacardic acid and oleic acid.

17. The method of claims 1 or 2, wherein the DNA-dependent DNA polymerase inhibitor activity is derived from a compound comprising actinomycin.

18. The method of claims 1 or 2, wherein the sample comprises total nucleic acid in a biological fluid.

19. The method of claim 18, wherein the biological fluid is selected from the group consisting of: plasma, serum, urine, saliva.

20. The method of claims 1 or 2, wherein the sample comprises viral RNA.

21. The method of claims 1 or 2, wherein the sample comprises cell or tissue lysates.

22. The method of claims 1 or 2, wherein the second primer comprises a tailed primer with a 5'-sequence which does not hybridize to the first primer extension product and which comprises a propromoter sequence such that RNA transcripts are produced comprising a sequence homologous to the target RNA; whereby multiple copies of the homologous sequence of the RNA sequence of interest are generated.

IDT Exhibit 1020.2292

23. The method of claims 1 or 2, wherein the producing of multiple copies in step (f) comprises hybridizing a PTO, comprising a propromoter sequence and a 3'-hybridizing region, to the 3'-end of a first primer extension product under conditions which allow transcription to occur by RNA polymerase, such that RNA transcripts are produced comprising sequences homologous to the target RNA; whereby multiple copies of the RNA sequence of interest are generated.

24. The method of claims 1 or 2, wherein the producing of multiple copies in step (f) comprises hybridizing a PTO, comprising a propromoter sequence and a 3'-hybridizing region, to the 3'-end of a second primer extension under conditions which allow transcription to occur by RNA polymerase, such that RNA transcripts are produced comprising sequences complementary to the target RNA; whereby multiple copies of the complementary sequence of the RNA sequence of interest are generated.

25. The method of claims 1 or 2, wherein the target RNA is mRNA.

26. The method of claims 1 or 2, wherein the first primer comprises a sequence complementary to poly-A on RNA.

27. The method of claims 1 or 2, wherein the sequence (B) which hybridizes to the target RNA comprises a random sequence.

28. The method of claims 1 or 2, wherein the first primer comprises a mixture of first primers comprising a first primer comprising a sequence complementary to poly-A on RNA, and further comprising a first primer comprising a sequence which hybridizes to the target RNA and comprises a random sequence.

29. The method of claim 15, wherein said fragment generated by cleaving RNA in the complex of step (b) with an enzyme that cleaves RNA from an RNA/DNA heteroduplex, by heat, or by chemical means.

30. The method of claims 1 or 2, wherein said method comprises generating multiple copies of two or more different sequences of interest.

31. The method of claim 30, wherein said method comprises at least two different first primers.

32. The method of claim 2, wherein cleaving the RNA in a heteroduplex of the RNA target and the first primer extension product is performed by RNase H.

33. A method of sequencing an RNA sequence of interest, said method comprising analyzing amplification products to determine sequence, said amplification products produced by the method of any of claims 1 or 2 in the presence of a mixture of NTPs and NTP analogs such that transcription is terminated upon incorporation of an NTP analog

IDT Exhibit 1020.2293

34. The method of claim 33, wherein the target RNA is mRNA, a whole transcriptome or substantial fraction thereof, or a set of whole transcriptomes or substantial fractions thereof.

35. A method of detecting a mutation in a target RNA, comprising analyzing sequences of amplification products for the presence of a mutation as compared to a reference polynucleotide sequence, said amplification products produced by the method of any of claims 1 or 2.

36. The method of claim 35, wherein the target RNA is mRNA.

37. A method of producing a nucleic acid immobilized to a substrate comprising immobilizing amplification products on a substrate, said amplification products produced by the method of any of claims 1 or 2.

38. The method of claim 37, wherein the target RNA is mRNA.

39. The method of claim 37, wherein the substrate is a microarray.

40. A method for performing expression analysis of one or more target RNA sequences comprising:

(a) collecting a sample of nucleic acid comprising RNA and DNA;

(b) optionally enriching the nucleic acid in the sample;

(c) contacting the nucleic acid with an inhibitor of DNA-dependent DNA polymerase activity;

(d) hybridizing to the target RNA a first primer comprising a sequence (A) that is not complementary to the target RNA, and a sequence (B) at the 3'-end which hybridizes to the target RNA;

(e) extending the first primer with at least one enzyme comprising RNA-dependent DNA polymerase activity in the presence of at least one compound comprising DNA-dependent DNA polymerase inhibitor activity, whereby a complex comprising a primer extension product and the target RNA is produced, whereby the first primer extension product comprises a sequence (Y) that is complementary to the target RNA and comprises sequence (A);

(f) disabling or removing at least one compound comprising DNA-dependent DNA polymerase inhibitor activity;

(g) producing multiple copies of a polynucleotide sequence complementary to the target RNA and/or complementary to sequence (Y) using sequence (A); and

(h) further analyzing the amplified products of step (g).

41. The method of claim 40, wherein step (h) further analyzing the products of step (g) includes whole transcriptome analysis, whole transcriptome profiling, microarray analysis, quantitative PCR, pyrosequencing, sequencing by ligation, dye-terminator sequencing, whole transcriptome sequencing, RNA sequencing by massively parallel sequencing, or digital PCR.

IDT Exhibit 1020.2294

42. The method of claim 40, wherein the DNA-dependent DNA polymerase inhibitor activity is derived from a compound comprising at least one of the following: actinomycin, a derivative of actinomycin, alpha-amanitin, aphidicolin, BP5, novobiocin, rifampicin, rifamycin, sulfoquinovosylmonoacylglycerol, sulfoquinovosyldiacylglycerol, ursane, oleanane triterpenoids, ursolic acid, oleanolic acid, mikanolide, dihydromikanolide, dehydroaltenusin, catapol, taxinine, cephalomanninine, dipeptide alcohols, corylifolin; bakuchiol; resveratrol; Neobavaisoflavone; daidzein; bakuchicin, levodopa, dopamine, anacardic acid and oleic acid.

43. The method of claim 40, wherein step (g) comprises PCR, linear PCR, in vitro transcription, or single primer isothermal amplification.

44. A method of generating at least one double stranded DNA product corresponding to a sequence of a target RNA or a portion thereof comprising:

(a) adding to a reaction mixture comprising target RNA and further comprising DNA:

(i) at least one first primer comprising a 3' annealing sequence and a 5' tail sequence;

(ii) a DNA-dependent DNA polymerase inhibitor; and

(iii) an RNA-dependent DNA polymerase;

(b) disabling or removing the at least one DNA-dependent DNA polymerase inhibitor from the reaction mixture;

(c) annealing and extending at least one all DNA second primer comprising an annealing sequence with a DNA-dependent DNA polymerase, wherein said double stranded product is labeled at least one end; and

(d) clonally amplifying one or both strands of the double-stranded product.

45. The method of claim 44, wherein the clonal amplification is performed on a solid substrate, in an emulsion, or both.

46. The method of claim 44, wherein the clonal amplification generates polonies.

47. A kit for amplifying a target RNA comprising: a tailed first primer, a reverse transcriptase enzyme, a DNA-dependent DNA polymerase inhibitor, and optionally a second primer.

48. The kit of claim 47 wherein the first tailed primer comprises a 3' random portion, and the second primer comprises a random primer.

49. The kit of claim 47 further comprising a composite amplification primer, and wherein the tailed first primer is a composite primer comprising a 3' DNA portion and a 5' RNA portion.

50. The kit of claim 47, wherein the kit further comprises instructions for the use of the kit.

IDT Exhibit 1020.2295

51. The kit of claim 47, wherein the kit further comprises a DNA-dependent DNA polymerase.

52. The kit of claim 47, wherein the kit further comprises reagents for DNA purification.

IDT Exhibit 1020.2296

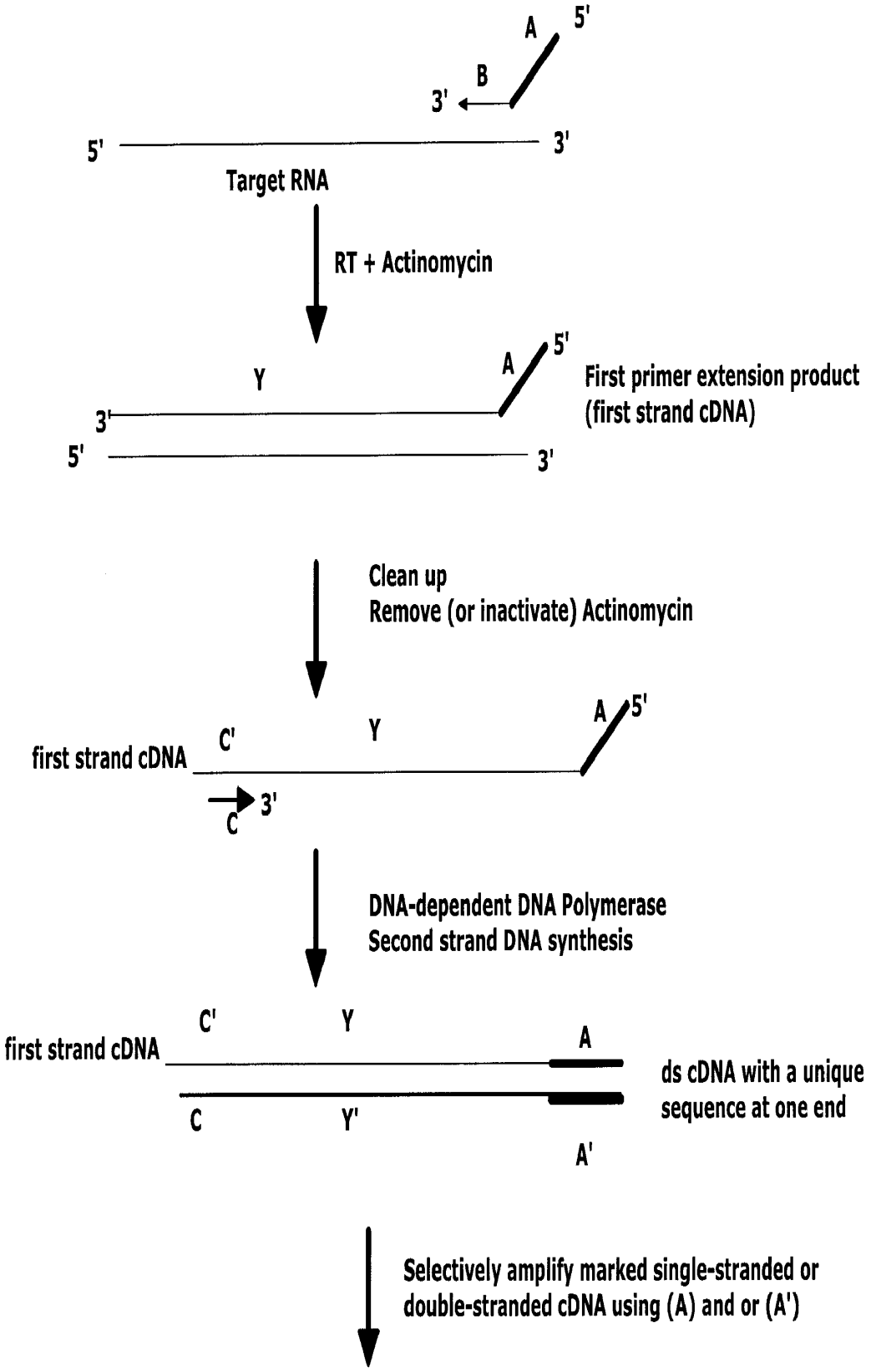

A
5'
B
3'
5'
3'
Target RNA
RT + Actinomycin
Y
A
5'
3'
5'
3'
First primer extension product
(first strand cDNA)
Clean up
Remove (or inactivate) Actinomycin
first strand cDNA
C'
Y
A
5'
C
3'
DNA-dependent DNA Polymerase
Second strand DNA synthesis
first strand cDNA
C'
Y
A
ds cDNA with a unique
sequence at one end
C
Y'
A'
Selectively amplify marked single-stranded or
double-stranded cDNA using (A) and or (A')

Figure 1

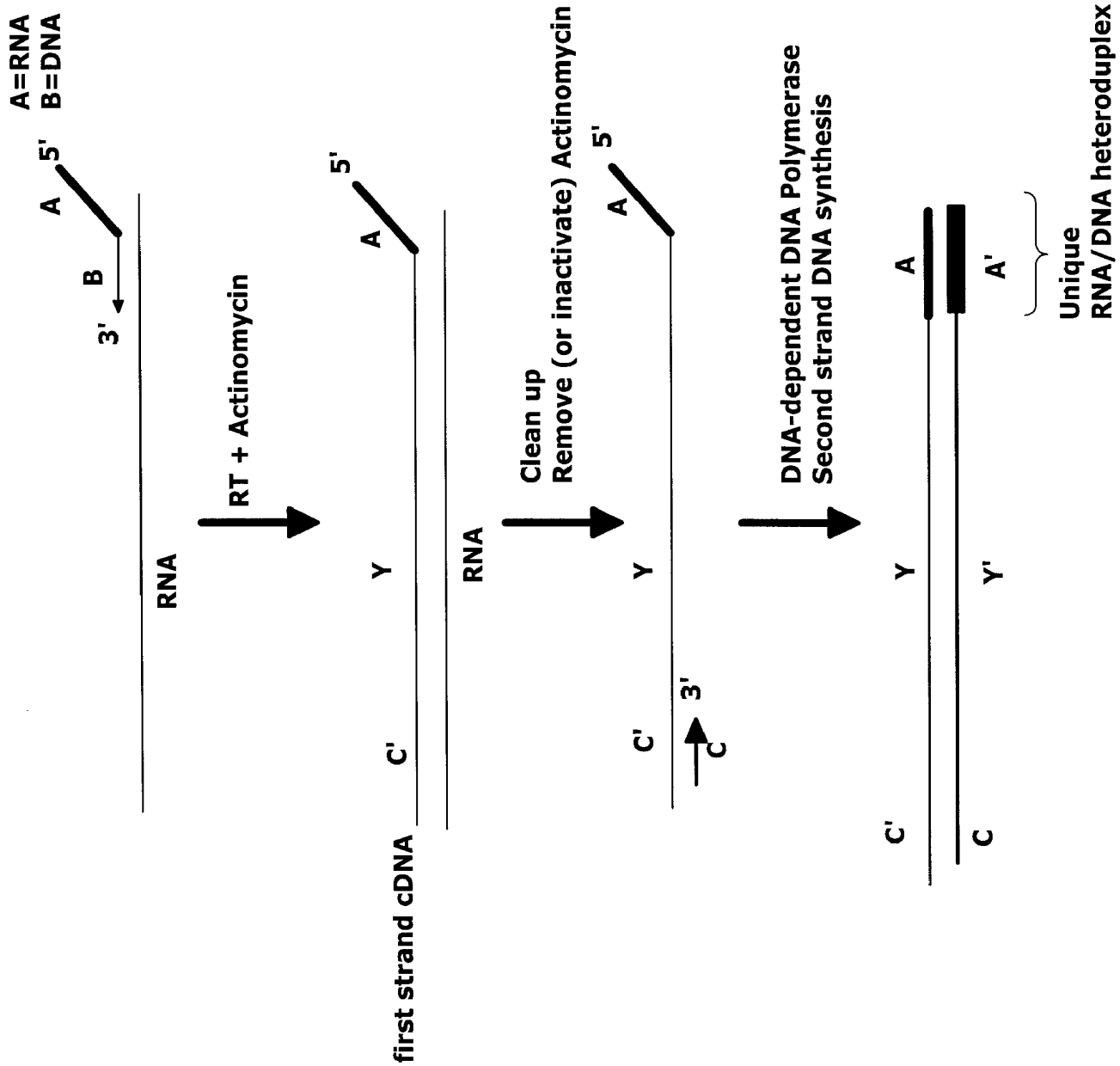

A=RNA
B=DNA
5'
A
B
3'
RNA
RT + Actinomycin
5'
A
Y
C'
first strand cDNA
RNA
Clean up
Remove (or inactivate) Actinomycin
5'
A
Y
C'
3'
C
DNA-dependent DNA Polymerase
Second strand DNA synthesis
A
A'
Y
Y'
C'
C
Unique
RNA/DNA heteroduplex

Figure 2a

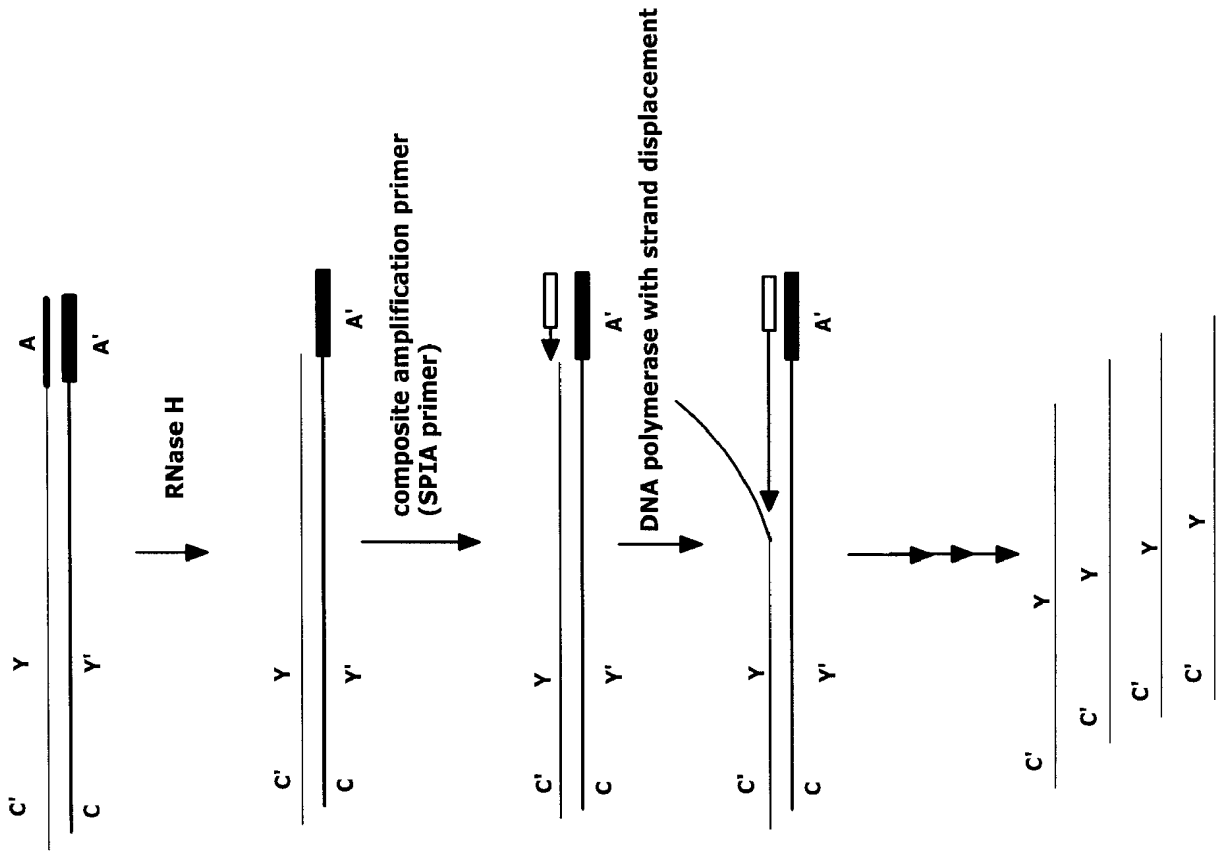

A
A'
Y
Y'
C'
C
RNase H
composite amplification primer
(SPIA primer)
DNA polymerase with strand displacement

Figure 2b

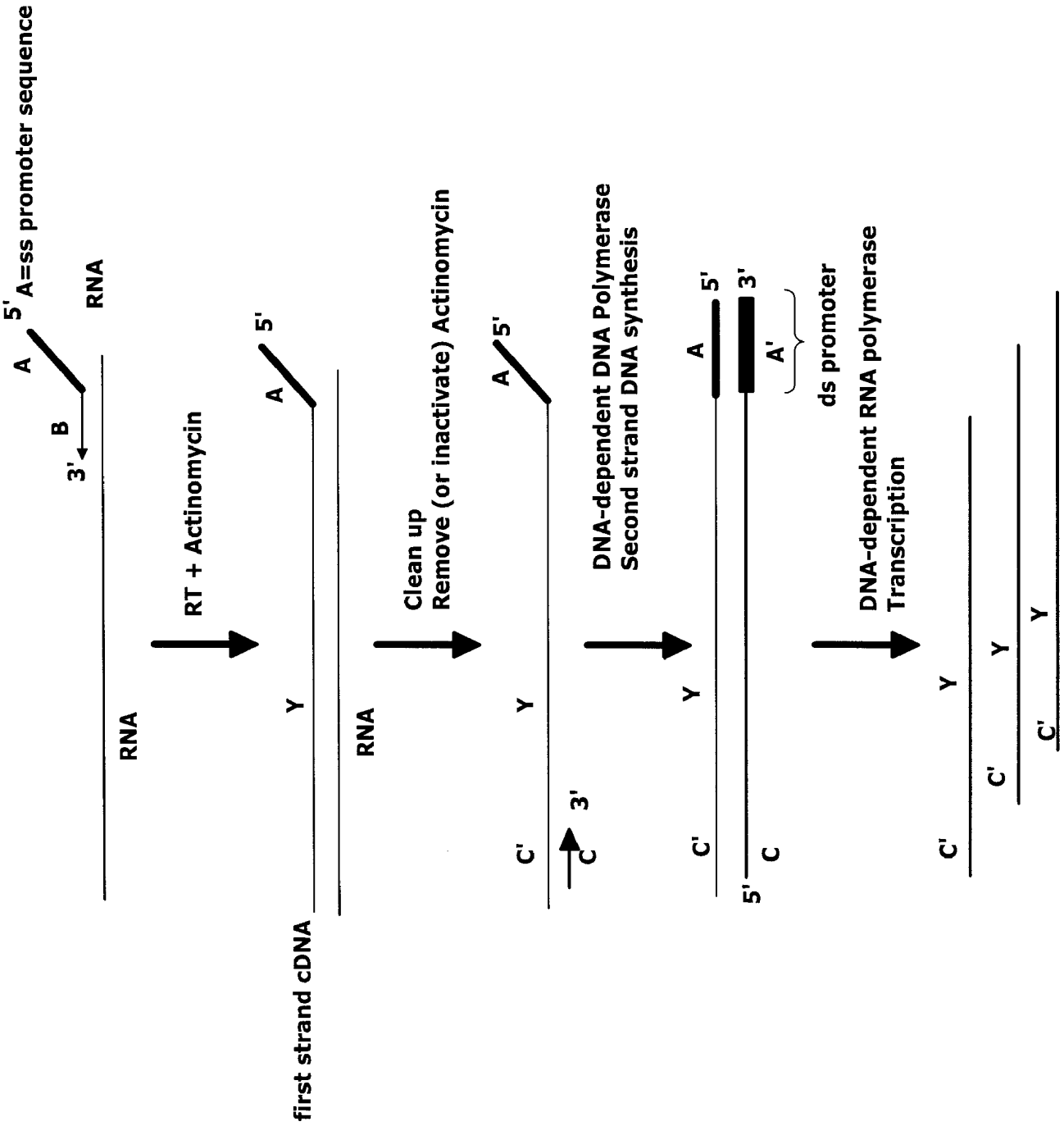

5' A=ss promoter sequence
A
B
3'
RNA
RT + Actinomycin
5'
A
Y
RNA
first strand cDNA
Clean up
Remove (or inactivate) Actinomycin
5'
A
Y
C'
3'
C
DNA-dependent DNA Polymerase
Second strand DNA synthesis
A
5'
Y
C'
3'
A'
5'
C
ds promoter
DNA-dependent RNA polymerase
Transcription
C'
Y
C'
Y
C'
Y

Figure 3

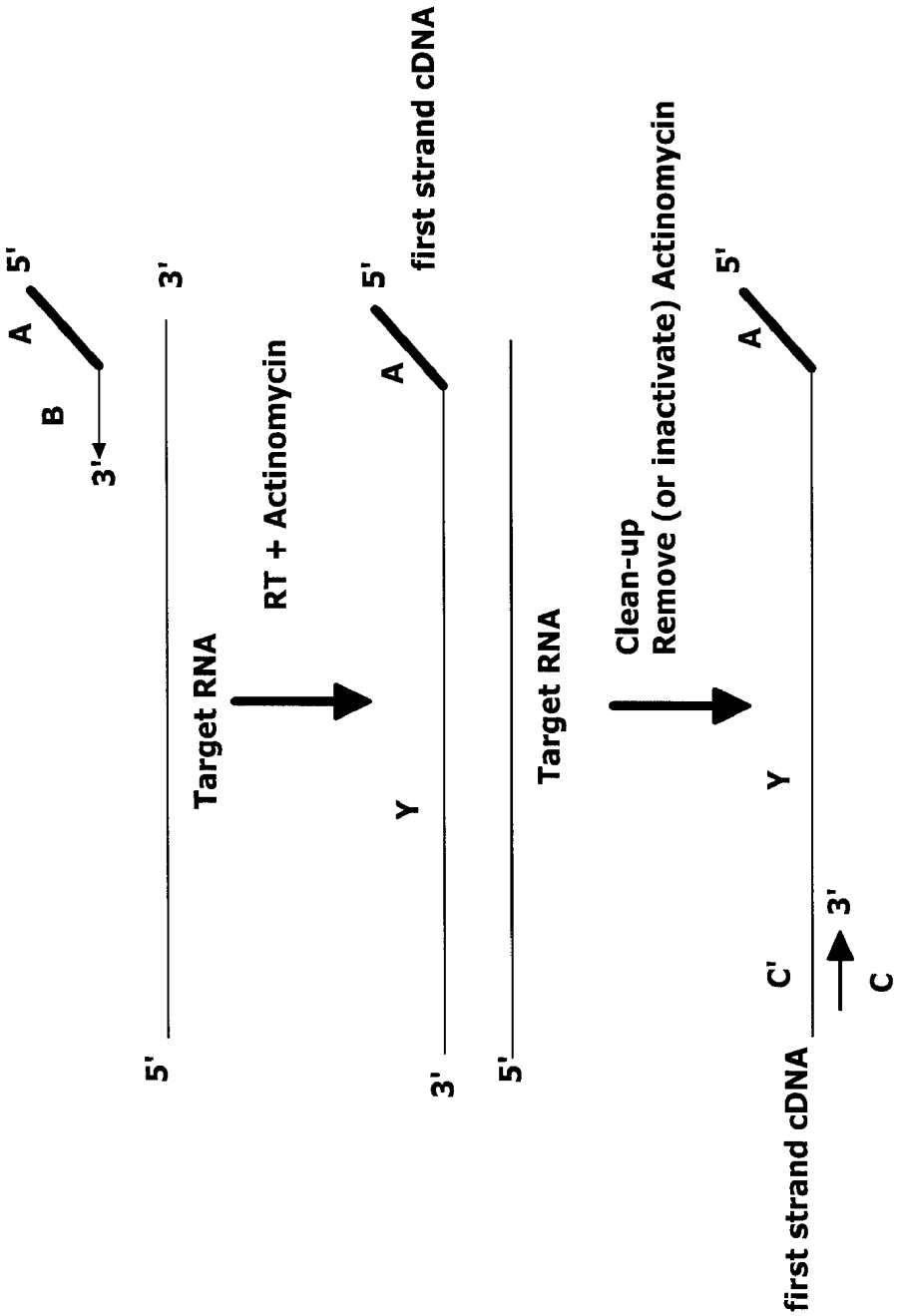

5'
A
B
3'
3'
Target RNA
5'
RT + Actinomycin
5'
A
Y
3'
first strand cDNA
Target RNA
5'
Clean-up
Remove (or inactivate) Actinomycin
5'
A
Y
C'
3'
C
first strand cDNA

Figure 4 a

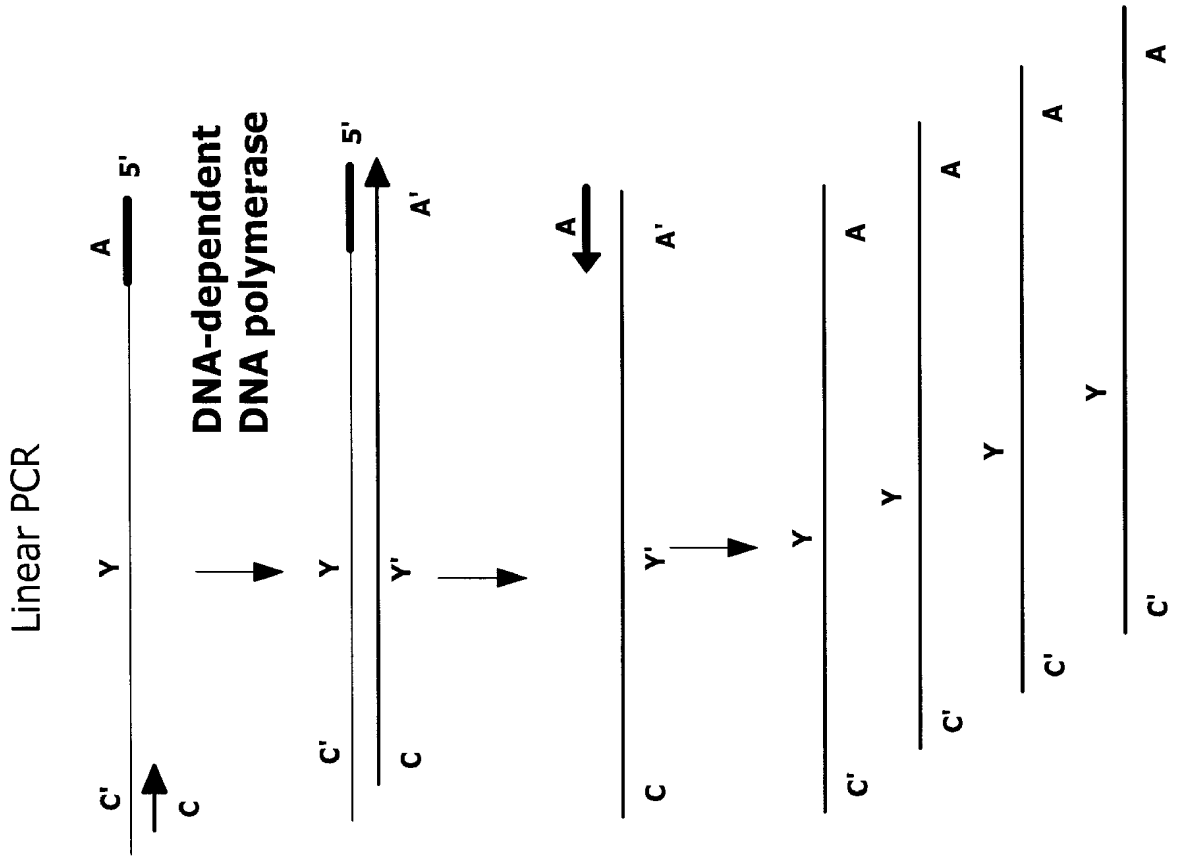


Figure 4b

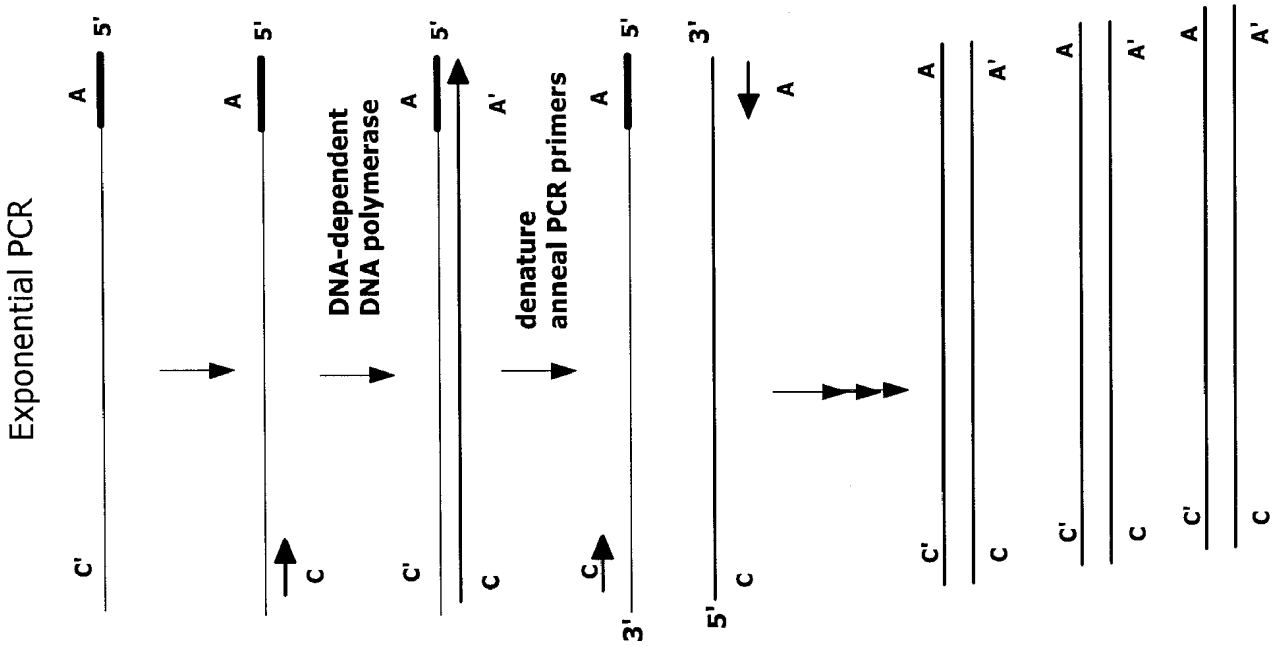

Exponential PCR
DNA-dependent
DNA polymerase
denature
anneal PCR primers
A
A'
C
C'
5'
3'

Figure 4c

Espacenet

Bibliographic data: WO2014144092 (A1) — 2014-09-18

SEQUENTIAL SEQUENCING

Inventor(s): AMORESE DOUG [US]; SCHROEDER BENJAMIN G [US]; SCOLNICK JONATHAN [US] ± (AMORESE, DOUG, ; SCHROEDER, BENJAMIN, G, ; SCOLNICK, JONATHAN)

Applicant(s): NUGEN TECHNOLOGIES INC [US] ± (NUGEN TECHNOLOGIES, INC)

Classification: **- international:** ***C12Q1/68***
- cooperative: **C12Q1/6874; G06F19/22**

Application number: WO2014US28356 20140314 Global Dossier

Priority number (s): US201361801600P 20130315

Also published as: EP2971130 (A1) EP2971130 (A4) US2014274738 (A1) US2016251711 (A1) US2016251712 (A1) more

Abstract of WO2014144092 (A1)

Described herein are improved methods, compositions and kits for next generation sequencing (NGS). The methods, compositions and kits described herein enable phasing of two or more nucleic acid sequences in a sample, i.e. determining whether the nucleic acid sequences (which can comprise regions of sequence variation) are located on the same chromosome and/or the same chromosomal fragment. Phasing information can be obtained by performing multiple, successive sequencing reactions from the same immobilized nucleic acid template. The methods, compositions and kits provided herein can be useful, for example, for haplotyping, SNP phasing, or for determining downstream exons in RNA-seq.

IDT Exhibit 1020.2304

```
ERROR: rangecheck
OFFENDING COMMAND: getinterval

STACK:

-10
0
(

)
0
(

)
1024
```

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 September 2014 (18.09.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/144092 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2014/028356

(22) International Filing Date:
14 March 2014 (14.03.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/801,600 15 March 2013 (15.03.2013) US

(71) Applicant: **NUGEN TECHNOLOGIES, INC.** [US/US]; 201 Industrial Road, Suite 310, San Carlos, CA 94070 (US).

(72) Inventors: **AMORESE, Doug**; 357 South El Monte Avenue, Los Altos, CA 94022 (US). **SCHROEDER, Benjamin, G.**; 1964 Shoreview Avenue, San Mateo, CA 94401 (US). **SCOLNICK, Jonathan**; 451 Kansas Street, #518, San Francisco, CA 94107 (US).

(74) Agents: **HSU, Jennifer** et al.; Wilson Sonsini Goodrich & Rosati, 650 Page Mill Road, Palo Alto, CA 94304-1050 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*

[Continued on next page]

(54) Title: SEQUENTIAL SEQUENCING

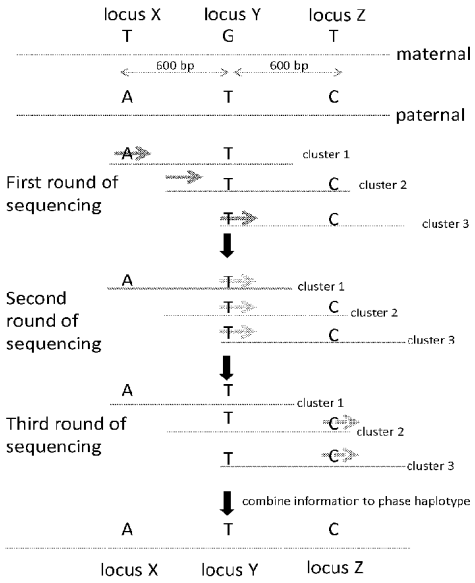


FIG. 3

(57) Abstract: Described herein are improved methods, compositions and kits for next generation sequencing (NGS). The methods, compositions and kits described herein enable phasing of two or more nucleic acid sequences in a sample, i.e. determining whether the nucleic acid sequences (which can comprise regions of sequence variation) are located on the same chromosome and/or the same chromosomal fragment. Phasing information can be obtained by performing multiple, successive sequencing reactions from the same immobilized nucleic acid template. The methods, compositions and kits provided herein can be useful, for example, for haplotyping, SNP phasing, or for determining downstream exons in RNA-seq.

WO 2014/144092 A1

IDT Exhibit 1020.2307

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

SEQUENTIAL SEQUENCING

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application No. 61/801,600, filed March 15, 2013, which application is incorporated herein by reference.

BACKGROUND

[0002] Next generation sequencing (NGS) can be utilized for a wide variety of life science applications, such as, e.g., diagnostics, epidemiology, and forensics. Widely used NGS technologies commonly involve sequencing of short reads (e.g., 100 nt reads) which can then be mapped to a reference genome. While such technologies can be useful for the detection of sequence variants (e.g., mutations), they are also associated with certain limitations. For example, NGS can be used to sequence genomes or genomic regions of diploid organisms. A genome of a diploid organism can comprise sets of chromosome pairs, wherein each chromosome pair comprises a maternal chromosome and a paternal chromosome. For example, a diploid organism can inherent a maternal allele and a paternal allele at a single locus. Thus, alleles detected at a plurality of loci which map to a particular chromosome might exist within a single chromosome (maternal or paternal) of a chromosome pair, or across both chromosomes of a chromosome pair. While many NGS platforms are useful for detecting alleles at a plurality of loci, they currently do not provide phasing and/or haplotype information e.g., to distinguish whether alleles detected at a plurality of loci that map to a particular chromosome are co-located on the same chromosome or are located on separate chromosomes in a chromosome pair. Determining whether a plurality of alleles are co-located on the same (maternal or paternal) chromosome or are located on different chromosomes can be useful for a variety of reasons, as discussed further below.

[0003] The pattern of alleles within each individual chromosome can be referred to as haplotype. Haplotyping can have many diagnostic and clinical applications. For example, two inactivating mutations across different loci within a single gene might be of little or no consequence if present on the same individual chromosome (i.e. chromosome of either maternal or paternal origin), because the other copy of the gene product will remain functional. On the other hand, if one of the inactivating mutations is present in the maternal chromosome and the other in the paternal chromosome, there can be no functional copy of the gene product, which can result in a negative phenotype (non-viability, increased risk for disease and others). Haplotyping can also be used to predict risk or susceptibility to specific

IDT Exhibit 1020.2309

genetic diseases, as many genetic associations are tied to haplotypes. For example, the various haplotypes of the human leukocyte antigen (HLA) system can be associated with genetic diseases ranging from autoimmune disease to cancers.

[0004] Another instance in which phasing information can be useful is distinguishing between functional genes and their non-functional pseudogene counterparts within the genome. One well known functional gene/pseudogene pair is the genes SMN1 and SMN2, which differ in sequence by only five nucleotides over many Kb of sequence, yet one of the nucleotide differences renders the SMN2 gene almost completely non-functional. Using short read sequencing, a mutation may be found in one of the two genes, but unless the mutation happens to occur within the sequencing read that also covers one of the known nucleotide differences between SMN1 and SMN2, it will be difficult to know which of the genes (the functional gene, or the nonfunctional pseudogene) is mutated.

[0005] The present NGS methods can employ short read sequencing to query regions of variable DNA sequence (polymorphisms etc.) interspersed within regions of conserved DNA sequence. As significant blocks of conserved sequence can be interspersed between the variable regions, short read sequencing does not lend itself to phasing analysis. Although methods have been developed to obtain phasing information, these methods (for example, Sanger sequencing and subcloning), can be labor intensive and/or costly.

[0006] There is a need for improved NGS methods that provide phasing information. Such methods can provide a highly parallel platform for performing multiple sequencing reactions from the same immobilized templates. Methods described herein fulfill this need.

SUMMARY OF THE INVENTION

[0007] Provided herein are novel methods, compositions and kits for phasing two or more nucleic acid sequences in a sample. Specifically, in one aspect methods and compositions are provided that can allow for determining whether two or more nucleic acid sequences (typically comprising regions of sequence variation) are located on the same nucleic acid template, chromosome, or chromosomal fragment. The methods and compositions disclosed herein can also be used to distinguish and differentiate between two closely related nucleic acid sequences by compiling and aligning data from sequential sequencing reads.

[0008] The methods, kits, and compositions described herein can employ sequential paired sequencing reads from the same immobilized nucleic acid template. The reads can be generated by successive rounds of priming, sequencing, denaturing and repriming, and the

IDT Exhibit 1020.2310

results from multiple reads originating from the same template can be compiled to obtain phasing information.

[0009] Additionally, the methods, kits, and compositions described herein can employ pools of oligonucleotides which may be used to prime sequencing reactions. Such sequencing reactions can target specific regions of specific nucleic acids for sequencing. These oligonucleotide pools can be used onboard a sequencer to extend the sequencing of nucleic acids that have already undergone a first round of sequencing.

[0010] In one aspect, provided herein is a method for relating multiple nucleic acid sequences (e.g., comprising regions of sequence variation) to the same nucleic acid template. In some embodiments, the method comprises: a) creating a nucleic acid library; b) sequencing the library with an oligonucleotide primer; c) denaturing the extension strand; d) performing a second round of sequencing by introducing a new oligonucleotide primer containing sequence complementary to conserved regions present in some of the nucleic acid templates within the nucleic acid library; e) repeating steps c) and d) as needed; and f) compiling sequencing data from the successive sequencing reads to differentiate between closely related nucleic acid sequences. In some embodiments, the nucleic acid library is a directional nucleic acid library. In some embodiments, the nucleic acid library is not a directional nucleic acid library.

[0011] In some embodiments, the nucleic acid library comprises closely related nucleic acid sequences as inserts. In some embodiments, the conserved regions within the nucleic acid inserts are located adjacent to variable regions. In some embodiments, alignment of multiple variable regions enables differentiating between and/or typing of related transcripts. In some embodiments, alignment of multiple variable regions enables differentiating between and/or typing of related micro-organisms. In some embodiments, the nucleic acid library is a directional nucleic acid library. In some embodiments, the nucleic acid library is not a directional nucleic acid library.

[0012] In another aspect, provided herein is a method for differentiating between closely related nucleic acid sequences (such as genes and pseudogenes) by using specific sets of oligonucleotide primers containing sequence complementary to a common region shared by the closely related sequences. In some embodiments, the method comprises: a) creating a sequencing library with related nucleic acid sequences as inserts; b) sequencing the library with an oligonucleotide primer; c) denaturing the extension strand; d) performing a second round of sequencing by introducing a new oligonucleotide primer containing sequence complementary to conserved regions present in some of the nucleic acid templates within the

IDT Exhibit 1020.2311

nucleic acid library; e) repeating steps c) and d) as needed; and f) compiling sequencing data from the successive sequencing reads to differentiate between closely related nucleic acid sequences.

[0013] Kits for performing any of the methods described herein are provided. Such kits can include reagents, enzymes and platforms for amplification and sequencing of nucleic acids. In one embodiment, a kit is provided comprising: a) an adaptor or several adaptors, b) one or more of oligonucleotide primers, and c) reagents for amplification. In another embodiment, the kit further comprises reagents for sequencing. A kit can include instructions for employing the kit components as well as the use of any other reagent not included in the kit.

[0014] Also provided herein is a method for relating at least two nucleic acid sequences or regions of sequence variation to the same nucleic acid template, the method comprising: (a) creating a nucleic acid library wherein a template is immobilized on a solid support; (b) generating a first sequencing read from the immobilized template with a first oligonucleotide primer, wherein a 3' terminal nucleotide of the first oligonucleotide primer binds to a first binding site in the template; (c) denaturing the first sequencing read from the immobilized template; (d) annealing a second oligonucleotide primer that is complementary to a region or regions within the immobilized template, wherein a 3' terminal nucleotide of the second oligonucleotide primer binds to a second binding site in the template that is more than 5 nt away from the first binding site in the template; (e) generating a second sequencing read with the second oligonucleotide primer; and (f) compiling data from the first sequencing read and the second sequencing read to map reads originating from the template. In some embodiments, the nucleic acid libraries are generated from amplicons. In some embodiments, the amplicons comprise conserved regions which flank a variable region. In some embodiments, alignment of multiple variable regions enables differentiation and/or typing of related transcripts. In some embodiments, alignment of multiple variable regions enables differentiation and/or typing of related micro-organisms. In some embodiments, the libraries are reduced complexity. In some embodiments, the reduced complexity is achieved by target enrichment. In some embodiments, the nucleic acid library comprises closely related nucleic acid sequences as inserts. In some embodiments, the nucleic acid library is a directional nucleic acid library. In some embodiments, the nucleic acid library is not a directional nucleic acid library.

[0015] Also provided herein is a method comprising: (a) coupling at least one nucleic acid template to a sequencing platform; (b) subjecting the at least one nucleic acid template to a first round of sequencing using a first sequencing primer, wherein the first sequencing primer

IDT Exhibit 1020.2312

comprises a 3' terminal nucleotide which anneals to a first binding site in the at least one nucleic acid template, wherein the first round of sequencing produces a first sequencing read of said at least one nucleic acid template; (c) removing said first annealed sequencing primer and said first sequencing read from said at least one nucleic acid template; and (d) subjecting said at least one nucleic acid template to a second round of sequencing using a second sequencing primer, wherein the second sequencing primer comprises a 3' terminal nucleotide which anneals to a second binding site in the at least one nucleic acid template which is more than 5 nt away from the first binding site, wherein the second round of sequencing produces a second sequencing read of the at least one nucleic acid template.

[0016] In some embodiments of any of the foregoing methods, the first and/or the second round of sequencing comprises sequencing by synthesis. In some embodiments, the first and/or the second round of sequencing comprises sequencing by ligation. In some embodiments, the coupling comprises conjugating the at least one nucleic acid template to a solid support of the sequencing platform. In some embodiments, the solid support comprises a bead, a well, an array, or a flow cell. In some embodiments, the at least one nucleic acid template comprises an adaptor oligonucleotide ligated to one or both ends of the at least one nucleic acid template. In some embodiments, the first sequencing primer is selectively hybridizable to a primer binding site of the adaptor oligonucleotide. In some embodiments, the second sequencing primer is not selectively hybridizable to the primer binding site of the adaptor oligonucleotide. In some embodiments, the second sequencing read comprises an informative locus. In some embodiments, the informative locus is a polymorphic locus. In some embodiments, the first sequencing read comprises a first informative locus and the second sequencing read comprises a second informative locus. In some embodiments, the method comprises determining that a first allele detected at the first informative locus and a second allele detected at the second informative locus are in phase. In some embodiments, the second sequencing primer is designed to hybridize to a location that is within x nt upstream of the informative locus, wherein x is a number that is less than an average sequencing read length produced by the first or second round of sequencing. In some embodiments, the removing comprises a denaturing step. In some embodiments, the removing further comprises a washing step. In some embodiments, the method further comprises subjecting the at least one nucleic acid template to one or more additional rounds of sequencing, wherein the one or more additional rounds of sequencing comprise the steps of : (i) removing a previously annealed sequencing primer and previously generated sequencing read from the at least one nucleic acid template; (ii) annealing an additional

IDT Exhibit 1020.2313

sequencing primer to the at least one nucleic acid template; and (iii) subjecting the at least one nucleic acid template to an additional round of sequencing using the additional sequencing primer, thereby producing an additional sequencing read of the at least one nucleic acid template. In some embodiments, the additional sequencing primer comprises a 3' terminal nucleotide which hybridizes to a binding site that is more than 5 nt away from a binding site of the 3' terminal nucleotide of the previously annealed sequencing primer. In some embodiments, a first allele detected by the first sequencing read, a second allele detected by the second sequencing read, and an additional allele detected by the additional sequencing read are determined to be in phase. In some embodiments, the additional sequencing primer and the previously annealed sequencing primer are designed to hybridize to genomic locations that map to a single chromosome. In some embodiments, the genomic locations that map to a single chromosome span a distance that is less than a length of the at least one nucleic acid template. In some embodiments, the genomic locations that map to a single chromosome span a distance that is less than three-quarters of the length of the at least one nucleic acid template. In some embodiments, the genomic locations that map to a single chromosome span a distance that is less than two-thirds of the length of the at least one nucleic acid template. In some embodiments, the genomic locations that map to a single chromosome span a distance that is less than one half of the length of the at least one nucleic acid template. In some embodiments, the genomic locations that map to a single chromosome span a distance that is about a sequencing read length of the sequencing platform. In some embodiments, the at least one nucleic acid template is a member of a sequencing library. In some embodiments, the sequencing library is a whole-genome sequencing library. In some embodiments, the sequencing library is a target-enriched sequencing library. In some embodiments, the sequencing library is a directional sequencing library. In some embodiments, the sequencing library is not a directional sequencing library. In some embodiments, the sequencing library comprises nucleic acid fragments of an average length. In some embodiments, the average length is about 50 bases to about 10 kb. In some embodiments, the first, second, and optionally additional sequencing reads comprise sequence information from a plurality of loci. In some embodiments, phasing of alleles detected at the plurality of loci is determined based upon the first, second, and optionally additional sequencing reads. In some embodiments, the at least one nucleic acid template comprises a plurality of nucleic acid templates. In some embodiments, the at least one nucleic acid template comprises at least 10,000 nucleic acid templates. In some embodiments, the at least one nucleic acid template comprises 3 billion nucleic acid

IDT Exhibit 1020.2314

templates. In some embodiments, the at least one nucleic acid template is coupled to a unique addressable location of the sequencing platform. In some embodiments, the method comprises compiling sequencing reads from the unique addressable location into one data file. In some embodiments, the method comprises determining phasing of a plurality of alleles detected across two or more unique addressable locations. In some embodiments, the two or more unique addressable locations each comprise two or more detected alleles, wherein the two or more detected alleles of each of the two or more unique addressable locations are determined to be in phase if at least one common detected allele is shared by the two or more unique addressable locations. In some embodiments, step (d) of a method disclosed herein comprises subjecting the at least one nucleic acid template to a second round of sequencing using a mixture of second sequencing primers. In some embodiments, the mixture of second sequencing primers comprises a pair of primers, wherein each primer in the primer pair is designed to sequence a forward or reverse complement strand containing a locus. In some embodiments, primers within the mixture are designed and/or selected to have minimal potential to hybridize to the same nucleic acid template.

[0017] In some embodiments of any of the aforementioned methods, the template comprises DNA. In some embodiments, the template comprises RNA. In some embodiments, one or more methods disclosed herein is used to provide phasing information for two or more alleles separated by a distance of over 150 bases, over 300 bases, over 500 bases, over 600 bases, over 700 bases, over 800 bases, over 900 bases, over 1000 bases (1kb), over 5 kb, over 10 kb, over 50 kb, over 100 kb, over 500 kb, over 1000 kb (1 Mb), over 5 Mb, over 10 Mb, over 50 Mb, over 100 Mb, over 150 Mb, over 200 Mb, over 250 Mb, or over 300 Mb.

[0018] Also provided herein is a kit, comprising: (a) one or more oligonucleotide primers for performing a sequencing reaction; and (b) instructions for practicing a method of any of the preceding claims using the one or more oligonucleotide primers.

[0019] Also provided herein is a computer readable medium comprising computer executable code that, upon execution by a computer processor, implements a method comprising: (a) accessing a data file in computer memory that comprises a plurality of sequencing reads corresponding to a first addressable location of a sequencing platform, wherein the sequencing reads are generating by subjecting a first nucleic acid template coupled to the addressable location to a first and at least a second round of sequencing, wherein the first and at least second rounds of sequencing are performed using a first and second sequencing primer, respectively; (b) using the data file, detecting a first detected allele

IDT Exhibit 1020.2315

at a first locus of the nucleic acid template and detecting a second detected allele at a second locus of the nucleic acid template; and (c) determining the first and second detected allele to be in phase based upon said detecting. In some embodiments, the method further comprises: (d) accessing a second data file in computer memory that comprises a plurality of sequencing reads corresponding to a second addressable location of a sequencing platform, wherein the sequencing reads are generated by subjecting a second nucleic acid template coupled to the addressable location to the first and the at least second round of sequencing; (e) using the second data file, detecting a first detected allele at a first locus of the second nucleic acid template and detecting a second detected allele at a second locus of the second nucleic acid template; and (f) determining the first and second detected alleles of the second nucleic acid template to be in phase with the first and second detected alleles of the first nucleic acid template if the first or second detected allele of the second nucleic acid template is shared in common with the first nucleic acid template.

INCORPORATION BY REFERENCE

[0020] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0021] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0022] **FIGS. 1A** and **1B** illustrate the current limitations of NGS sequencing in obtaining phasing information for a plurality of alleles.

[0023] **FIG. 2** depicts an exemplary workflow of a method described herein.

[0024] **FIG. 3** depicts an exemplary embodiment of a method described herein.

[0025] **FIG. 4** depicts another exemplary embodiment of a method described herein.

[0026] **FIG. 5** depicts an exemplary computer system disclosed herein.

[0027] **FIG. 6** depicts sequential sequencing method as applied to 16S microbial rRNA characterization, as described in Example 1.

IDT Exhibit 1020.2316

[0028] **FIG. 7** depicts the use of specific oligonucleotide pools and the generation of mated pairs of sequencing reads to differentiate between two closely related nucleotide sequences, such as a gene/pseudogene pair.

DETAILED DESCRIPTION OF THE INVENTION

[0029] Throughout this application, various embodiments are presented in a range format. It should be understood that the description in range format is merely for convenience and brevity and should not be construed as an inflexible limitation on the scope of the invention. Accordingly, the description of a range should be considered to have specifically disclosed all the possible subranges as well as individual numerical values within that range. For example, description of a range such as from 1 to 6 should be considered to have specifically disclosed subranges such as from 1 to 3, from 1 to 4, from 1 to 5, from 2 to 4, from 2 to 6, from 3 to 6 etc., as well as individual numbers within that range, for example, 1, 2, 3, 4, 5, and 6. This applies regardless of the breadth of the range.

[0030] Unless otherwise specified, terms and symbols of genetics, molecular biology, biochemistry and nucleic acid used herein follow those of standard treatises and texts in the field, e.g. Kornberg and Baker, *DNA Replication,* Second Edition (W.H. Freeman, New York, 1992); Lehninger, *Biochemistry,* Second Edition (Worth Publishers, New York, 1975); Strachan and Read, Human *Molecular Genetics,* Second Edition (Wiley-Liss, New York, 1999); Eckstein, editor, *Oligonucleotides and Analogs: A Practical Approach* (Oxford University Press, New York, 1991); Gait, editor, *Oligonucleotide Synthesis: A Practical Approach* (IRL Press, Oxford, 1984); Sambrook, Fritsch & Maniatis, *Molecular Cloning: A Laboratory Manual*, Fourth Edition (2012); *Oligonucleotide Synthesis* (M. J. Gait, ed., 1984); *Nucleic Acid Hybridization* (B. D. Hames & S. J. Higgins, eds., 1984); *A Practical Guide to Molecular Cloning* (B. Perbal, 1984); and a series, *Methods in Enzymology* (Academic Press, Inc.) and the like.

Definitions

[0031] As used in the specification and claims, the singular forms "a", "an" and "the" can include plural references unless the context clearly dictates otherwise. For example, the term "a primer" can include a plurality of primers, including mixtures thereof.

[0032] The terms "determining", "measuring", "evaluating", "assessing," "assaying," and "analyzing" can be used interchangeably herein to refer to any form of measurement, and include determining if an element is present or not. These terms can include both quantitative and/or qualitative determinations. Assessing may be relative or absolute. "Assessing the

IDT Exhibit 1020.2317

presence of" can include determining the amount of something present, as well as determining whether it is present or absent.

[0033] As used herein, the term "phasing" can refer to the process of determining whether two or more nucleic acid sequences are located on the same nucleic acid template, the same fragment of a chromosome, or the same chromosome.

[0034] As used herein, the term "haplotype" can refer to a haploid genotype, e.g., a combination or set of nucleic acid sequences detected at a plurality of loci which are typically inherited as a unit and are linked, for example during a translocation event. A haplotype can provide a distinctive genetic pattern of an individual. A haplotype can be determined for one locus, several loci, or an entire chromosome depending on the number of recombination events that occur between a given set of loci. The nucleic acid sequences are not limited to any specific type and can include, by way of example only, conserved genetic sequences (e.g., non-variant) or variant genetic sequences (e.g., polymorphisms). For example, single nucleotide polymorphisms (SNPs), short tandem repeats (STRs), including , for example, di-, tri-, and tetra-nucleotide repeats, copy number variations (CNVs), insertions, deletions, inversions, methylations, etc. can be considered variant genetic sequences. Such markers can be found in various publications and databases including, for example, the ATCC short tandem repeat (STR) database, the Human SNP Database, the NCBI dbSNP database, Celera Human SNP database, SNP Database of the Genome Analysis Group, and the like.

[0035] The term "phased alleles" can refer to the distribution of the particular alleles on a nucleic acid template, a single chromosomal region, and/or a single chromosome. Accordingly, the "phase" of two alleles can refer to a characterization or determination of whether the alleles are located on a single chromosome or two separate chromosomes (e.g., a maternally or paternally inherited chromosomes). Unless otherwise stated, "haplotype" and "phased alleles" can be considered synonymous.

[0036] "Nucleotides" and "nt" can be used interchangeably herein to generally refer to biological molecules that can form nucleic acids. Nucleotides can have moieties that contain not only the known purine and pyrimidine bases, but also other bases that have been modified. Such modifications include acylated purines or pyrimidines, methylated purines or pyrimidines, alkylated riboses, or other heterocycles.

[0037] As used herein, the term "oligonucleotide" can refer to a polynucleotide chain, typically less than 200 residues long, most typically between 15 and 100 nucleotides long, but also intended to encompass longer polynucleotide chains. Oligonucleotides can be single-or double-stranded. The terms "oligonucleotide probe" or "probe", as used herein, can refer to

IDT Exhibit 1020.2318

an oligonucleotide capable of hybridizing to a complementary nucleotide sequence. As used herein, the term "oligonucleotide" can be used interchangeably with the terms "primer", "adaptor" and "probe".

[0038] The term "polynucleotide", "nucleic acid", or grammatical equivalents, can refer to two or more nucleotides covalently linked together. A nucleic acid described herein can contain phosphodiester bonds, although in some cases, as outlined below (for example in the construction of primers and probes such as label probes), nucleic acid analogs are included that can have alternate backbones, comprising, for example, phosphoramide (Beaucage et al., Tetrahedron 49(10):1925 (1993) and references therein; Letsinger, J. Org. Chem. 35:3800 (1970); Sprinzl et al., Eur. J. Biochem. 81:579 (1977); Letsinger et al., Nucl. Acids Res. 14:3487 (1986); Sawai et al, Chem. Lett. 805 (1984), Letsinger et al., J. Am. Chem. Soc. 110:4470 (1988); and Pauwels et al., Chemica Scripta 26:141 91986)), phosphorothioate (Mag et al., Nucleic Acids Res. 19:1437 (1991); and U.S. Pat. No. 5,644,048), phosphorodithioate (Briu et al., J. Am. Chem. Soc. 111:2321 (1989), O-methylphosphoroamidite linkages (see Eckstein, Oligonucleotides and Analogues: A Practical Approach, Oxford University Press), and peptide nucleic acid (also referred to herein as "PNA") backbones and linkages (see Egholm, J. Am. Chem. Soc. 114:1895 (1992); Meier et al., Chem. Int. Ed. Engl. 31:1008 (1992); Nielsen, Nature, 365:566 (1993); Carlsson et al., Nature 380:207 (1996), all of which are incorporated by reference. Other analog nucleic acids include those with bicyclic structures including locked nucleic acids (also referred to herein as "LNA"), Koshkin et al., J. Am. Chem. Soc. 120.13252 3 (1998); positive backbones (Denpcy et al., Proc. Natl. Acad. Sci. USA 92:6097 (1995); non-ionic backbones (U.S. Pat. Nos. 5,386,023, 5,637,684, 5,602,240, 5,216,141 and 4,469,863; Kiedrowshi et al., Angew. Chem. Intl. Ed. English 30:423 (1991); Letsinger et al., J. Am. Chem. Soc. 110:4470 (1988); Letsinger et al., Nucleoside & Nucleotide 13:1597 (1994); Chapters 2 and 3, ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research", Ed. Y. S. Sanghui and P. Dan Cook; Mesmaeker et al., Bioorganic & Medicinal Chem. Lett. 4:395 (1994); Jeffs et al., J. Biomolecular NMR 34:17 (1994); Tetrahedron Lett. 37:743 (1996)) and non-ribose backbones, including those described in U.S. Pat. Nos. 5,235,033 and 5,034,506, and Chapters 6 and 7, ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research", Ed. Y. S. Sanghui and P. Dan Cook, all of which are incorporated by reference. Nucleic acids containing one or more carbocyclic sugars are also included within the definition of nucleic acids (see Jenkins et al., Chem. Soc. Rev. (1995) pp 169-176, hereby incorporated by reference). Several nucleic acid analogs are described in Rawls, C & E News

IDT Exhibit 1020.2319

Jun. 2, 1997 page 35, hereby incorporated by reference. "Locked nucleic acids" are also included within the definition of nucleic acid analogs. LNAs can be a class of nucleic acid analogues in which the ribose ring is "locked" by a methylene bridge connecting the 2'-O atom with the 4'-C atom. All of these references are hereby expressly incorporated by reference. These modifications of the ribose-phosphate backbone can be done to increase the stability and half-life of such molecules in physiological environments. For example, PNA:DNA and LNA-DNA hybrids can exhibit higher stability and thus can be used in some embodiments. The target nucleic acids can be single stranded or double stranded, as specified, or contain portions of both double stranded or single stranded sequence. Depending on the application, the nucleic acids can be DNA (including, e.g., genomic DNA, mitochondrial DNA, and cDNA), RNA (including, e.g., mRNA and rRNA) or a hybrid, where the nucleic acid contains any combination of deoxyribo- and ribo-nucleotides, and any combination of bases, including uracil, adenine, thymine, cytosine, guanine, inosine, xathanine hypoxathanine, isocytosine, isoguanine, etc.

[0039] As used herein, the terms “hybridization”, “hybridizing” and “annealing” can be used interchangeably and can refer to the pairing of complementary nucleic acids.

[0040] The term “primer”, as used herein, can refer to an oligonucleotide, generally with a free 3’ hydroxyl group, that is capable of hybridizing with a template (such as a target polynucleotide, target DNA, target RNA or a primer extension product) and is also capable of promoting polymerization of a polynucleotide complementary to the template. A primer can contain a non-hybridizing sequence that constitutes a tail of the primer. A primer can still hybridize to a target even though its sequences are not fully complementary to the target.

[0041] “Complementary”, as used herein, can refer to complementarity to all or only to a portion of a sequence. The number of nucleotides in the hybridizable sequence of a specific oligonucleotide primer can be such that stringency conditions used to hybridize the oligonucleotide primer can prevent excessive random non-specific hybridization. The number of nucleotides in the hybridizing portion of the oligonucleotide primer can be at least as great as the defined sequence on the target polynucleotide that the oligonucleotide primer hybridizes to, namely, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least about 20, or from about 6 to about 10 or 6 to about 12, or 12 to about 200 nucleotides, or about 10 to about 50 nucleotides. In general, the target polynucleotide can be larger than the oligonucleotide primer or primers as described previously.

IDT Exhibit 1020.2320

[0042] The term "genomic sequence", as used herein, can refer to a sequence that occurs in a genome. RNAs can be transcribed from a genome; this term can encompass sequence that exist in the nuclear genome of an organism, as well as sequences that are present in a cDNA copy of an RNA (e.g., an mRNA) transcribed from such a genome.

[0043] The term "target polynucleotide,", "target region", or "target", as used herein, can refer to a polynucleotide of interest under study. A target polynucleotide can contain one or more sequences or loci that are of interest and under study. A target polynucleotide can comprise, for example, a genomic region. The genomic region can comprise all or a portion of a chromosome. The genomic region can comprise a plurality of loci. The genomic region can comprise a plurality of loci for which phasing and/or haplotype information is desired.

[0044] The terms "template", "template strand", "template DNA" and "template nucleic acid" can be used interchangeably herein to refer to a strand of nucleic acid (e.g., DNA) that is copied by an amplification cycle.

[0045] The term "mutation", as used herein, can refer to a change of the nucleotide sequence of a genome. Mutations can involve large regions of a genome (e.g., copy number variation). Mutations can involve small regions of a genome. Examples of mutations involving small sections of a genome include, e.g., point mutations or single nucleotide polymorphisms, multiple nucleotide polymorphisms, insertions (e.g., insertion of one or more nucleotides at a locus), multiple nucleotide changes, deletions (e.g., deletion of one or more nucleotides at a locus), and inversions (e.g., reversal of a sequence of one or more nucleotides).

[0046] The term "locus", as used herein, can refer to a location of a gene, a nucleotide, or a sequence on a chromosome and/or in a genome. A locus can be considered "polymorphic" or to have a "polymorphism" if there exists in a population two or more forms of a nucleotide or sequence at the locus (e.g., if the locus comprises variant genetic sequences). An "allele" of a locus, as used herein, can refer to an alternative form of a nucleotide or sequence at the locus. A "wild-type allele" can refer to an allele that has the highest frequency in a population of subjects. A "mutant allele" can refer to an allele that has a lower frequency that a "wild-type allele". An "informative" locus, as used herein, can refer to a locus which can be used to determine phasing information of two or more alleles. The informative locus can be a polymorphic locus.

[0047] The term "single nucleotide polymorphism", or "SNP", as used herein, can refer to a type of genomic sequence variation resulting from a single nucleotide substitution within a

IDT Exhibit 1020.2321

sequence. "SNP alleles" or "alleles of a SNP" can refer to alternative forms of the SNP at a given locus.

[0048] The term "genotyping", as used herein, can refer to a process of determining differences in the genetic make-up (genotype) of an individual, e.g., by examining the individual's DNA sequence using a biological assay. The biological assay can comprise a sequencing reaction.

[0049] As used herein, a "sample" or "nucleic acid sample" can refer to any substance containing or presumed to contain nucleic acid. The sample can be a biological sample obtained from a subject.

[0050] An "addressable location", as used herein, can refes to a location that can be recorded and/or tracked throughout any of the procedures carried out during one or more methods described herein.

OVERVIEW

[0051] Provided herein are methods, compositions, and kits useful for sequence analysis of nucleic acids. The subject methods, compositions, and kits can be useful for obtaining accurate phasing and/or haplotyping information using NGS technologies. For example, the methods, compositions, and kits provided herein can be used to determine phasing of a plurality of alleles that are separated by a distance that is greater than a distance of an NGS sequence read or paired read. The methods and kits provided herein can be used to obtain phasing information for detected alleles that map to two or more loci. For example, the methods and kits provided herein can be used for determining whether detected alleles that map to two or more loci are located on the same nucleic acid template, on the same chromosome or chromosome region, or on different chromosomes or chromosomal regions in a chromosome pair. The methods and kits can be used to determine a plurality of polymorphic markers at a plurality of loci within a nucleic acid template. The methods and kits described herein can also be used to obtain phasing information for a plurality of alleles that are detected on different nucleic acid templates. One or more phasing methods disclosed herein can be used to provide phasing information for two or more alleles. The two or more alleles may be separated by a distance of over 150 bases, over 300 bases, over 500 bases, over 600 bases, over 700 bases, over 800 bases, over 900 bases, over 1000 bases (1kb), over 5 kb, over 10 kb, over 50 kb, over 100 kb, over 500 kb, over 1000 kb (1 Mb), over 5 Mb, over 10 Mb, over 50 Mb, over 100 Mb, over 150 Mb, over 200 Mb, over 250 Mb, or over 300 Mb.

IDT Exhibit 1020.2322

[0052] The methods and kits described herein can also be used to differentiate between closely related nucleic acid sequences, e.g., to distinguish a gene from a closely related pseudogene. Such methods can be useful, for example, for haplotyping, SNP phasing, determining downstream exons in RNA-seq, and in genetic diagnostics applications. The methods, kits and compositions described herein can utilize sequential sequencing of a nucleic acid template. Altogether, the phasing methods disclosed herein provide an improvement over the existing methods by offering a highly parallel, efficient method for obtaining phasing information.

[0053] **FIGS. 1A** and **1B** illustrate current limitations of NGS sequencing in obtaining phasing information for a plurality of alleles. **FIG. 1A** depicts a chromosome pair comprising a maternal and paternal chromosome obtained from genomic DNA from a subject. The chromosomes comprise a plurality of loci: locus X, locus Y, which is located, e.g., 600 bp downstream of locus X, and locus Z, which is located, e.g., 600 bp downstream of locus Y. The maternal chromosome comprises the following alleles: allele T at locus X (X_T, allele G at locus Y (Y_G), and allele T at locus Z (Z_T). By contrast, the paternal chromosome comprises the following alleles, allele A at locus X (X_A), allele T at locus Y (Y_T), and allele C at locus Z (Z_C). FIG. 1B depicts a current embodiment NGS sequencing method which is unable to determine phasing of alleles at loci X, Y, and Z. Genomic DNA from the subject is fragmented into ~1000 bp DNA fragments. The fragments are spatially separated and clonally amplified into clusters. By way of example only, clusters 1-3 tile overlapping regions which span loci X, Y, and Z. The clusters then undergo next generation sequencing, e.g., on an Illumina platform. Next generation sequencing typically produces sequencing reads from one or both ends of the DNA fragments, due to the sequencing adaptors attached to one or both ends of the fragments, wherein the sequencing adaptors comprise sequencing primer binding sites (not shown in the figure). The sequencing reads can have read lengths of, e.g., 100 nt, so the next generation sequencing in this example does not provide sequencing information across the entire template molecule. In this particular example, paired end sequencing is employed, producing sequencing reads from both ends of the template DNA fragments, each read having a read length of 100 nt (as depicted by thick gray arrows in **FIG. 1B**). Accordingly, the DNA template fragment of cluster 1 comprises X_A and Y_T, however, only X_A is detected by the sequencing reads. The DNA template fragment of cluster 2 comprises Y_T and Z_C, however, only Z_C is detected by the sequencing reads. The DNA template fragment of cluster 3 comprises Y_T and Z_C, however, only Y_T is detected by the sequencing reads. Because sequencing reads from clusters 1-3 only detect a

IDT Exhibit 1020.2323

single allele at a single locus for each cluster, phasing information is not generated for the detected alleles X_A, Y_T, and Z_C, e.g., whether they are located on the same template molecule, the same chromosomal region, or the same chromosome.

[0054] **FIG. 2** depicts an exemplary workflow of a phasing method described herein. The method can comprise a first step 210 of coupling a nucleic acid template to a sequencing platform. The coupling can comprise attaching the nucleic acid template to a solid support of the sequencing platform. Coupling can in some cases include formation of a covalent bond (e.g., formed by ligation), or a noncovalent bond (e.g., formed by hybridization) between the template and the solid support of the sequencing platform. In some cases, coupling does not involve formation of a covalent bond or noncovalent bond between the template and the solid support of the sequencing platform. In some cases, coupling can include associating two entities, e.g., placing an entity, such as a nucleic acid template, in a well. In some cases, the nucleic acid template is coupled to a unique addressable location of the sequencing platform. For example, a plurality of nucleic acid templates can be coupled to a sequencing platform, wherein each of the plurality of templates is coupled to a unique addressable location of the sequencing platform. Coupling of nucleic acid templates to unique addressable locations of a sequencing platform can comprise attaching the templates to the unique addressable locations and/or can comprise confining the templates within the unique addressable locations (e.g., confining to a well). For example, a template can be attached to a solid support (e.g., a bead) and the solid support confined within the unique addressable location, such as, e.g., depositing the bead in a well. The method can further comprise a second step 220 of subjecting the nucleic acid template to a first sequencing reaction using a first sequencing primer. The first sequencing primer can initiate sequencing at a first location on the nucleic acid template. The first sequencing reaction can generate a first sequencing read. The first sequencing read can provide the sequence of a first region of the nucleic acid template. The method can further comprise a second step 230 of subjecting the nucleic acid template to a second sequencing reaction using a second sequencing primer. The second sequencing primer can initiate sequencing at a second location on the nucleic acid template. The second location can be distinct from the first location. In some cases, a 3' terminal nucleotide of the second primer can hybridize to a location that is more than 5 nt away from a binding site of a 3' terminal nucleotide of the first primer. The second sequencing reaction can generate a second sequencing read. The second sequencing read can provide the sequence of a second region of the nucleic acid template which is distinct from the first region of the nucleic acid template. In some embodiments, the nucleic acid template is optionally subjected to one or

IDT Exhibit 1020.2324

more additional rounds of sequencing (depicted as step 240) using additional sequencing primers.

[0055] Nucleic acid templates can be subjected to any number of rounds of sequencing as described herein. For example, nucleic acid templates can be subjected to 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38. 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 100000, or more than 100000 rounds of sequencing according to one or more phasing methods described herein. Nucleic acid templates can be subjected to about 2 to about 8 rounds of sequencing, about 6 to about 20 rounds of sequencing, about 10 to about 100 rounds of sequencing, about 50 to about 500 rounds of sequencing, about 100 to about 1000 rounds of sequencing, about 500 to about 2000 rounds of sequencing, about 1000 to about 10000 rounds of sequencing, or more than about 10000 rounds of sequencing according to one or more phasing methods described herein.

[0056] The rounds of sequencing can produce any number of sequencing reads for the nucleic acid template. For example, about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 100000, or more than 100000 sequencing reads can be generated for any given nucleic acid template. About 1 to about 8 sequencing reads, about 6 to about 20 sequencing reads, about 10 to about 100 sequencing reads, about 50 to about 500 sequencing reads, about 100 to about 1000 sequencing reads, about 500 to about 2000 sequencing reads, about 1000 to about 10000 sequencing reads, or more than about 10000 sequencing reads can be generated for any given nucleic acid template,

[0057] In some embodiments, the first, second, and/or any additional rounds of sequencing are conducted using a plurality of primers. For example, any round of sequencing designed to sequence a particular locus can comprise use of a pair of primers, wherein each primer in the primer pair is designed to sequence a forward or reverse complement strand containing

IDT Exhibit 1020.2325

the locus. For example, any round of sequencing can comprise sequencing using a mixture of primers. Primers within a mixture can be designed and/or selected to have minimal potential to hybridize to the same nucleic acid template. For example, primers within a mixture can be designed to hybridize to locations that are at least 1 kb, 2 kb, 3kb, 4 kb, 5 kb, 10 kb, 20 kb, 30 kb, 40 kb, 50 kb, 60 kb, 70 kb, 80 kb, 90 kb, 100 kb, 200 kb, 300 kb, 400 kb, 500 kb, 600 kb, 700 kb, 800 kb, 900 kb, 1000 kb (1Mb), 2 Mb, 3 Mb, 4 Mb, 5 Mb, 6 Mb, 7 Mb, 8 Mb, 9 Mb, 10 Mb, , 20 Mb, 30 Mb, 40 Mb, 50 Mb, 60 Mb, 70 Mb, 80 Mb, 90 Mb, 100 Mb, or 200 Mb apart. Primers within a mixture can be designed to hybridize to different chromosomes. A mixture of primers used for any round of sequencing can comprise any number of primers, for example, can comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38. 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 100000, or more than 100000 primers. A mixture of priemrs used for any round of sequencing can comprise about 1 to about 8 primers, about 6 to about 20 primers, about 10 to about 100 primers, about 50 to about 500 primers, about 100 to about 1000 primers, about 500 to about 2000 primers, about 1000 to about 10000 primers, or more than about 10000 primers.

[0058] In some cases, the first, second, and optionally additional sequencing reads can be compiled to produce sequence information for the nucleic acid template. For example, first, second, and optionally additional sequencing reads from a unique addressable location can be compiled to produce sequence information for the nucleic acid template coupled to that unique addressable location.

[0059] **FIG. 3** depicts an exemplary embodiment of a phasing method described herein, and highlights advantages of the subject method for obtaining phasing information for a plurality of alleles. **FIG. 3** depicts the same chromosome pair comprising a maternal and paternal chromosome obtained from genomic DNA from a subject as depicted in **FIG. 1**. As in **FIG. 1**, the chromosomes comprise a plurality of loci: locus X, locus Y, which is located, e.g., 600 bp downstream of locus X, and locus Z, which is located, e.g., 600 bp downstream of locus Y. As in **FIG. 1**, the maternal chromosome comprises the following alleles: allele T at locus X (X_T), allele G at locus Y (Y_G), and allele T at locus Z (Z_T). As in **FIG. 1**, the paternal chromosome comprises the following alleles, allele A at locus X (X_A), allele T at

IDT Exhibit 1020.2326

locus Y (Y_T), and allele C at locus Z (Z_C). As in **FIG. 1**, genomic DNA from the subject is fragmented into ~1000 bp DNA fragments. As in **FIG. 1**, the fragments are spatially separated and clonally amplified into clusters. As in **FIG. 1**, clusters 1-3 tile overlapping regions which span loci X, Y, and Z. The clusters are then subjected to sequential rounds of sequencing according to a phasing method described herein. A first round of sequencing is performed using a first sequencing primer. The first sequencing primer can be designed to selectively hybridize to a sequencing primer binding site of a sequencing adaptor attached to a first end of the cluster templates. The first round of sequencing using the first sequencing primer provides a first set of sequencing reads (dark gray arrows) of a first end of the cluster templates. Because locus X happens to be located near the end of the first end of cluster 1, Allele A_A is detected by the first read of cluster 1. The extension strands resulting from extension of the first sequencing primer are then denatured and removed from the clusters. In a next step, a second round of sequencing is performed using a second sequencing primer. The second sequencing primer can be designed to hybridize to a genomic location upstream of locus Y and to sequence locus Y. Accordingly, the second round of sequencing produces a second set of sequencing reads (medium gray arrows). Because clusters 1, 2, and 3 comprise locus Y, the second set of sequencing reads comprise reads from clusters 1, 2, and 3. The second set of sequencing reads detect allele T at locus Y (Y_T) in clusters 1, 2, and 3. The extension strands resulting from extension of the second sequencing primer are then denatured and removed from the clusters. In a next step, a third round of sequencing is performed using a third sequencing primer. The third sequencing primer can be designed to hybridize to a genomic location upstream of locus Z and to sequence locus Z. Accordingly, the third round of sequencing produces a third set of sequencing reads (stippled medium gray arrows). Because clusters 2 and 3, but not cluster 1, comprise locus Z, the third set of sequencing reads comprise reads from clusters 2 and 3 but not from cluster 1. The third set of sequencing reads detect allele C at locus Z (Z_C) in clusters 2 and 3. Sequence data from the first, second, and third sets of sequencing reads is compiled and used to determine phasing of the detected alleles at loci X, Y, and Z. Because the first and second set of sequencing reads reveal X_A and Y_T in cluster 1, it can be determined that X_A and Y_T are in phase (e.g., on the same nucleic acid fragment). Because the second and third sets of sequencing reads reveal Y_T and Z_C in clusters 2 and 3, it can be determined that Y_T and Z_C are in phase (e.g., on the same nucleic acid fragment). Moreover, because the first, second, and third clusters share a common allele (Y_T), it can be determined that X_A, Y_T, and Z_C are all in

IDT Exhibit 1020.2327

phase. Because X_A, Y_T, and Z_C are all in phase, a haplotype of the subject can be determined to be $X_A Y_T Z_C$.

[0060] **FIG. 4** depicts another exemplary embodiment of a phasing method described herein. **FIG. 4** depicts the same chromosome pair comprising a maternal and paternal chromosome obtained from genomic DNA from a subject as depicted in **FIG. 1**. As in **FIG. 1**, the chromosomes comprise a plurality of loci: locus X, locus Y, which is located, e.g., 600 bp downstream of locus X, and locus Z, which is located, e.g., 600 bp downstream of locus Y. As in **FIG. 1**, the maternal chromosome comprises the following alleles: allele T at locus X (X_T), allele G at locus Y (Y_G), and allele T at locus Z (Z_T). As in **FIG. 1**, the paternal chromosome comprises the following alleles, allele A at locus X (X_A), allele T at locus Y (Y_T), and allele C at locus Z (Z_C). As in **FIG. 1**, genomic DNA from the subject is fragmented into ~1000 bp DNA fragments. As in **FIG. 1**, the fragments are spatially separated and clonally amplified into clusters. As in **FIG. 1**, clusters 1-3 tile overlapping regions which span loci X, Y, and Z. The clusters are then subjected to sequential rounds of sequencing according to a phasing method described herein. A first round of sequencing is performed using a first sequencing primer. The first sequencing primer can be designed to selectively hybridize to a genomic location upstream of locus X and to sequence locus X. The first round of sequencing produces a first set of sequencing reads. Because cluster 1, but not clusters 2 or 3, comprise locus X, the first set of sequencing reads comprises a read from cluster 1 but not from clusters 2 or 3. The first set of sequencing reads detects allele A at locus X(X_A). The extension strands resulting from extension of the first sequencing primer are then denatured and removed from the clusters. In a next step, a second round of sequencing is performed using a second sequencing primer. The second sequencing primer can be designed to hybridize to a genomic location upstream of locus Y and to sequence locus Y. Accordingly, the second round of sequencing produces a second set of sequencing reads (medium gray arrows). Because clusters 1, 2, and 3 all comprise locus Y, the second set of sequencing reads comprise reads from clusters 1, 2, and 3. The second set of sequencing reads detect allele T at locus Y ((Y_T) in clusters 1, 2, and 3. The extension strands resulting from extension of the second sequencing primer are then denatured and removed from the clusters. In a next step, a third round of sequencing is performed using a third sequencing primer. The third sequencing primer can be designed to hybridize to a genomic location upstream of locus Z and to sequence locus Z. Accordingly, the third round of sequencing produces a third set of sequencing reads (stippled medium gray arrows). Because clusters 2 and 3, but not cluster 1, comprise locus Z, the third set of sequencing reads

IDT Exhibit 1020.2328

comprise reads from clusters 2 and 3 but not from cluster 1. The third set of sequencing reads detect allele C at locus Z (Z_C) in clusters 2 and 3. Sequence data from the first, second, and third sets of sequencing reads is compiled and used to determine phasing of the detected alleles at loci X, Y, and Z. Because the first and second set of sequencing reads reveal X_A and Y_T in cluster 1, it can be determined that X_A and Y_T are in phase. Because the second and third sets of sequencing reads reveal Y_T and Z_C in clusters 2 and 3, it can be determined that Y_T and Z_C are in phase. Moreover, because the first, second, and third clusters share a common allele (Y_T), it can be determined that X_A, Y_T, and Z_C are all in phase. Because X_A, Y_T, and Z_C are all in phase, a haplotype of the subject can be determined to be X_A Y_T Z_C.

[0061] Reference will now be made in detail to exemplary embodiments. While the disclosed methods and compositions will be described in conjunction with the exemplary embodiments, it will be understood that these exemplary embodiments are not intended to limit the methods and compositions described herein, but can encompass alternatives, modifications and equivalents, which may be included in the spirit and scope described herein

Exemplary nucleic acid templates

[0062] Nucleic acid templates to be sequenced by a method provided herein can comprise DNA. The DNA can be complex DNA, for example genomic DNA. The DNA can be from a linear or circular genome. The DNA can be plasmid DNA, cosmid DNA, bacterial artificial chromosome (BAC), or yeast artificial chromosome (YAC). The DNA can be from more than one individual or organism. The DNA can be double stranded or single stranded. The input DNA can be part of chromatin. The DNA can be associated with histones. The DNA can comprise one or more chromosomes. For example, if the DNA is from a human, the DNA can comprise one or more of chromosome 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, X, or Y. The DNA can be mitochondrial DNA. The DNA can be cell-free DNA. The DNA can be cDNA. The cDNA can be generated from RNA, e.g., mRNA. Nucleic acid tempaltes to be sequenced by a method provided herein can comprise RNA.

[0063] Nucleic acid templates to be sequenced by a method provided herein can be obtained from and/or can be in a nucleic acid sample. The nucleic acid sample can be obtained from a subject. A subject can be any biological entity of interest from which phasing and/or haplotype information is desired. The subject can be an animal, plant, fungus, insect, bacteria, algae, virus, cell, tissue, and the like. The animal can be human, non-human primate, dog, cat, cow, pig, sheep, guinea pig, hamster, bird, frog, fish, rat, mouse, other

IDT Exhibit 1020.2329

rodent or other animal. A nucleic acid sample can also comprise a mixture of genomes of different species such as host-pathogen, bacterial populations, viral populations, and the like. For example, a DNA sample can comprise DNA (e.g., cDNA) made from a mixture of genomes of different species. Alternatively, the nucleic acid sample can be from a synthetic source.

[0064] The nucleic acid sample can be obtained from a biological sample from the subject. The biological sample can be, e.g., a liquid sample, a solid sample, or a combination thereof. Exemplary liquid samples include, but are not limited to ascites, buccal sample, cavity rinse, cerebrospinal fluid, whole blood, plasma, serum, sweat, tears, sputum, saliva, urine, or organ rinse. The sample can be a cell-free sample. Exemplary cell-free samples include, but are not limited to plasma, serum, plasma, urine, sweat, saliva, sputum, and tears. Exemplary solid biological samples include, e.g., tissue biopsy, feces, and the like. A biological sample can also comprise cultured cells or constituents of culture cells (including, e.g., conditioned medium resulting from the growth of cells in cell culture medium, recombinant cells and cell components). The sample can be a single cell. Nucleic acid samples can be obtained from biological samples by any means known to those of skill in the art. For example, nucleic acid can be extracted from a biological sample using liquid extraction (e.g, Trizol, DNAzol) techniques. Nucleic acid can also be extracted using commercially available kits. Nucleic acids can be concentrated by, e.g., centrifugation and/or precipitation. Nucleic acids can be purified by binding to a selective membrane and then by elution. In some cases, gentle handling steps may be taken to avoid shearing of chromosomal or intact DNA during the extraction process. Exemplary gentle handling strategies can include, e.g., pipetting with a wide-bore pipette tip and avoidance of vigorous pipetting and/or vortexing. In some cases, gentle handling steps are unnecessary for preparation of nucleic acid samples.

[0065] The nucleic acid sample can be fragmented by any means known in the art or otherwise described herein. The nucleic acids can be fragmented to result in a population of nucleic acid fragments of a desired length. The nucleic acid fragments can be double stranded or single stranded fragments. The length of the nucleic acid fragments can be any length. A nucleic acid fragment to be sequenced by a method described herein can have a length that is less than 300 Mb, less than 200 Mb, less than 100 Mb, less than 50 Mb, less than 40 Mb, less than 30 Mb, less than 20 Mb, less than 10 Mb, less than 5 Mb, less than 3 Mb, less than 2 Mb, less than 1 Mb (1000 kb), less than 900 kb, less than 800 kb, less than 700 kb, less than 600 kb, less than 500 kb, less than 400 kb, less than 300 kb, less than 200 kb, less than 100 kb (100,000 bases), less than 100,000 bases, less than 90,000 bases, less

IDT Exhibit 1020.2330

than 80,000 bases, less than 70,000 bases, less than 60,000 bases, less than 50,000 bases, less than 40000 bases, less than 30,000 bases, less than 20,000 bases, less than 10,000 bases, less than 9000 bases, less than 8000 bases, less than 7000 bases, less than 6000 bases, less than 5000 bases, less than 4000 bases, less than 3000 bases, less than 2000 bases, less than 1500 bases, less than 1400 bases, less than 1300 bases, less than 1200 bases, less than 1100 bases, less than 1000 bases, less than 900 bases, less than 800 bases, less than 700 bases, less than 600 bases, less than 500 bases, less than 400 bases, less than 300 bases, less than 250 bases, less than 200 bases, less than 150 bases, less than 100 bases, less than 50 bases, less than 40 bases, less than 30 bases, less than 20 bases, or less than 10 bases. In some cases, a nucleic acid fragment to be sequenced by a method described herein has a length that is about 10 bases, about 20 bases, about 30 bases, about 40 bases, about 50 bases, about 100 bases, about 150 bases, about 200 bases, about 250 bases, about 300 bases, about 400 bases, about 500 bases, about 600 bases, about 700 bases, about 800 bases, about 900 bases, about 1000 bases, about 1100 bases, about 1200 bases, about 1300 bases, about 1400 bases, about 1500 bases, about 1600 bases, about 1700 bases, about 1800 bases, about 1900 bases, about 2000 bases, about 3000 bases, about 4000 bases, about 5000 bases, about 6000 bases, about 7000 bases, about 8000 bases, about 9000 bases, about 10000 bases, about 20000 bases, about 30000 bases, about 40000 bases, about 50000 bases, about 60000 bases, about 70000 bases, about 80000 bases, about 90000 bases, about 100000 bases (100 kb), about 200 kb, about 300 kb, about 400 kb, about 500 kb, about 600 kb, about 700 kb, about 800 kb, about 900 kb, about 1000 kb (1Mb), about 2 Mb, about 3 Mb, about 5 Mb, about 10 Mb, about 10 Mb, about 30 Mb, about 40 Mb, about 50 Mb, about 100 Mb, about 200 Mb, or about 300 Mb. In some cases, a nucleic acid fragment to be sequenced by a method described herein has a length that is more than about 10 bases, more than about 20 bases, moer than about 30 bases, more than about 40 bases, more than about 50 bases, more than about 100 bases, more than about 150 bases, more than about 200 bases, more than about 250 bases, more than about 300 bases, more than about 400 bases, more than about 500 bases, more than about 600 bases, more than about 700 bases, more than about 800 bases, more than about 900 bases, more than about 1000 bases, more than about 1100 bases, more than about 1200 bases, more than about 1300 bases, more than about 1400 bases, more than about 1500 bases, more than about 1600 bases, more than about 1700 bases, more than about 1800 bases, more than about 1900 bases, more than about 2000 bases, more than about 3000 bases, more than about 4000 bases, more than about 5000 bases, more than about 6000 bases, more than about 7000 bases, more than about 8000 bases, more than about 9000 bases, more than about 10000 bases, more than about

IDT Exhibit 1020.2331

20000 bases, more than about 30000 bases, more than about 40000 bases, more than about 50000 bases, more than about 60000 bases, more than about 70000 bases, more than about 80000 bases, more than about 90000 bases, more than about 100000 bases (100 kb), more than about 200 kb, more than about 300 kb, more than about 400 kb, more than about 500 kb, more than about 600 kb, more than about 700 kb, more than about 800 kb, more than about 900 kb, more than about 1000 kb (1Mb), more than about 2 Mb, more than about 3 Mb, more than about 5 Mb, more than about 10 Mb, more than about 10 Mb, more than about 30 Mb, more than about 40 Mb, more than about 50 Mb, more than about 100 Mb, more than about 200 Mb, or more than about 300 Mb. A nucleic acid fragment to be sequenced by a method described herein can have a length that is between about 10 to about 100 bases, about 50 to about 200 bases, about 100 to about 250 bases, about 150 to about 500 bases, about 300 to about 700 bases, about 400 to about 800 bases, about 600 to about 900 bases, about 800 to about 1000 bases, about 200 to about 1000 bases, about 500 bases to about 10000 bases, or about 2000 bases to about 100000 bases. In some embodiments, the nucleic acid fragment has a length that is over 9000 bases. For example, the nucleic acid fragments can be about 0.9 to about 2 kb, about 1.5 to about 4 kb, about 2 to about 6 kb, about 5 to about 8 kb, or about 7 to about 10 kb, about 10 to about 100 kb, about 50 kb to about 1 Mb, about 1 Mb to about 50 Mb, about 10 Mb to about 300 Mb. One of skill in the art will readily understand that the length of the nucleic acid templates to be sequenced by a method disclosed herein can be any length. Length of nucleic acid templates to be sequenced by a method disclosed herein can vary depending on the sequencing platform used in practicing a method disclosed herein. In some embodiments, the nucleic acid sample is not fragmented.

[0066] The nucleic acid sample can be enriched for target polynucleotides of interest. Target polynucleotides can comprise polynucleotides that map to genomic regions of interest. A genomic region of interest can be any genomic region from which phasing and/or haplotype information is desired. The genomic region of interest can comprise a region of a chromosome. The genomic region of interest can comprise a whole chromosome. The chromosome can be a diploid chromosome. In a human genome, for example, the diploid chromosome can be any of chromosomes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23. In some cases, the chromosome can be an X or Y chromosome. In some cases, the genomic region of interest comprises a portion of a chromosome. A genomic region of interest can be of any length. The genomic region of interest can have a length that is between, e.g., about 1 to about 10 bases, about 5 to about 50 bases, about 10 to about 100 bases, about 70 to about 300 bases, about 200 bases to about 1000 bases (1 kb), about 700

IDT Exhibit 1020.2332

bases to about 2000 bases, about 1 kb to about 10 kb, about 5 kb to about 50 kb, about 20 kb to about 100 kb, about 50 kb to about 500 kb, about 100 kb to about 2000 kb (2 Mb), about 1 Mb to about 50 Mb, about 10 Mb to about 100 Mb, about 50 Mb to about 300 Mb. For example, a genomic region of interest can be over 1 base, over 10 bases, over 20 bases, over 50 bases, over 100 bases, over 200 bases, over 400 bases, over 600 bases, over 800 bases, over 1000 bases (1kb), over 1.5 kb, over 2 kb, over 3 kb, over 4 kb, over 5 kb, over 10 kb, over 20 kb, over 30 kb, over 40 kb, over 50 kb, over 60 kb, over 70 kb, over 80 kb, over 90 kb, over 100 kb, over 200 kb, over 300 kb, over 400 kb, over 500 kb, over 600 kb, over 700 kb, over 800 kb, over 900 kb, over 1000 kb (1Mb), over 2 Mb, over 3 Mb, over 4 Mb, over 5 Mb, over 6 Mb, over 7 Mb, over 8 Mb, over 9 Mb, over 10 Mb, , over 20 Mb, over 30 Mb, over 40 Mb, over 50 Mb, over 60 Mb, over 70 Mb, over 80 Mb, over 90 Mb, over 100 Mb, or over 200 Mb.

[0067] A genomic region of interest can comprise one or more informative loci. An informative locus can be a polymorphic locus, e.g., comprising two or more alleles. In some cases, the two or more alleles comprise a minor allele. In particular embodiments, genomic regions of interest comprise one or more HLA loci.

[0068] Enrichment of target polynucleotides can be by any means known in the art or otherwise described herein. For example, nucleic acids can be enriched using target-specific primers, or by hybridization to target-selective probes. Target-selective probes can comprise a nucleic acid sequence that is selectively hybridizable to a target polynucleotide of interest. The target-selective probes can comprise a capture moiety which enables capture by a capture reagent. Exemplary capture moiety/reagents include, e.g., biotin/avidin, streptavidin, neutravidin, digoxigenin/wheat germ agglutinin, histidine/cobalt, iron. Nucleic acids can be enriched by a target-selective sequencing library preparation method. Target-selective sequencing library preparation kits and technology are commercially available from, e.g., NuGEN, Agilent, Illumina, or Nimblegen. Target-selective sequencing library preparation methods from Agilient are described in, e.g., US Patent No. 7,867,703 and US Patent Application Publication No. 20120289426, which are hereby incorporated by reference. Target-selective sequencing library preparation methods from Illumina are described in, e.g., US Patent Application Pub No. 20070141604, hereby incorporated by reference. Target-selective sequencing library preparation methods from Nimblegen are described in, e.g., US Patent Application Pub. No. 20090105081, hereby incorporated by reference. Target-selective sequencing library preparation methods from NuGEN are described in, e.g., US Patent Application Pub. No. 20130231253, hereby incorporated by reference. Other

IDT Exhibit 1020.2333

exemplary target-selective sequencing library preparation methods are described in US Patent Application Publication No. 20120003657, hereby incorporated by reference. Enrichment of target genomic regions can, in some cases, comprise isolation of intact chromosomes or chromosomal regions from a biological sample. In some embodiments, the nucleic acid sample is not enriched for target polynucleotides, e.g., represents a whole genome.

[0069] The nucleic acid template to be sequenced can be a member of a sequencing library. Sequencing libraries can comprise any number of library members. Sequencing libraries can comprise 2, 10, 50, 100, 500, 1000, 5000, 10000, 50000, 100000, 500000, 1000000, 5000000, 10000000, 50000000, 100000000, 500000000, 1000000000, 2000000000, 3000000000, 4000000000, 5000000000, 6000000000, 7000000000, 8000000000, 9000000000, 10000000000, or more than 10000000000 members. In some embodiments, nucleic acid libraries are sequenced according to one or more methods described herein. The sequencing library can be prepared from a nucleic acid sample by any means known to those of skill in the art or otherwise described herein. In some embodiments, library preparation comprises attachment of one or more adaptor oligonucleotides to one or more ends of nucleic acid fragments. In some cases, adaptor oligonucleotides are attached to both ends of nucleic acid fragments. The term "adaptor", as used herein, can refer to an oligonucleotide. In some cases, the attachment of an adaptor to a nucleic acid template can enable the generation of amplification-ready products of the nucleic acid template. An adaptor can have various designs. In some embodiments, an adaptor comprises a primer binding site for a platform-specific sequencing primer. In some embodiments, an adaptor comprises a barcode sequence. The barcode sequence can be a unique sequence of nucleotides that can encode information about an assay. A barcode sequence can, for example, encode information relating to the identity of an interrogated allele, identity of a target region of interest, identity of a sample, identity of a subject, or any combination thereof.

[0070] Attachment of adaptor oligonucleotides to nucleic acid fragments can be by any means known in the art, including, e.g., hybridization, amplification, ligation, and the like. Various ligation processes and reagents are known in the art and can be useful for attachment of adaptor oligonucleotides to nucleic acid fragments. For example, blunt ligation can be employed. Similarly, a single dA nucleotide can be added to the 3'-end of the double-stranded DNA product, by a polymerase lacking 3'-exonuclease activity and can anneal to an adaptor comprising a dT overhang (or the reverse). This design allows the hybridized components to be subsequently ligated (e.g., by T4 DNA ligase). Other ligation strategies and the corresponding reagents and known in the art and kits and reagents for carrying out

IDT Exhibit 1020.2334

efficient ligation reactions are commercially available (e.g, from New England Biolabs, Roche).

[0071] In some cases, it is useful to preserve information about the direction of single-stranded nucleic acid molecules while generating double-stranded polynucleotides. This may be accomplished by creation of a strand specific, e.g., directional sequencing library. The term "strand specific" or "directional", as used herein, can refer to the ability to differentiate in a double-stranded polynucleotide between the original template strand and the strand that is complementary to the original template strand. In some cases, one of the strands of the double-stranded polynucleotide is synthesized so that it has at least one modified nucleotide incorporated into it along the entire length of the strand. In some embodiments, the incorporation of the modified nucleotide marks the strand for degradation or removal. In some embodiments, the methods provided herein contemplate construction of directional nucleic acid libraries as described in pending US Application, Serial No. 13/643,056, titled COMPOSITIONS AND METHODS FOR DIRECTIONAL NUCLEIC ACID AMPLIFICATION AND SEQUENCING, Serial No. 13/643,056, hereby incorporated by reference in its entirety. In some cases, information about the direction of single-stranded nucleic acid molecules is not preserved while generating double-stranded polynucleotides. In some cases, the sequencing library is not a directional sequencing library.

[0072] Nucleic acid library members can optionally be amplified prior to sequencing. Methods of amplification are well known in the art. In some embodiments, the amplification is exponential, e.g. in the enzymatic amplification of specific double stranded sequences of DNA by a polymerase chain reaction (PCR). In other embodiments the amplification method is linear. In other embodiments the amplification method is isothermal. In some embodiments, the amplification comprises digital PCR. The digital PCR can be droplet digital PCR. In some embodiments, the amplification comprises single primer isothermal amplification (SPIA). In some embodiments, the amplification comprises reverse transcription PCR. In some embodiments, the amplification comprises real-time PCR. In some embodiments, the amplification comprises quantitative PCR (Q-PCR).

[0073] Nucleic acid library members can be clonally amplified prior to sequencing. Clonal amplification can involve generation of identical copies of individual nucleic acid templates in a nucleic acid library. Nucleic acid templates can be clonally amplified by any means known to those of skill in the art or otherwise described herein. In some cases, clonal amplification can comprise spatial separating individual nucleic acid template molecules and performing amplification of the separated nucleic acid template molecules. In certain

IDT Exhibit 1020.2335

instances, clonal amplification can comprise emulsion PCR (454 Life Sciences, Life Technologies). Emulsion PCR can comprise attachment of an individual nucleic acid template to a bead or particle. Beads and/or particles can be isolated in a water-in-oil emulsion, wherein droplets of the water-in-oil emulsion comprise a bead or particle with a template molecule attached thereon, and reaction components for carrying out an amplification reaction. Amplification of templates within the emulsion can result in clonal amplification of the templates. Clonal amplification can comprise bridge amplification (Illumina, Inc.). Bridge amplification is described, e.g, in US Patent No. 7,985,565, which is hereby incorporated by reference. Clonal amplification can comprise template walking (Life Technologies), described in US Patent Application Publication No. 20120156728, hereby incorporated by reference.

Primers

[0074] Primers can be oligonucleotides that are employed in an extension reaction by a polymerase along a polynucleotide template, such as in PCR or cDNA synthesis, for example. An oligonucleotide primer can be a synthetic polynucleotide that is single stranded, containing a sequence at its 3'-end that is capable of hybridizing with a sequence of the target polynucleotide. In some cases, the 3' region of the primer that hybridizes with the target nucleic acid has at least 80%, 90%, 95%, or 100% complementarity to a sequence or primer binding site.

[0075] Primers can be sufficiently long to prime synthesis of an extension strand along a template. Primers used in one or more methods described herein can have a length. The length can be any length, e.g., about 5 to about 15 nt, about 10 to about 20 nt, about 15 to about 30 nt, about 20 to about 40 nt, about 30 to about 60 nt, about 50 to about 100 nt, or more than about 100 nt.

[0076] Primers used in one or more methods described herein can include primers designed to hybridize to a portion or all of a sequencing adaptor. In some cases, primers described herein are designed to hybridize to a primer-binding sequence of a sequencing adaptor. Exemplary adaptors are described herein. In some cases, a first sequencing primer to be used in a sequential sequencing method (e.g., primer or primers to be used in a first round of sequencing) is designed to hybridize to a portion or all of a sequencing adaptor.

[0077] Primers used in one or more methods can include primers designed to sequence one or more loci from which phasing information is desired. The one or more loci can comprise one or more informative loci. For example, in some cases, the first sequencing primer, second sequencing primer (e.g., primer or primers to be used in a first or second round of

IDT Exhibit 1020.2336

sequencing, respectively), or any subsequent sequencing primer used in a sequential sequencing method is designed to sequence an informative locus.

[0078] A primer designed to sequence an informative locus can be designed to hybridize to a genomic location that is a distance upstream of the informative locus to be sequenced. The distance can be of any length. A skilled artisan will readily appreciate that the distance can vary depending on the read length of the sequencing technology utilized in the sequential sequencing method. In some cases, the distance is shorter than a read length of the sequencing technology utilized in the sequential sequencing method. The primer designed to sequence an informative locus can be designed to hybridize to a genomic location that is about, e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 100000, or more than 100000 nt upstream of the informative locus. The primer designed to sequence an informative locus can be designed to hybridize to a genomic location that is about 1 to about 10, about 5 to about 50, about 20 to about 100, about 40 to about 200, about 80 to about 500, about 100 to about 1000, about 500 to about 10000, about 5000 to about 100000, or more than 100000 nt upstream of the informative locus.

[0079] In some cases, the primer designed to sequence an informative locus can be designed to hybridize to a conserved region of the genome. The conserved region can be a region that does not comprise a polymorphic locus, e.g., is substantially not polymorphic. In some embodiments, the conserved region exhibits a polymorphism (e.g., sequence variant) at a population frequency that is less than 5%, less than 4%, less than 3%, less than 2%, less than 1%, less than 0.9%, less than 0.8%, less than 0.7%, less than 0.6%, less that 0.5%, less than 0.4%, less than 0.3%, less than 0.2%, less than 0.1%, less than 0.05%, less than 0.01%, less than 0.05%, less than 0.001%, less than 0.005%, less than 0.0001%, %, less than 0.0005%, less than 0.00001%,%, less than 0.00005%, less than 0.000001%, less than 0.000005%, less than 0.0000001%, less than 0.0000005%, or less than 0.00000001%. In some cases, the primer designed to sequence an informative locus can be designed to hybridize to an unconserved region of the genome.

[0080] Primers used in a first, second, or additional round of sequencing can hybridize to distinct locations on the nucleic acid template. For example, a first primer used in a first

IDT Exhibit 1020.2337

round of sequencing may comprise a 3' terminal nucleotide that hybridizes to a first location of a nucleic acid template. A second primer used in a second round of sequencing may comprise a 3' terminal nucleotide that hybridizes to a second location that is more than 5 nt away from the first location of the nucleic acid template (e.g., hybridizes to a location that is at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38. 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 100000, or more than 100000 nt away from the first location of the nucleic acid template. In some cases, a third primer used in a third round of sequencing may comprise a 3' terminal nucleotide that hybridizes to a third location that is more than 5 nt away from the second location of the nucleic acid template (e.g., hybridizes to a nucleotide that is at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38. 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 100000, or more than 100000 nt away from the second location of the nucleic acid template. Any primer used in a subsequent round of sequencing can have a 3' terminal nt that hybridizes to a location of the nucleic acid template that is at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38. 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 100000, or more than 100000 nt away from a binding site of the 3' terminal nt of the previous sequencing primer used in the previous round of sequencing.

[0081] Primers used in a first, second, or additional round of sequencing can initiate sequencing at distinct locations on the nucleic acid template. For example, a first primer used in a first round of sequencing may initiate sequencing at a first location of a nucleic acid template. A second primer used in a second round of sequencing may initiate sequencing at a

IDT Exhibit 1020.2338

second location that is more than 5 nt away from the first location of the nucleic acid template (e.g., initiates sequencing at a location that is at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 100000, or more than 100000 nt away from the first location of the nucleic acid template. In some cases, a third primer used in a third round of sequencing may initiate sequencing at a third location that is more than 5 nt away from the second location of the nucleic acid template (e.g., initiates sequencing at a location that is at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 100000, or more than 100000 nt away from the second location of the nucleic acid template. Any primer used in a subsequent round of sequencing can initiate sequencing at a location of the nucleic acid template that is more than 5 nt away from the location of initiated sequencing for the previous round of sequencing, e.g., at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 100000, or more than 100000 nt away from the location of initiated sequencing for the previous round of sequencing.

[0082] In some embodiments, standard or universal sequencing primers are used. For example, primers designed to hybridize to a primer binding site of a sequencing adaptor can be used. A primer can be a polyA or polyT oligonucleotide. In some embodiments, sequence-specific primers that hybridize to a conserved region or conserved regions within the nucleic acid inserts in the sequencing library are used. In some embodiments, the sequence-specific primers are designed to hybridize to conserved regions adjacent to regions of variable sequence within the nucleic acid inserts, thereby enabling differentiating between

IDT Exhibit 1020.2339

closely related sequences. In some embodiments, a set of oligonucleotide primers that hybridize to sequences shared in closely related sequences, such as gene/pseudogene pairs, are used.

[0083] Primers can be prepared by any suitable method. Methods for preparing oligonucleotides of specific sequence are known in the art, and include, for example, cloning and restriction of appropriate sequences or direct chemical synthesis. Chemical synthesis methods may include, for example, the phosphodiester method disclosed by Brown et al., 1979, Methods in Enzymology 68:109, the phosphotriester method described by Narang et al., 1979, Methods in Enzymology 68:90, the diethylphosphoramidate method disclosed in Beaucage et al., 1981, Tetrahedron Letters 22:1859, and the solid support method disclosed in U.S. Pat. No. 4,458,066, all of which publications are hereby incorporated by reference. Primers and/or reporter probes can also be obtained from commercial sources such as Amersham Pharmacia Biotech, Operon Technologies, Sigma, IDT Technologies, and Life Technologies. Computer programs can also be used to design primers, including but not limited to Array Designer Software (Arrayit Inc.), NetPrimer, Oligonucleotide Probe Sequence Design Software for Genetic Analysis (Olympus Optical Co.), Primer3, and DNAsis from Hitachi Software Engineering.

Methods of Sequencing

[0084] The methods provided herein contemplate sequential sequencing of nucleic acid templates. Sequencing methods are also well known in the art. Sequencing can be performed by a sequencing platform. The sequencing platform can be a next-generation sequencing (NGS) platform. In some cases, the NGS platform can sequence clonally amplified DNA templates or single DNA molecules in a massively parallel fashion (e.g. as described in Volkerding et al. Clin Chem 55:641-658 [2009]; Metzker M Nature Rev 11:31-46 [2010]), hereby incorporated by reference. Sequential sequencing methods provided herein can also employ Sanger sequencing.

[0085] The next-generation sequencing platform can be a commercially available platform. Commercially available platforms useful for performing a method described herein include, e.g., platforms for sequencing-by-synthesis, ion semiconductor sequencing (ChemFET sequencing), reversible dye terminator sequencing, pyrosequencing, sequencing by ligation, single-molecule sequencing, and sequencing by hybridization.

[0086] Platforms for sequencing by synthesis are available from, e.g., 454 Life Sciences, Illumina, Helicos Biosciences, and Qiagen. Illumina platforms can include, e.g., Illumina's Solexa platform, Illumina's Genome Analyzer, Illumina's MiSeq, Illumina's HiSeq,

IDT Exhibit 1020.2340

Illumina's NextSeq, and are described in Gudmundsson et al (Nat. Genet. 2009 41:1122-6), Out et al (Hum. Mutat. 2009 30:1703-12) and Turner (Nat. Methods 2009 6:315-6), U.S. Patent Application Publication Nos. 20080160580 and 20080286795, U.S. Patent Nos. 6,306,597, 7,115,400, and 7,232,656, which are hereby incorporated by reference. 454 Life Science platforms include, e.g., the GS Flex and GS Junior, and are described in U.S. Patent No. 7,323,305, hereby incorporated by reference. Platforms for ion semiconductor sequencing include, e.g., the Ion Torrent Personal Genome Machine (PGM) and are described in U.S. Patent No. 7,948,015, hereby incorporated by reference. Platforms for pryosequencing include the GS Flex 454 system and are described in U.S. Patent Nos. 7,211,390; 7,244,559; 7,264,929, hereby incorporated by reference. Platforms for sequencing by ligation include, e.g., the SOLiD sequencing platform and are described in U.S. Patent No. 5,750,341, hereby incorporated by reference. Platforms for single-molecule sequencing include the SMRT system from Pacific Bioscience and the Helicos True Single Molecule Sequencing platform.

[0087] For example, a sequencing technique that can be used in one or more phasing methods described herein is the method commercialized by Illumina, as described, e.g., in U.S. Pat. Nos. 5,750,341; 6,306,597; and 5,969,119. Such method can employ cluster amplification of library members onto a flow cell and a sequencing-by-synthesis approach. In a round of sequencing, cluster-amplified library members can be subjected to repeated cycles of polymerase-directed single base extension. Single-base extension can involve incorporation of reversible-terminator dNTPs, each dNTP labeled with a different removable fluorophore. The reversible-terminator dNTPs can be 3' modified to prevent further extension by the polymerase. After incorporation of the dNTP, the incorporated nucleotide can be identified by fluorescence imaging. Following fluorescence imaging, the fluorophore can be removed and the 3' modification can be removed resulting in a 3' hydroxyl group, thereby allowing another cycle of single base extension.

[0088] In one embodiment of a phasing method employing Illumina sequencing methodology, a sequencing library is prepared from a nucleic acid sample obtained from a subject. In some cases, the sequencing library is a directional sequencing library. In some cases, the sequencing library is not a directional sequencing library. Single-stranded nucleic acids from the library can optionally be amplified, for example, by PCR, prior to coupling to the Illumina sequencing platform. To couple template nucleic acids to the Illumina platform, the nucleic acids can be denatured into single-stranded templates. The single-stranded templates can be randomly attached to the inside surface of flow-cell channels. In some

IDT Exhibit 1020.2341

embodiments the single-stranded templates are randomly attached to unique addressable locations of the flow-cell. Unlabeled nucleotides can be added to initiate solid-phase bridge amplification to produce dense clusters of clonally amplified templates. In some embodiments, each cluster is coupled to a unique addressable location of the flow-cell. To initiate a first round of sequencing, labeled reversible terminator dNTPs, a first sequencing primer, and DNA polymerase can be added to the Illumina flow cell. The clusters can be subjected to repeated cycles of polymerase-directed single base extension and label detection as described above, thereby producing a first sequencing read from one or more clusters. Upon completion of the first round of sequencing, the extended strands and sequencing primers can be denatured from the cluster templates and removed, e.g., by washing. To initiate a second round of sequencing, labeled reversible terminator dNTPS, a second sequencing primer, and DNA polymerase can be added to the Illumina flow cell. The clusters can then be subjected to repeated cycles of polymerase-directed single base extension and label detection as described above, thereby producing a second sequencing read from one or more clusters. Subsequent rounds of sequencing can optionally be performed by removing the extended strands and primers from the preceding round of sequencing, using subsequent primers that are designed to hybridize, for example, downstream or upstream of the preceding sequencing primer. Sequence reads for the first, second, and optionally additional rounds of sequencing can be compiled for each cluster. The compiling can be used to generate sequence information for the clonally amplified templates. Sequence information for the clonally amplified templates can be used to determine phasing of one or more alleles detected by the sequencing.

[0089] In some embodiments, one or more phasing methods described herein employ sequencing by ligation methods commercialized by Applied Biosystems (e.g., SOLiD sequencing). Library preparation for SOLiD sequencing can comprise ligation of adaptors attached to the 5' and 3' ends of the fragments to generate a fragment library. Alternatively, internal adaptors can be introduced by ligating adaptors to the 5' and 3' ends of the fragments, circularizing the fragments, digesting the circularized fragment to generate an internal adaptor, and attaching adaptors to the 5' and 3 ' ends of the resulting fragments to generate a mate-paired library. Next, nucleic acid templates can be clonally amplified on beads via PCR. Following PCR, the templates can be denatured. Beads can be enriched for beads with extended templates. The clonally amplified templates on the beads can be subjected to a 3' modification that permits bonding to a glass slide. The beads can be deposited on a glass slide for sequencing by ligation. In some embodiments, each bead is deposited onto a unique

IDT Exhibit 1020.2342

addressable location of the glass slide. A round of SOLiD sequencing can comprise the following steps: (1) hybridization of a first sequencing primer to the clonally amplified templates. The first sequencing primer can be designed to hybridize to a sequencing primer binding site on one or more adaptors attached to the template nucleic acids. (2) Following hybridization, a set of four fluorescently labeled di-base probes can compete for ligation to the sequencing primer. Specificity of the di-base probe can be achieved by interrogating every 1st and 2nd base in each ligation reaction. (3) Cycles of ligation, detection and cleavage can be performed with the number of cycles determining the eventual read length. (4) Following a series of ligation cycles, the extension product can be removed and the template can be reset with a primer complementary to the n-1 position for a second round of ligation cycles. For SOLiD sequencing, one round of sequencing can comprise five rounds of primer reset and ligation cycles as described above, wherein each subsequent primer hybridizes to an n-1 position of the adaptor sequencing.

[0090] In one embodiment of a phasing method described herein employing SOLiD sequencing methodology, a sequencing library is prepared from a nucleic acid sample obtained from a subject using a library preparation method for SOLiD sequencing as described herein. In some cases, the sequencing library is a directional sequencing library. In some cases, the sequencing library is not a directional sequencing library. The sequencing library members can be deposited onto unique addressable locations of the SOLiD platform as described herein. A first round of SOLiD sequencing can be performed as described herein. Following the first round of sequencing, the extension product can be removed and a second set of sequencing primers can be used for a second round of SOLiD sequencing. The second round of SOLiD sequencing can comprise multiple cycles of primer hybridization and ligation as described above. The second set of sequencing primers can comprise a primer "primer A" that is designed to hybridize to a genomic location residing 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, or more than 50 nucleotides upstream of an informative locus "A" as well as 2, 3, 4, 5, or more additional sequencing primers that are designed to hybridize to an n-1 genomic location, n-2 genomic location, n-3 genomic location, n-4 genomic location, n-5 genomic location, etc., of "Primer A" Following a second round of SOLiD sequencing, the extension product can be removed and a third set of sequencing primers can be used for a third round of SOLiD sequencing. The third set of sequencing primers can comprise a primer "primer B" that is designed to hybridize to a genomic location residing 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21,

IDT Exhibit 1020.2343

22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, or more than 50 nucleotides upstream of an informative locus "B"., as well as 2, 3, 4, 5, or more additional sequencing primers that are designed to hybridize to an n-1 location, n-2 location, n-3 location, n-4 location, n-5 location, etc., of "Primer B". The informative locus "B" can be located downstream or upstream of informative locus "A". The third round of SOLiD sequencing can comprise multiple cycles of primer hybridization and ligation as described above. Additional rounds of SOLiD sequencing can be performed to sequence further downstream informative loci, as described above. Sequence reads for the first, second, and optionally additional rounds of sequencing can be compiled for each addressable location. The compiling can be used to generate sequence information for the clonally amplified templates. Sequence information for the clonally amplified templates can be used to determine phasing of one or more alleles detected by the sequencing.

[0091] In other embodiments, one or more phasing methods disclosed herein employ sequencing by synthesis using the methods commercialized by 454/Roche Life Sciences, including but not limited to the methods and apparatus described, e.g., in Margulies et al., *Nature(2005)* 437:376-380 (2005); and U.S. Pat. Nos. 7,244,559; 7,335,762; 7,211,390; 7,244,567; 7,264,929; and 7,323,305, hereby incorporated by reference. 454 sequencing library preparation can comprise ligation of adaptors to the ends of nucleic acid fragments. The adaptors can serve as primers for amplification and can provide a sequencing primer binding site for sequencing-by-synthesis. One of the adaptors can comprise a capture reagent, e.g., a biotin. The nucleic acid fragments can be attached to capture beads, e.g., streptavidin-coated beads. The fragments attached to the beads can be PCR amplified within droplets of an oil-water emulsion, resulting in multiple copies of clonally amplified nucleic acid fragments on each bead. Following 454 library preparation, library members can be coupled to a 454 sequencing platform by capture of beads in wells of the 454 platform, which can be pico-liter sized. In some cases, a well captures one bead. Each well can accordingly be an addressable location of the 454 platform. Pyrosequencing can be performed on each DNA fragment in parallel. Pyrosequencing can comprise hybridization of a sequencing primer to a sequencing primer binding site on an adaptor, followed by extension of the primer in a pyrosequencing reaction. During the pyrosequencing reaction, dTNPs are added such that only one out of four of the possible A/T/C/G nucleotides are added and available at a time. Pyrosequencing can detect release of pyrophosphate (PPi) upon nucleotide incorporation. PPi can be converted to ATP by ATP sulfurylase in the presence of adenosine 5' phosphosulfate. Luciferase can use ATP to convert luciferin to oxyluciferin, thereby

IDT Exhibit 1020.2344

generating a light signal that is detected. A detected light signal can be used to identify the incorporated nucleotide. The intensity of the light signal can be used to determine how many bases are incorporated. Following detection of the light signal, unincorporated dNTPs can be removed and the next one out of four possible A/T/C/G nucleotides can be added.

[0092] In one embodiment of a phasing method employing pyrosequencing methodology (e.g., 454 sequencing), a sequencing library is prepared from a nucleic acid sample obtained from a subject. In some cases, the sequencing library is a directional sequencing library. In some cases, the sequencing library is not a directional sequencing library. Sequencing library members can be coupled to addressable locations of a pyrosequencing platform (e.g., a 454 sequencing platform) as described above. A first round of pyrosequencing can be performed as described above, generating a first sequencing read from the library members. Following the first round of pyrosequencing, the extension strands and primers are removed from the library members, e.g., by denaturation and washing. A second round of sequencing can then be performed by hybridization of a second sequencing primer to one or more library members. The second sequencing primer can be designed to hybridize to a location upstream of an informative locus "A" and to sequence informative locus "A". A second pyrosequencing reaction can then be performed as described above. Subsequent removal of the second sequencing primer and additional rounds of pyrosequencing can be performed to sequence additional informative loci. The additional informative loci can be located downstream or upstream of locus "A". Sequence reads for the first, second, and optionally additional rounds of sequencing can be compiled for each addressable location. The compiling can be used to generate sequence information for the clonally amplified templates. Sequence information for the clonally amplified templates can be used to determine phasing of two or more alleles detected by the sequencing.

[0093] In other embodiments, one or more phasing methods disclosed herein may employ the sequencing methods commercialized by Helicos BioSciences Corporation (Cambridge, Mass.) as described, e.g., in U.S. Patent Application Ser. No. 11/167,046, and U.S. Pat. Nos. 7,501,245; 7,491,498; 7,276,720; and in U.S. Patent Application Publication Nos. US20090061439; US20080087826; US20060286566; US20060024711; US20060024678; US20080213770; and US20080103058, hereby incorporated by reference. For example, a method can employ the true Single-Molecule Sequencing technology by Helicos Biosciences Corporation. tSMS library preparation can comprise ligation of a polyA adaptor to 3 ' end of DNA fragments. The adapted fragments can then be coupled to a Helicos tSMS sequencer by the following steps: (1) fragments are hybridized to poly-T oligonucleotides immobilized on

IDT Exhibit 1020.2345

the tSMS flow cell; (2) the flow cell is loaded into an instrument, e.g., HeliScope™ sequencer. Each fragment can be hybridized to a unique addressable location on the tSMS flow cell. A laser can illuminate the surface of the flow cell, revealing the position of each template. A CCD camera can map the position of the templates on the flow cell surface. tSMS sequencing can then proceed by subjecting library members to repeated cycles of polymerase-directed single base extension. A round of tSMS sequencing reaction can commence by introducing a DNA polymerase and a fluorescently labeled nucleotide. The polymerase can incorporate the labeled nucleotides to the primer in a template directed manner. The polymerase and unincorporated nucleotides can be removed. The templates that have directed incorporation of the fluorescently labeled nucleotide can be discerned by imaging the flow cell surface. After imaging, a cleavage step can remove the fluorescent label, and the process can be repeated with other fluorescently labeled nucleotides until a desired read length is achieved. Sequence information can be collected with each nucleotide addition step.

[0094] In one embodiment of a phasing method employing the Helicos tSMS methodology, a tSMS sequencing library is prepared as described above from a nucleic acid sample obtained from a subject. In some cases, the sequencing library is a directional sequencing library. In some cases, the sequencing library is not a directional sequencing library. The sequencing library members can be coupled to unique addressable locations on the tSMS platform as described herein, e.g., by hyridization to polyT oligonucleotides immobilized on the flow cell. Reverse complement complement strands of the hybridized library members can be synthesized, thus creating immobilized templates for sequencing. The original library members hybridized to the flowcell can be denatured and/or removed. A first round of tSMS sequencing can be performed on immobilized templates as described above, generating a first sequencing read from the immobilized templates. Following the first round of tSMS sequencing, the first sequencing reads can be denatured and/or removed. The immobilized templates can be retained on the flow cell. A second round of tSMS sequencing can then be performed by hybridization of a sequencing primer to the immobilized templates. The second sequencing primer can be designed to hybridize to a location upstream of an informative locus "A" and to sequence informative locus "A". A second tSMS sequencing reaction can then be performed as described above. Subsequent denaturation and/or removal of the sequencing primer and sequence reads, followed by additional rounds of tSMS sequencing can be performed on the immobilized templates to sequence additional informative loci. The additional informative loci can be located

IDT Exhibit 1020.2346

downstream or upstream of locus "A". Sequence reads for the first, second, and optionally additional rounds of sequencing can be compiled for each addressable location. The compiling can be used to generate sequence information for the templates. Sequence information for the templates can be used to determine phasing of one or more alleles detected by the sequencing.

[0095] Another example of a sequencing technique that can be used in the methods described herein is semiconductor sequencing provided by Ion Torrent (e.g., using the Ion Personal Genome Machine (PGM)). Ion Torrent technology can use a semiconductor chip with multiple layers, e.g., a layer with micro-machined wells, an ion-sensitive layer, and an ion sensor layer. Sequencing library preparation for the Ion Torrent platform can involves ligation of two distinct adaptors at both ends of a DNA fragment. Individual library members can be attached to beads, as described herein. The library members can then be clonally amplified by, e.g., emulsion PCR. The clonally amplified library members can be coupled to an Ion Torrent sequencing platform, for example, by introducing individual beads comprising the clonally amplified library members into wells of the Ion Torrent platform. Each well can be considered a unique addressable location of the Ion Torrent platform. A round of sequencing can comprise annealing a first sequencing primer to the library members. The first sequencing primer can be designed to hybridize to a sequencing primer binding region of an Ion Torrent library adaptor. To initiate sequencing of the nucleic acids on the beads, one type of deoxyribonucleotide (e.g., dATP, dCTP, dGTP, or dTTP) can be introduced into the wells. When one or more nucleotides are incorporated by DNA polymerase, protons (hydrogen ions) can be released in the well, which can be detected by the ion sensor. The semiconductor chip can then be washed and the process can be repeated with a different deoxyribonucleotide. A plurality of nucleic acids can be sequenced in the wells of a semiconductor chip. The semiconductor chip can comprise chemical-sensitive field effect transistor (chemFET) arrays to sequence DNA (for example, as described in U.S. Patent Application Publication No. 20090026082). Incorporation of one or more triphosphates into a new nucleic acid strand at the 3' end of the sequencing primer can be detected by a change in current by a chemFET. An array can have multiple chemFET sensors.

[0096] In one embodiment of a phasing method described herein employing Ion Torrent methodology, an Ion Torrent sequencing library is prepared as described above from a nucleic acid sample obtained from a subject. In some cases, the sequencing library is a directional sequencing library. In some cases, the sequencing library is not a directional sequencing library. Sequencing library members can be coupled to unique addressable

IDT Exhibit 1020.2347

locations of the Ion Torrent platform as described herein. A first round of Ion Torrent sequencing can be performed on library members as described above, generating a first sequencing read from the library members. Upon completion of the first round of sequencing, the extended strands and sequencing primers can be denatured from the templates and removed, e.g., by washing. A second round of sequencing can be performed using a second sequencing primer. The second sequencing primer can be designed to hybridize to a location upstream of an informative locus "A" and to sequence informative locus "A". A second sequencing reaction can then be performed as described above. Subsequent removal of the second sequencing primer and additional rounds of sequencing can be performed to sequence additional informative loci. The additional informative loci can be located downstream or upstream of locus "A". Sequence reads for the first, second, and optionally additional rounds of sequencing can be compiled for each addressable location. The compiling can be used to generate sequence information for the clonally amplified templates. Sequence information for the clonally amplified templates can be used to determine phasing of one or more alleles detected by the sequencing.

[0097] Sequential sequencing by any of the means described herein can produce one or more sequencing reads for nucleic acid templates. The sequencing reads can have a length which may vary depending on the particular sequencing technology utilized. NGS platforms can provide sequence reads that vary in size from 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, tens to hundreds, or thousands of base pairs. In some embodiments of the method described herein, the sequence reads are about 1 base long, about 2 bases long, about 3 bases long, about 4 bases long, about 5 bases long, about 6 bases long, about 7 bases long, about 8 bases long, about 9 bases long, about 10 bases long, about 12 bases long, about 14 bases long, about 16 bases long, about 18 bases long, about 20 bases long, about 25 bases long, about 30 bases long, about 35 bases long, about 40 bases long, about 45 bases long, about 50 bases long, about 55 bases long, about 60 bases long, about 65 bases long, about 70 bases long, about 75 bases long, about 80 bases long, about 85 bases long, about 90 bases long, about 95 bases long, about 100 bases long, about 110 bases long, about 120 bases long, about 130, about 140 bases long, about 150 bases long, about 200 bases long, about 250 bases long, about 300 bases long, about 350 bases long, about 400 bases long, about 450 bases long, about 500 bases long, about 600 bases long, about 700 bases long, about 800 bases long, about 900 bases long, about 1000 bases, about 2000 bases, about 3000 bases, about 4000 bases, about 5000 bases, about 6000 bases, about 7000 bases, about 8000 bases, about 9000 bases, about 10,000 bases long, about 20,000 bases long, about 30,000 bases long, about 40,000 bases long, about

IDT Exhibit 1020.2348

50,000 bases long, about 60,000 bases long, about 70,000 bases long, about 80,000 bases long, about 90,000 bases long, about 100,000 bases, about 150,000 bases long, about 200,000 bases long, about 250,000 bases long, or more than 250,000 bases long.

[0098] In some cases, a sequencing read is at least 1 base long, at least 2 bases long, at least 3 bases long, at least 4 bases long, at least 5 bases long, at least 6 bases long, at least 7 bases long, at least 8 bases long, at least 9 bases long, at least 10 bases long, at least 12 bases long, at least 14 bases long, at least 16 bases long, at least 18 bases long, at least 20 bases long, at least 25 bases long, at least 30 bases long, at least 35 bases long, at least 40 bases long, at least 45 bases long, at least 50 bases long, at least 55 bases long, at least 60 bases long, at least 65 bases long, at least 70 bases long, at least 75 bases long, at least 80 bases long, at least 85 bases long, at least 90 bases long, at least 95 bases long, at least 100 bases long, at least 110 bases long, at least 120 bases long, at least 130 bases long, at least 140 bases long, at least 150 bases long, at least 200 bases long, at least 250 bases long, at least 300 bases long, at least 350 bases long, at least 400 bases long, at least 450 bases long, at least 500 bases long, at least 600 bases long, at least 700 bases long, at least 800 bases long, at least 900 bases long, at least 1000 bases long, at least 10,000 bases long, at least 20,000 bases long, at least 30,000 bases long, at least 40,000 bases long, at least 50,000 bases long, at least 60,000 bases long, at least 70,000 bases long, at least 80,000 bases long, at least 90,000 bases long, at least 100,000 bases, at least 150,000 bases long, at least 200,000 bases long, or more than 200,000 bases long.

[0099] A sequencing read can be between about 1 to about 10 bases long, between about 5 to about 50 bases long, between about 20 to about 100 bases long, between about 50 to about 200 bases long, between about 100 to about 500 bases long, between about 200 to about 1000 bases long, between about 500 to about 2000 bases long, between about 1000 to about 5000 bases long, between about 2000 to about 10000 bases long, between about 5000 to about 50000 bases long, abetween about 10000 to about 100000 bases long, or more than 100000 bases long.

Sequence analysis

[00100] Sequencing reads (e.g., sequence tags) can be aligned to assemble a partial or complete map of the subject's genome. Sequence tags can be aligned (e.g., mapped), for example, to a reference genome. The reference genome can be a human reference genome. The human reference genome can be the NCBI36/hg18 sequence, available on the world wide web at genome.ucsc.edu/cgi-bin/hgGateway?org=Human&db=hgl 8&hgsid=166260105). Sequencing reads can be mapped using an alternative source of

IDT Exhibit 1020.2349

public sequence information. Other sources of public sequence information include GenBank, dbEST, dbSTS, EMBL (the European Molecular Biology Laboratory), and the DDBJ (the DNA Databank of Japan). In some cases, the reference genome can be an artificial target sequence genome comprising polymorphic target sequences.

[00101] Mapping of the sequence tags can be achieved, for example, by aligning the sequence of the tag to a sequence of the reference genome. Such methods can be used to map a sequencing read to a genomic region, a chromosomal region, or a chromosome. A number of computer algorithms can be used for aligning sequences. Such algorithms include, without limitation, BLAST (Altschul et al., 1990), BLITZ (MPsrch) (Sturrock & Collins, 1993), BOWTIE (Langmead et al, Genome Biology 10:R25.1-R25.10 [2009]), FASTA (Person & Lipman, 1988), SAMtools (SAMtools, Bioinformatics, 2009, 25(16):2078-9), and ELAND (Illumina, Inc., San Diego, CA, USA). In one embodiment, sequence reads produced by Illumina sequencing are processed by bioinformatic alignment analysis for the Illumina Genome Analyzer, which uses the Efficient Large-Scale Alignment of Nucleotide Databases (ELAND) software. Additional software includes the Burroughs-Wheeler block sorting compression procedure which involves block sorting or preprocessing to make compression more efficient.

[00102] Reads from the sequential rounds of sequencing can be compiled for each sequenced template to assemble complete or partial sequence information of the template. Reads from the sequential rounds of sequencing can be compiled according to unique addressable locations of the sequencing platform. For example, all reads generated from a unique addressable location of a sequencing platform can be compiled into a data file. Reads can be compiled with aid of a computer processor. Compiled reads from a unique addressable location can be used to determine partial or complete sequence information for a nucleic acid template (either single template or clonally amplified template). The partial or complete sequence information can comprise information over a longer distance than an average read length of the sequencing platform. Because the sequence information is over a longer distance than an average read length of the sequencing platform, the probability that the sequence information comprises information over two or more informative loci can be increased.

[00103] Compiled reads from a unique addressable location may comprise sequence information for two or more detected alleles. Such compiled reads can be used to determine that the two or more detected alleles are in phase, since the detected alleles are located on the sample template or clonally amplified template. Alleles detected across two or more

IDT Exhibit 1020.2350

addressable locations can also be determined to be in phase, if the two or more addressable locations each share at least one detected allele in common. Alleles detected across a set of addressable locations comprising any number of addressable locations can be determined to be in phase, if the addressable locations in the set share at least one detected allele in common with at least one other addressable location in the set.

[00104] Compiling of reads, determination of phasing information, and/or haplotype analysis can be performed with the aid of a computer system. In some embodiments, the computer system executes instructions contained in a computer-readable medium. The computer system can comprise a processor. The processor can be associated with one or more controllers, calculation units, and/or other units of a computer system, or implanted in firmware. In some embodiments, one or more steps of the method are implemented in hardware. In some embodiments, one or more steps of the method are implemented in software. Software routines may be stored in any computer readable memory unit such as flash memory, RAM, ROM, magnetic disk, laser disk, or other storage medium as described herein or known in the art. Software may be communicated to a computing device by any known communication method including, for example, over a communication channel such as a telephone line, the internet, a wireless connection, or by a transportable medium, such as a computer readable disk, flash drive, etc. The one or more steps of the methods described herein may be implemented as various operations, tools, blocks, modules and techniques which, in turn, may be implemented in firmware, hardware, software, or any combination of firmware, hardware, and software. When implemented in hardware, some or all of the blocks, operations, techniques, etc. may be implemented in, for example, an application specific integrated circuit (ASIC), custom integrated circuit (IC), field programmable logic array (FPGA), or programmable logic array (PLA).

[00105] **FIG. 5** depicts an exemplary computer system adapted to implement one or more methods provided herein. The system **500** includes a central computer server **501** that is programmed to implement exemplary methods described herein. The server **501** includes a central processing unit (CPU, also "processor") **505** which can be a single core processor, a multi core processor, or plurality of processors for parallel processing. The server **501** also includes memory **510** (e.g. random access memory, read-only memory, flash memory); electronic storage unit **515** (e.g. hard disk); communications interface **520** (e.g. network adaptor) for communicating with one or more other systems; and peripheral devices **525** which may include cache, other memory, data storage, and/or electronic display adaptors. The memory **510**, storage unit **515**, interface **520**, and peripheral devices **525** are in

IDT Exhibit 1020.2351

communication with the processor **505** through a communications bus (solid lines), such as a motherboard. The storage unit **515** can be a data storage unit for storing data. The server **501** is operatively coupled to a computer network ("network") **530** with the aid of the communications interface **520**. The network **530** can be the Internet, an intranet and/or an extranet, an intranet and/or extranet that is in communication with the Internet, a telecommunication or data network. The network **530** in some cases, with the aid of the server **501**, can implement a peer-to-peer network, which may enable devices coupled to the server **501** to behave as a client or a server.

[00106] The storage unit **515** can store files, such as subject reports, and/or communications with the caregiver, sequencing data, data about individuals, or any aspect of data generated by one or more phasing methods disclosed herein.

[00107] The server can communicate with one or more remote computer systems through the network **530**. The one or more remote computer systems may be, for example, personal computers, laptops, tablets, telephones, Smart phones, or personal digital assistants.

[00108] In some situations the system **500** includes a single server **501**. In other situations, the system includes multiple servers in communication with one another through an intranet, extranet and/or the Internet.

[00109] The server **501** can be adapted to store sequencing information, and/or information on a subject, such as, for example, polymorphisms, mutations, patient history of the subject, demographic data and/or other information of potential relevance. Such information can be stored on the storage unit **515** or the server **501** and such data can be transmitted through a network.

[00110] Methods as described herein can be implemented by way of machine (or computer processor) executable code (or software) stored on an electronic storage location of the server **501**, such as, for example, on the memory **510**, or electronic storage unit **515**. During use, the code can be executed by the processor **505**. In some cases, the code can be retrieved from the storage unit **515** and stored on the memory **510** for ready access by the processor **505**. In some situations, the electronic storage unit **515** can be precluded, and machine-executable instructions are stored on memory **510**. Alternatively, the code can be executed on a second computer system **540**.

[00111] Aspects of the systems and methods provided herein, such as the server **501**, can be embodied in programming. Various aspects of the technology may be thought of as "products" or "articles of manufacture" typically in the form of machine (or processor) executable code and/or associated data that is carried on or embodied in a type of machine

IDT Exhibit 1020.2352

readable medium. Machine-executable code can be stored on an electronic storage unit, such memory (e.g., read-only memory, random-access memory, flash memory) or a hard disk. "Storage" type media can include any or all of the tangible memory of the computers, processors or the like, or associated modules thereof, such as various semiconductor memories, tape drives, disk drives and the like, which may provide non-transitory storage at any time for the software programming. All or portions of the software may at times be communicated through the Internet or various other telecommunication networks. Such communications, for example, may enable loading of the software from one computer or processor into another, for example, from a management server or host computer into the computer platform of an application server. Thus, another type of media that may bear the software elements includes optical, electrical, and electromagnetic waves, such as used across physical interfaces between local devices, through wired and optical landline networks and over various air-links. The physical elements that carry such waves, such as wired or wireless likes, optical links, or the like, also may be considered as media bearing the software. As used herein, unless restricted to non-transitory, tangible "storage" media, terms such as computer or machine "readable medium" can refer to any medium that participates in providing instructions to a processor for execution.

[00112] Hence, a machine readable medium, such as computer-executable code, may take many forms, including but not limited to, tangible storage medium, a carrier wave medium, or physical transmission medium. Non-volatile storage media can include, for example, optical or magnetic disks, such as any of the storage devices in any computer(s) or the like, such may be used to implement the system. Tangible transmission media can include: coaxial cables, copper wires, and fiber optics (including the wires that comprise a bus within a computer system). Carrier-wave transmission media may take the form of electric or electromagnetic signals, or acoustic or light waves such as those generated during radio frequency (RF) and infrared (IR) data communications. Common forms of computer-readable media therefore include, for example: a floppy disk, a flexible disk, hard disk, magnetic tape, any other magnetic medium, a CD-ROM, DVD, DVD-ROM, any other optical medium, punch cards, paper tame, any other physical storage medium with patterns of holes, a RAM, a ROM, a PROM and EPROM, a FLASH-EPROM, any other memory chip or cartridge, a carrier wave transporting data or instructions, cables, or links transporting such carrier wave, or any other medium from which a computer may read programming code and/or data. Many of these forms of computer readable media may be involved in carrying one or more sequences of one or more instructions to a processor for execution.

IDT Exhibit 1020.2353

[00113] The computer readable medium can comprise computer executable code. The computer executable code can implement a method comprising: (a) accessing a data file in computer memory that comprises a plurality of sequencing reads corresponding to a first addressable location of a sequencing platform, wherein the sequencing reads are generating by subjecting a first nucleic acid template coupled to the addressable location to a first and at least a second round of sequencing, wherein the first and at least second rounds of sequencing are performed using a first and second sequencing primer, respectively; (b) using said data file, detecting a first detected allele at a first locus of the nucleic acid template and detecting a second detected allele at a second locus of the nucleic acid template; and (c) determining the first and second detected allele to be in phase based upon said detecting. The method executed by the computer executable code can further comprise the steps of (d) accessing a second data file in computer memory that comprises a plurality of sequencing reads corresponding to a second addressable location of a sequencing platform, wherein the sequencing reads are generating by subjecting a second nucleic acid template coupled to the addressable location to the first and the at least second round of sequencing; (e) using the second data file, detecting a first detected allele at a first locus of the second nucleic acid template and detecting a second detected allele at a second locus of the second nucleic acid template; and (f) determining the first and second detected alleles of the second nucleic acid template to be in phase with the first and second detected alleles of the first nucleic acid template if the first or second detected allele of the second nucleic acid template is shared in common with the first nucleic acid template.

Kits

[00114] Any of the compositions described herein may be comprised in a kit. In a non-limiting example, the kit, in a suitable container, comprises: an adaptor or several adaptors, one or more of oligonucleotide primers and reagents for ligation, primer extension and amplification. The kit may also comprise means for purification, such as a bead suspension, and nucleic acid modifying enzymes. Kits can further comprise instructions for practicing any of the phasing methods described herein.

Products

[00115] Products based on the methods provided herein may be commercialized by the Applicants under the Ovation® family. Ovation is a registered trademark of NuGEN Technologies, Inc.

[00116] The information gained from the phasing of alleles in a genome as provided by practicing the methods described herein can be of great utility in research and discovery as

IDT Exhibit 1020.2354

well as for clinical, diagnostic, forensic, archeological, geneological, and/or epidemiological applications. For example, haplotype information can be used to aid in the diagnosis, prognosis, and/or theranosis of a disease. For other example, haplotype information of a potential tissue or organ donor and organ recipient can aid in the assessment of HLA compatibilities of the donor and recipient for decreasing transplant rejection. In some cases, information gained from phasing of alleles in a genome of a subject can be used to tailor a healthcare or therapeutic regimen for the subject. In some cases, one or more methods described herein can be used to construct a database of haplotypes associated with one or more disease or biological trait. A haplotype obtained from a subject using a method as described herein can be compared against a haplotype database, thereby allowing for diagnosis and/or prognosis of a disease or condition. In some cases, haplotype information can be used to aid in the treatment of an infectious disease. For example, bacteria or virus infections can be associated with a plurality of drug resistant markers. Use of one or more phasing methods described herein can help determine if the plurality of drug resistant markers are due to co-infection with single-drug resistant strains or due to infection with a single strain that is multi-drug resistant. One or more phasing methods described herein can also be used to characterize T-cell receptor and/or immunoglobulin repertoire in a subject. For example, one or more phasing methods described herein can be used to assess V(D)J recombination in a subject. One or more phasing methods described herein can be used for diagnosing, predicting, determining or assessing the genetic characteristics of a fetus or embryo. A phasing method described herein can be used to assess genetic phasing within the genome of a developing fetus. In some cases, one or more phasing methods can be used to combat bioterrorism and/or to assess an epidemic or pandemic (e.g., by characterizing or identifying an infectious agent such as a virus or bacterium). In some cases, one or more phasing methods can be used for paternity testing, to establish a family tree, or to determine ancestry. In some cases, one or more phasing methods can be used to determine alternative splicing of RNA transcripts.

EXAMPLES

Example 1 — Characterization of the human oral microbiome by sequential sequencing of bacterial 16S ribosomal operon

[00117] This example describes the characterization of the human oral microbiome by sequencing of the 16S rRNA gene sequences of a number of related bacterial organisms. 16S rRNA gene sequences contain species-specific hypervariable regions that can provide means

IDT Exhibit 1020.2355

for bacterial identification. An exemplary embodiment of 16S rRNA sequential sequencing, described below, is depicted in **FIG. 6**.

Sample nucleic acid

[00118] Microbial genomic DNA is isolated from human saliva using the OMNIgene-DISCOVER sample collection kit (DNA Genotek) according to the manufacturer's instructions. The different line shadings and hash marks indicate different microbial genomes (microbial DNA). The relative numbers (e.g., 4 hatched lines vs 1 gray line) reflect the different ratios of the organisms in the saliva sample. Extracted DNA is then fragmented via sonication to an average length of 1000 bp and purified using Agencourt AMPure XP beads (Beckman Coulter Genomics). In some cases, extracted DNA is not subjected to fragmentation. For example, the microbial genetic DNA can be optionally enriched for target polynucleotides comprising the 16S genomic regions using primers to conserved 16S regions of the genome, thereby generating 16S amplicons.

Generation 16 S libraries with ligated adapters

[00119] The NuGEN Ovation Ultralow Library System (NuGEN Technologies) is used to attach library adaptors to the DNA fragments, thereby generating next generation sequencing libraries from 100 ng of the purified sample according to manufacturer's instructions. Ligation products with a minimum length of 200 bp in length are purified by selective binding to Agencourt AMPure XP beads.

Sequential primer sequencing

[00120] Library members are clonally amplified by bridge amplification on an Illumina flowcell, thereby generating spatially separated clusters, wherein the clusters comprise identical copies of an original sequencing library member. In a first round of sequencing, all clusters, including those containing 16 S ribosomal DNA fragments, are sequenced by Illumina sequencing system using an Illumina standard forward primer. The standard forward primer can be designed to hybridize to a sequencing primer binding site of a library adaptor. Alternatively, a custom primer may be used. The custom primer can be designed to hybridize to a location upstream of and to sequence an informative locus. In such cases, clusters comprising a primer binding site for the first custom primer can be sequenced. The first round of sequencing can generate a first sequencing read ("Read 1"). Following the first sequencing read, the DNA is denatured to wash away the newly synthesized strand comprising Read 1. In cases where the first sequencing primer is a standard forward primer, Read 1 can contain sequence information for the forward end of the cluster templates. In cases where the first sequencing primer is a custom primer, Read 1 can comprise sequence

IDT Exhibit 1020.2356

information for the informative locus. A second round of sequencing can then be performed using a second sequencing primer. The second sequencing primer can be a custom primer that is designed to hybridize to a location upstream of and to sequence an informative locus. In cases where the first round of sequencing is performed using a custom primer that sequences an informative locus, the second sequencing primer is a designed to hybridize to a location upstream of and to sequence informative locus which is downstream of the first informative locus. The second round of sequencing can produce a second sequencing read "Read 2" in clusters comprising a primer binding site for the second sequencing primer. Read 2 can comprise sequence information for the informative locus sequenced by the second sequencing primer. Successive rounds of denaturation, re-priming and sequencing are optionally performed with primers that are designed to sequence additional informative loci. Sequence reads from successive priming and sequencing are compiled for each cluster and aligned to map reads originating from the same nucleic acid fragments.

Example 2 — Genomic DNA sequencing — distinguishing between a gene and its pseudogene using sequential sequencing

[00121] FIG. 7 depicts an exemplary application of a method described herein to distinguish a gene from a pseudogene in a gene/pseudogene pair. Genomic DNA sequencing libraries are prepared using the NuGEN's Encore system. These libraries are sequenced on a DNA sequencing system such as those made by Illumina or Ion Torrent. Clusters include, but are not limited to cluster 1, from a different genomic region from a gene/pseudogene pair, cluster 2, which comprises a portion of the gene of a gene/psuedogene pair, and cluster 3, which comprises a portion of the pseudogene of the gene/pseudogene pair. A first round of sequencing is performed using a first sequencing primer, thereby generating a first sequencing read. In some cases, the first sequencing primer is a standard forward primer as described in Example 1. The first sequencing read reveals that clusters 2 and 3 exhibit different alleles at a certain locus (a T in cluster 2 and a C in cluster 3). However, it is impossible to determine from read 1 which cluster represents the gene or the pseudogene. Following a first sequencing read, the DNA is denatured to wash away the extension strand. An oligonucleotide designed to anneal to a common sequence shared by the gene and pseudogene is injected into the sequencer to act as the second sequencing primer. The second sequencing primer can be designed to hybridize to a conserved sequence shared by the gene and pseudogene, wherein extension of the primer produces a second sequencing read that extends across a locus that can distinguish the gene and pseudogene. An exemplary gene/pseudogene pair which can be distinguished using a method described herein is the

IDT Exhibit 1020.2357

SMN1 gene/SMN2 pseudogene pair. In some embodiments, the second sequencing primer comprises a sequence which is designed to hybridize to a conserved squence shared by SMN1 and SMN2, wherein extension of the primer produces a second sequencing read that extends across a locus that can distinguish SMN1 and SMN2. In one embodiment, the second sequencing primer comprises the sequence CTTCCTTTATTTTCCTTACAGGGT (SEQ. ID. NO: 1). The second sequencing read reveals a downstream C base call at the locus in cluster 3. The C base call reveals that cluster 3 maps to the pseudogene but that cluster 2 maps to the gene. An oligonucleotide set may include primers that will sequence through one of the nucleotide differences between the gene and pseudogene as well as primers that will generate sequence to read nucleotide differences, and therefore determine whether a sequencing read is from a gene or its pseudogene. A combination of such primers will allow multiple gene/pseudogene pairs across the genome to be analyzed simultaneously for genetic mutations.

Example 3 — Targeted DNA sequencing library

[00122] A targeted DNA sequencing library is made using a target-selective library preparation kit from NuGEN, Agilent, Illumina, or Nimblegen, described herein. These libraries are sequenced on a DNA sequencing system such as those made by Illumina or Ion Torrent. Following a first sequencing read, the DNA is denatured to wash away the extension strand. A pool of primers that hybridize to common sequences in gene/pseudogene pairs are injected into the sequencer to act as a priming site for a second sequencing read. A primer set may include primers that will sequence through one of the nucleotide differences between a gene and its pseudogene (e.g., SMN1 and SMN2) as well as primers that will generate sequence to read nucleotide differences, and therefore determine whether a sequencing read is, for example, from a gene or pseudogene. A combination of such primers will allow multiple gene/pseudogene pairs across the genome to be analyzed simultaneously for genetic mutations. This type of technology is useful for genetic diagnostics.

Example 4 — RNA-sequencing library

[00123] An RNA sequencing library is made from NuGEN's Encore Complete RNA-Seq Library System. The library is sequenced on an Illumina DNA sequencer. Following the first sequencing read, a pool of primers that will hybridize to specific exons of interest is injected into the sequencing machine. These primers are used to generate a second sequencing read in a downstream exon. The second, targeted sequencing read provides information about which exons have been spliced together to generate a particular RNA transcript.

IDT Exhibit 1020.2358

[00124] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments provided herein described herein may be employed in practicing the invention. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

IDT Exhibit 1020.2359

CLAIMS

WHAT IS CLAIMED IS:

1. A method for relating at least two nucleic acid sequences or regions of sequence variation to the same nucleic acid template, the method comprising:

a) creating a nucleic acid library wherein a template is immobilized on a solid support;

b) generating a first sequencing read from the immobilized template with a first oligonucleotide primer, wherein a 3' terminal nucleotide of the first oligonucleotide primer binds to a first binding site in the template;

c) denaturing the first sequencing read from the immobilized template;

d) annealing a second oligonucleotide primer that is complementary to a region or regions within the immobilized template, wherein a 3' terminal nucleotide of the second oligonucleotide primer binds to a second binding site in the template that is more than 5 nt away from the first binding site in the template;

e) generating a second sequencing read with the second oligonucleotide primer; and

f) compiling data from the first sequencing read and the second sequencing read to map reads originating from the template.

2. The method of claim 1, wherein the nucleic acid libraries are generated from amplicons.

3. The method of claim 2, wherein the amplicons comprise conserved regions which flank a variable region.

4. The method of claim 3, wherein alignment of multiple variable regions enables differentiation and/or typing of related transcripts.

5. The method of claim 3, wherein alignment of multiple variable regions enables differentiation and/or typing of related micro-organisms.

6. The method of claim 1, wherein libraries are of reduced complexity.

7. The method of claim 6, wherein reduced complexity is achieved by target enrichment.

8. The method of any of the preceding claims, wherein the nucleic acid library comprises closely related nucleic acid sequences as inserts.

IDT Exhibit 1020.2360

9. A method, comprising:

a) coupling at least one nucleic acid template to a sequencing platform;

b) subjecting the at least one nucleic acid template to a first round of sequencing using a first sequencing primer, wherein the first sequencing primer comprises a 3' terminal nucleotide which anneals to a first binding site in the at least one nucleic acid template, wherein the first round of sequencing produces a first sequencing read of the at least one nucleic acid template;

c) removing the first annealed sequencing primer and the first sequencing read from the at least one nucleic acid template; and

d) subjecting the at least one nucleic acid template to a second round of sequencing using a second sequencing primer, wherein the second sequencing primer comprises a 3' terminal nucleotide which anneals to a second binding site in the at least one nucleic acid template which is more than 5 nt away from the first binding site, wherein the second round of sequencing produces a second sequencing read of the at least one nucleic acid template.

10. The method of claim 9, wherein the first and/or the second round of sequencing comprises sequencing by synthesis.

11. The method of claim 9, wherein the first and/or the second round of sequencing comprises sequencing by ligation.

12. The method of claim 9, wherein the coupling comprises conjugating the at least one nucleic acid template to a solid support of the sequencing platform.

13. The method of claim 12, wherein the solid support comprises a bead, a well, an array, or a flow cell.

14. The method of claim 9, wherein the at least one nucleic acid template comprises an adaptor oligonucleotide ligated to one or both ends of the at least one nucleic acid template.

15. The method of claim 14, wherein the first sequencing primer is selectively hybridizable to a primer binding site of the adaptor oligonucleotide.

16. The method of claim 14, wherein the second sequencing primer is not selectively hybridizable to the primer binding site of the adaptor oligonucleotide.

17. The method of claim 9, wherein the second sequencing read comprises an informative locus.

18. The method of claim 17, wherein the informative locus is a polymorphic locus.

IDT Exhibit 1020.2361

19. The method of claim 17, wherein the first sequencing read comprises a first informative locus and wherein the second sequencing read comprises a second informative locus.

20. The method of claim 19, comprising determining that a first allele detected at the first informative locus and a second allele detected at the second informative locus are in phase.

21. The method of claim 17, wherein the second sequencing primer is designed to hybridize to a location that is within *x* nt upstream of the informative locus, wherein *x* is a number that is less than an average sequencing read length produced by the first or second round of sequencing.

22. The method of claim 9, wherein the removing comprises a denaturing step.

23. The method of claim 22, wherein the removing further comprises a washing step.

24. The method of claim 9, further comprising subjecting the at least one nucleic acid template to one or more additional rounds of sequencing, wherein the one or more additional rounds of sequencing comprise the steps of : (i) removing a previously annealed sequencing primer and previously generated sequencing read from the at least one nucleic acid template; (ii) annealing an additional sequencing primer to the at least one nucleic acid template; and (iii) subjecting the at least one nucleic acid template to an additional round of sequencing using the additional sequencing primer, thereby producing an additional sequencing read of the at least one nucleic acid template.

25. The method of claim 24, wherein the additional sequencing primer comprises a 3' terminal nucleotide which hybridizes to a binding site that is more than 5 nt away from a binding site of the 3' terminal nucleotide of the previously annealed sequencing primer.

26. The method of claim 24, wherein a first allele detected by the first sequencing read, a second allele detected by the second sequencing read, and an additional allele detected by the additional sequencing read are determined to be in phase.

27. The method of claim 24, wherein the additional sequencing primer and the previously annealed sequencing primer are designed to hybridize to genomic locations that map to a single chromosome.

28. The method of claim 27, wherein the genomic locations that map to a single chromosome span a distance that is less than a length of the at least one nucleic acid template.

29. The method of claim 9, wherein the at least one nucleic acid template is a member of a sequencing library.

IDT Exhibit 1020.2362

30. The method of claim 29, wherein the sequencing library is a whole-genome sequencing library.

31. The method of claim 29, wherein the sequencing library is a target-enriched sequencing library.

32. The method of claim 29, wherein the sequencing library comprises nucleic acid fragments of an average length.

33. The method of claim 32, wherein the average length is about 50 nucleotides to about 10 kb.

34. The method of any of claims 9-33, wherein the first, second, and optionally additional sequencing reads comprise sequence information from a plurality of loci.

35. The method of claim 34, wherein phasing of alleles detected at the plurality of loci is determined based upon the first, second, and optionally additional sequencing reads.

36. The method of any of claims 9-33, wherein the at least one nucleic acid template comprises a plurality of nucleic acid templates.

37. The method of any of claims 9-33, wherein the at least one nucleic acid template comprises at least 10,000 nucleic acid templates.

38. The method of claim 37, wherein the at least one nucleic acid template comprises 3 billion nucleic acid templates.

39. The method of any of claims 9-33, wherein the at least one nucleic acid template is coupled to a unique addressable location of the sequencing platform.

40. The method of claim 39, comprising compiling sequencing reads from the unique addressable location into one data file.

41. The method of claim 39, comprising determining phasing of a plurality of alleles detected across two or more unique addressable locations.

42. The method of claim 41, wherein the two or more unique addressable locations each comprise two or more detected alleles, wherein the two or more detected alleles of each of the two or more unique addressable locations are determined to be in phase if at least one common detected allele is shared by the two or more unique addressable locations.

43. The method of any of claims 9-33, wherein step (d) comprises subjecting the at least one nucleic acid template to a second round of sequencing using a mixture of second sequencing primers.

IDT Exhibit 1020.2363

44. The method of claim 43, wherein the mixture of second sequencing primers comprises a pair of primers, wherein each primer in the primer pair is designed to sequence a forward or reverse complement strand containing a locus.

45. The method of claim 43, wherein primers within the mixture are designed and/or selected to have minimal potential to hybridize to the same nucleic acid template.

46. A kit, comprising:

a) one or more oligonucleotide primers for performing a sequencing reaction; and

b) instructions for practicing a method of any of the preceding claims using the one or more oligonucleotide primers.

47. A computer readable medium comprising computer executable code that, upon execution by a computer processor, implements a method comprising:

a) accessing a data file in computer memory that comprises a plurality of sequencing reads corresponding to a first addressable location of a sequencing platform, wherein the sequencing reads are generating by subjecting a first nucleic acid template coupled to the addressable location to a first and at least a second round of sequencing, wherein the first and at least second rounds of sequencing are performed using a first and second sequencing primer, respectively;

b) using the data file, detecting a first detected allele at a first locus of the nucleic acid template and detecting a second detected allele at a second locus of the nucleic acid template; and

c) determining the first and second detected allele to be in phase based upon the detecting.

48. The computer readable medium of claim 47, wherein the method further comprises:

d) accessing a second data file in computer memory that comprises a plurality of sequencing reads corresponding to a second addressable location of a sequencing platform, wherein the sequencing reads are generated by subjecting a second nucleic acid template coupled to the addressable location to the first and the at least second round of sequencing;

e) using the second data file, detecting a first detected allele at a first locus of the second nucleic acid template and detecting a second detected allele at a second locus of the second nucleic acid template; and

IDT Exhibit 1020.2364

f) determining the first and second detected alleles of the second nucleic acid template to be in phase with the first and second detected alleles of the first nucleic acid template if the first or second detected allele of the second nucleic acid template is shared in common with the first nucleic acid template.

IDT Exhibit 1020.2365

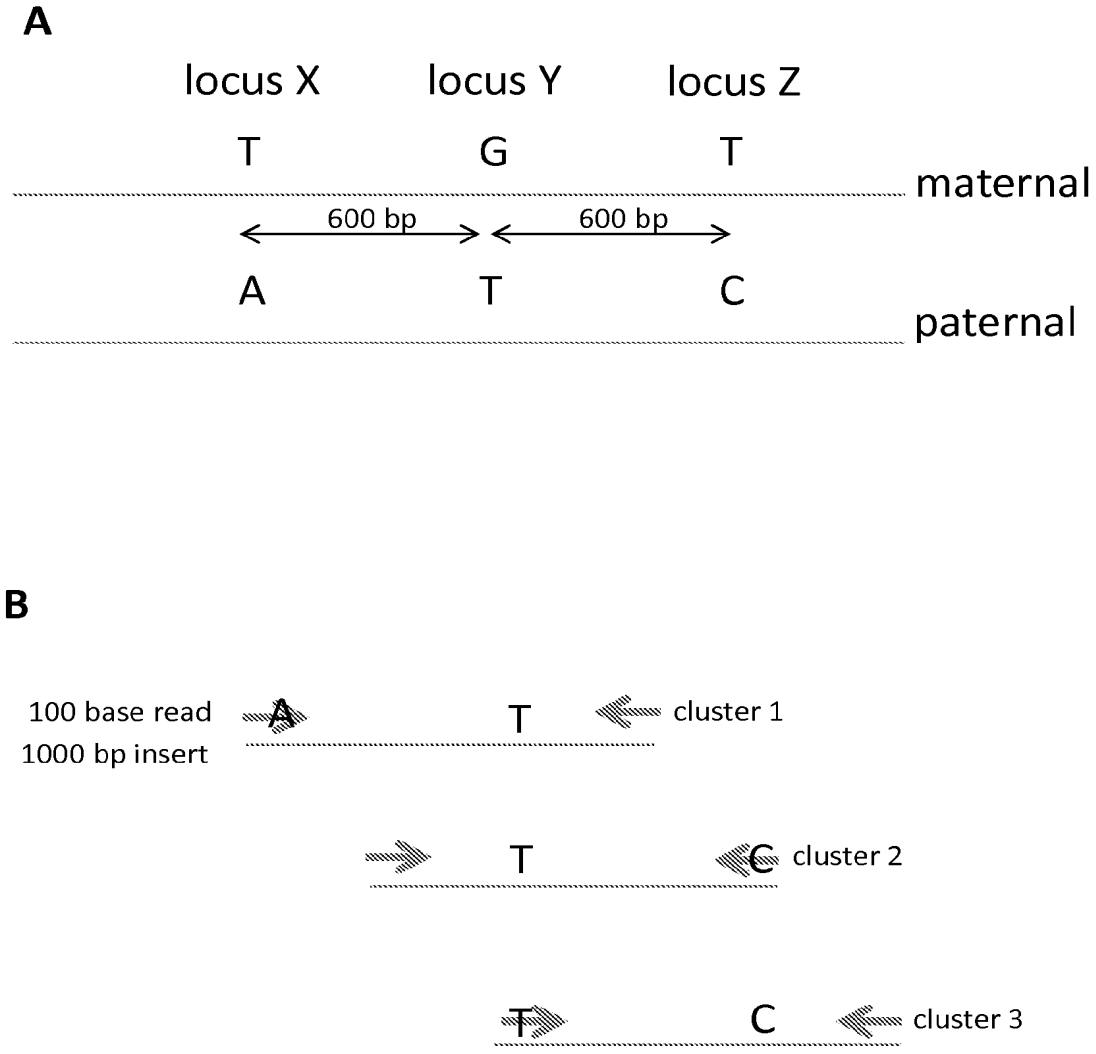

A
locus X
locus Y
locus Z
T
G
T
maternal
600 bp
600 bp
A
T
C
paternal
B
100 base read
1000 bp insert
A
T
cluster 1
T
C
cluster 2
T
C
cluster 3

FIG. 1

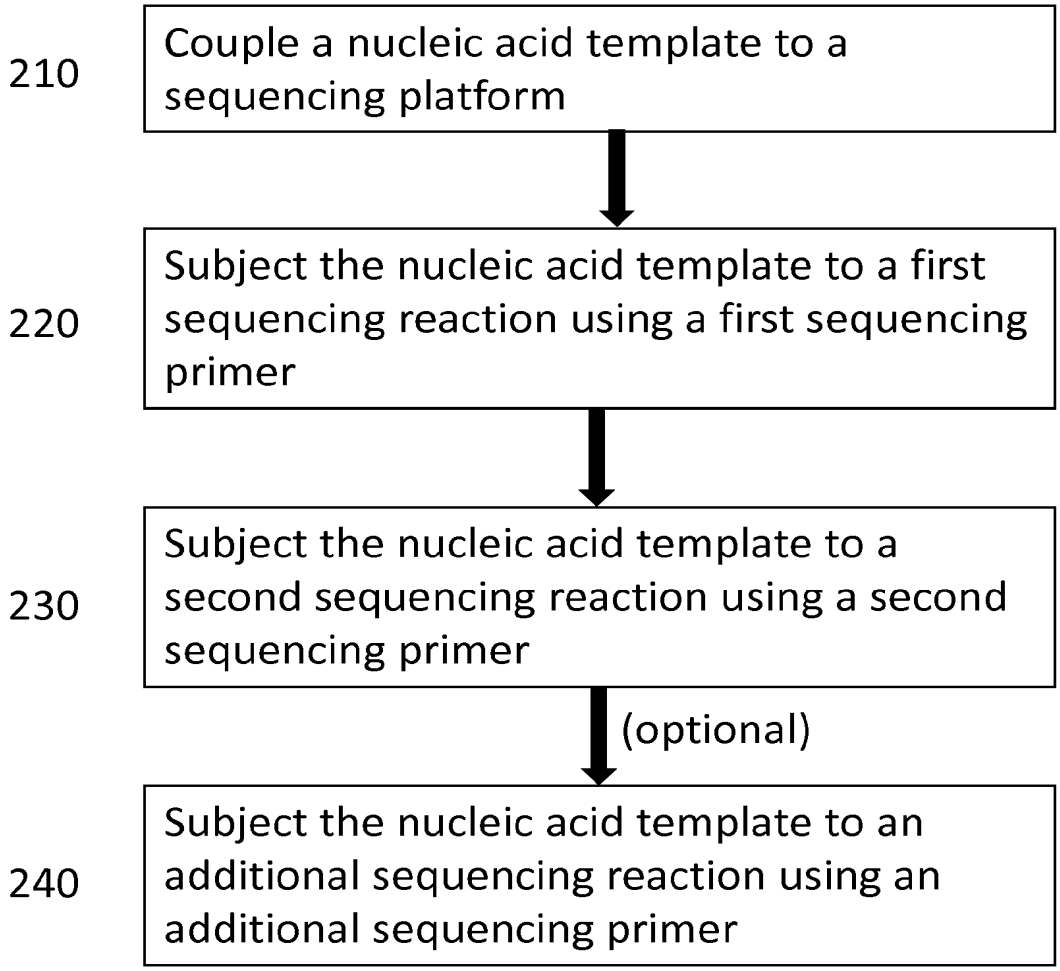

210
Couple a nucleic acid template to a sequencing platform
220
Subject the nucleic acid template to a first sequencing reaction using a first sequencing primer
230
Subject the nucleic acid template to a second sequencing reaction using a second sequencing primer
(optional)
240
Subject the nucleic acid template to an additional sequencing reaction using an additional sequencing primer

FIG. 2

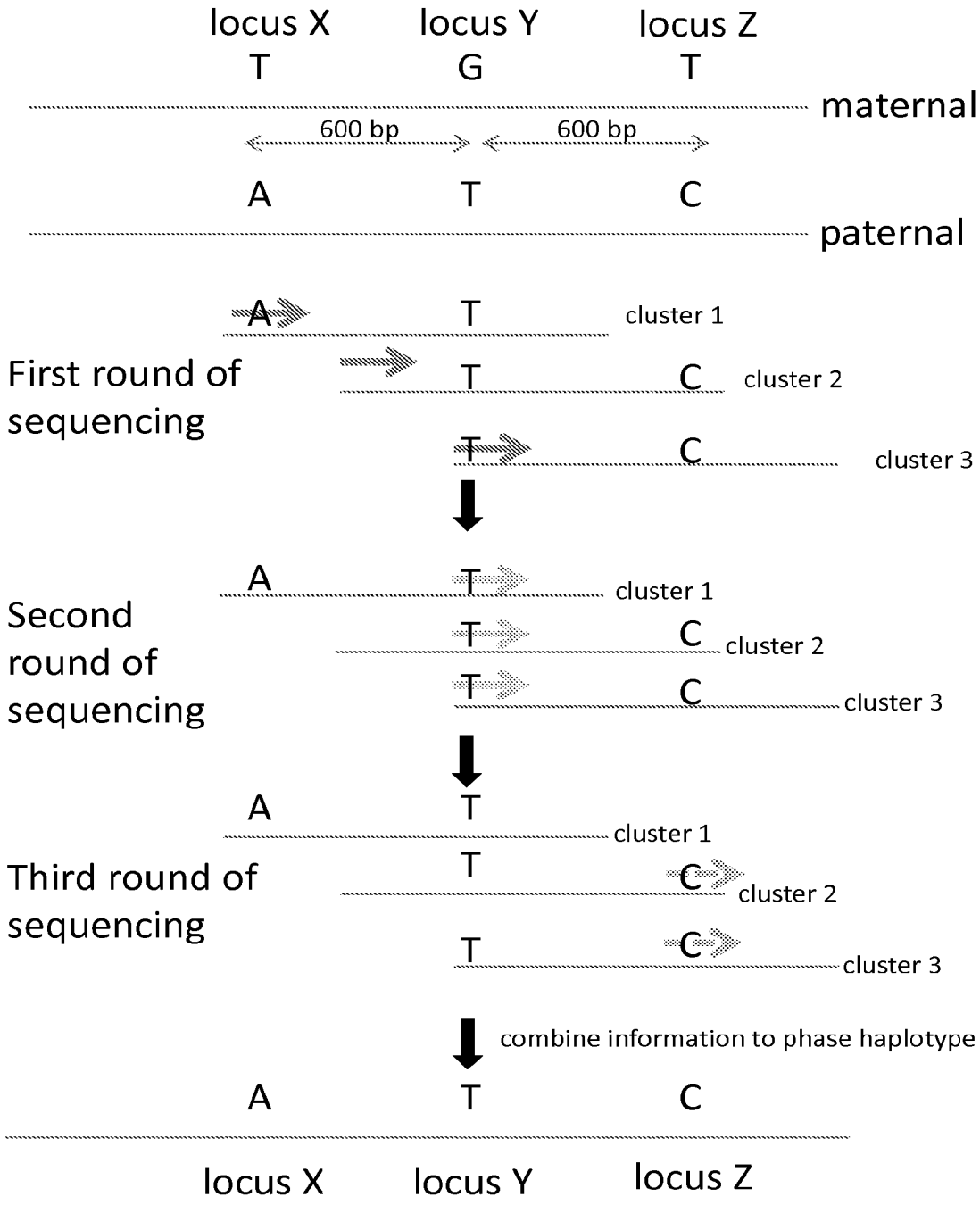

locus X
locus Y
locus Z
T
G
T
maternal
600 bp
600 bp
A
T
C
paternal
First round of sequencing
A
T
cluster 1
T
C
cluster 2
T
C
cluster 3
Second round of sequencing
A
T
cluster 1
T
C
cluster 2
T
C
cluster 3
Third round of sequencing
A
T
cluster 1
T
C
cluster 2
T
C
cluster 3
combine information to phase haplotype
A
T
C
locus X
locus Y
locus Z

FIG. 3

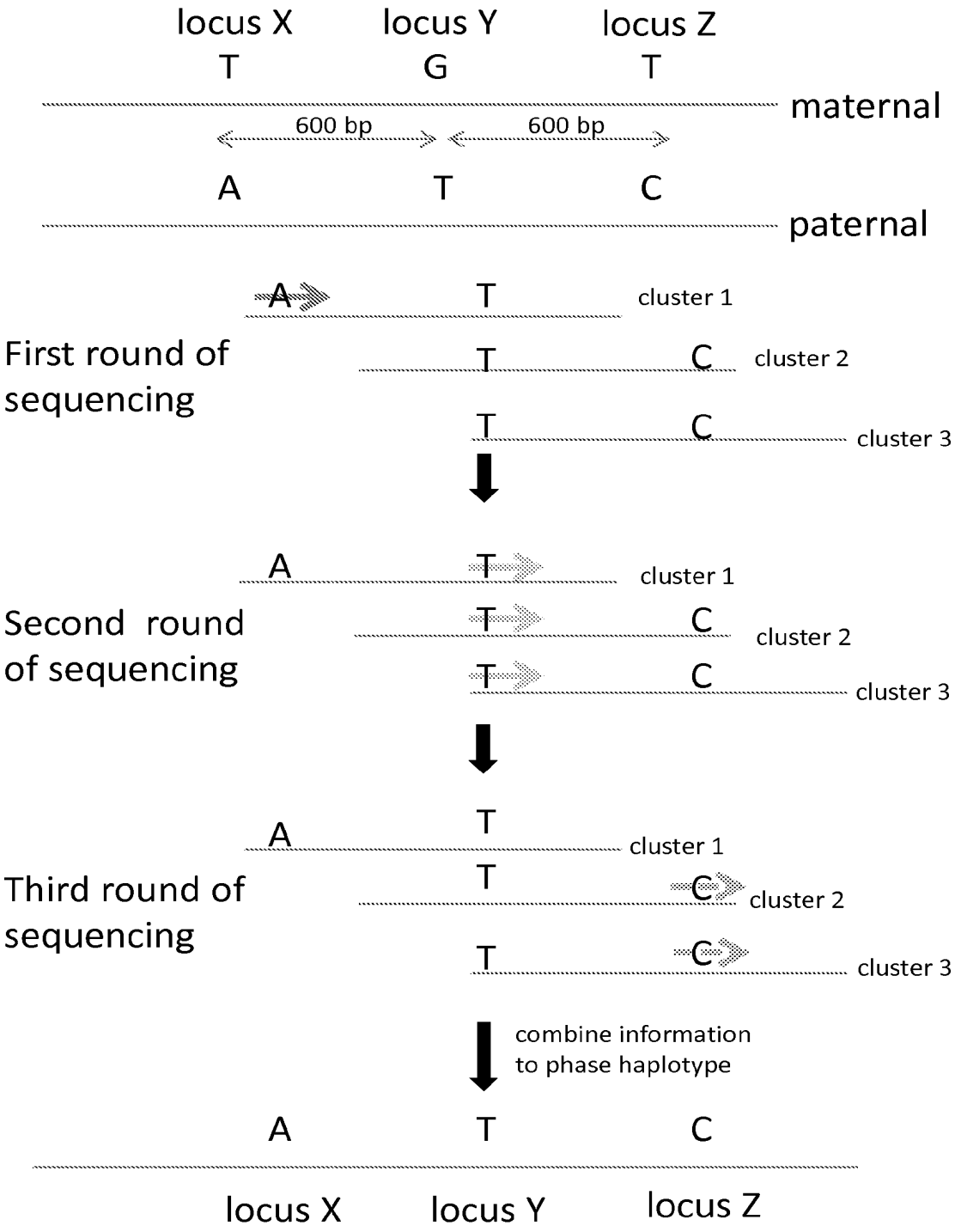

locus X
locus Y
locus Z
T
G
T
maternal
600 bp
600 bp
A
T
C
paternal
First round of sequencing
cluster 1
cluster 2
cluster 3
Second round of sequencing
Third round of sequencing
combine information to phase haplotype
locus X
locus Y
locus Z

FIG. 4

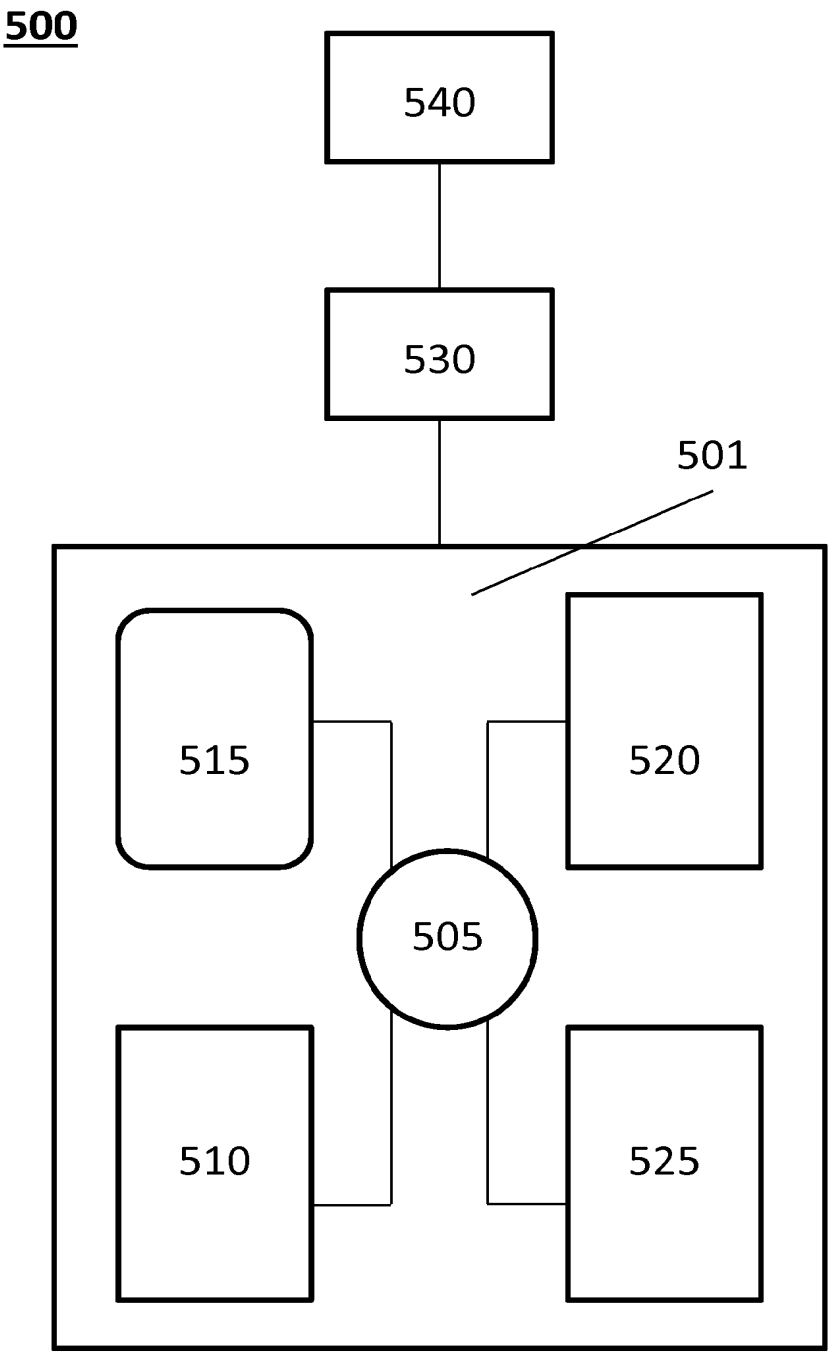


FIG. 5

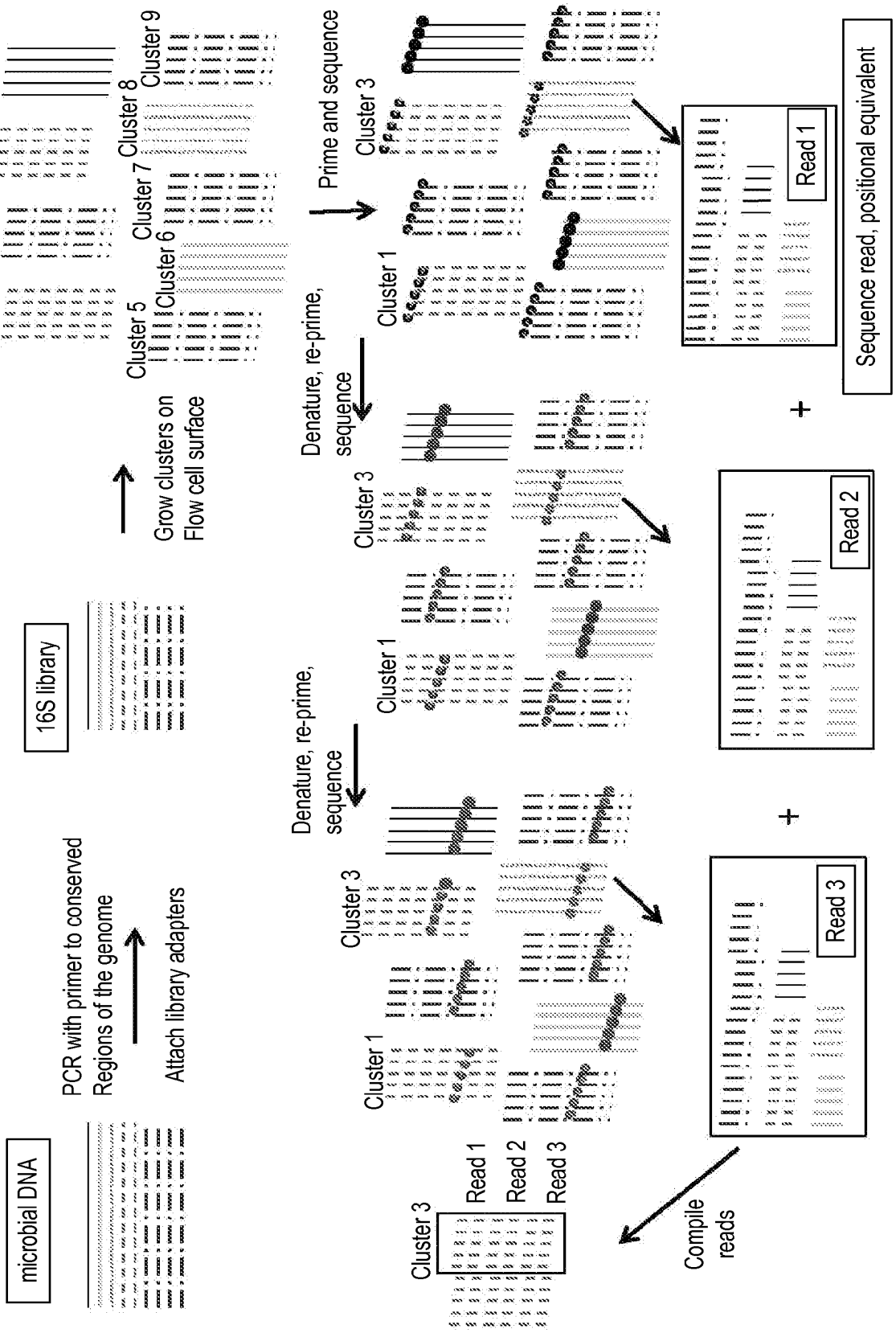


FIG. 6

IDT Exhibit 1020.2371

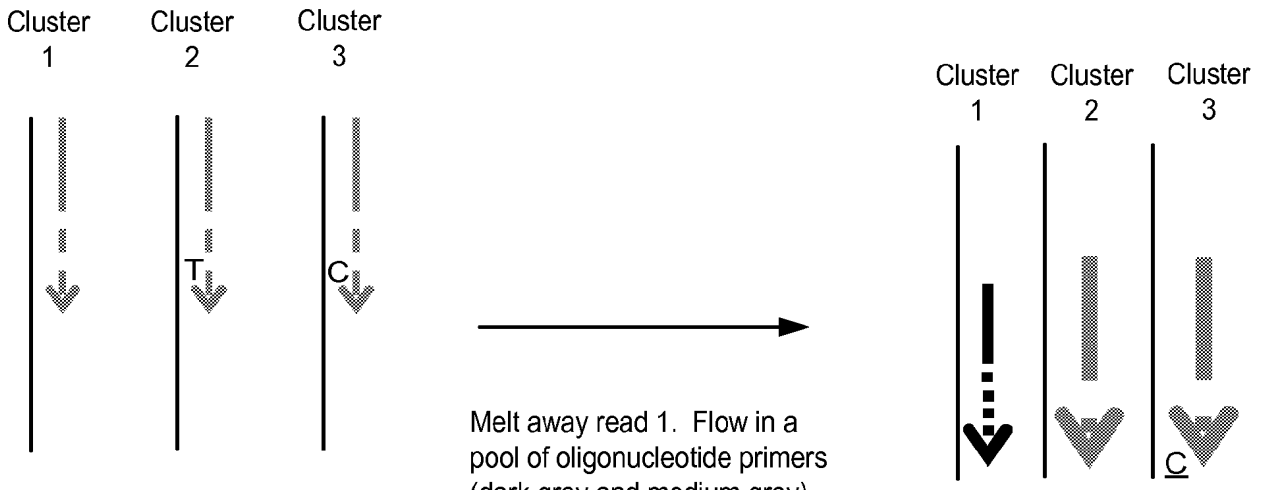


FIG. 7

IDT Exhibit 1020.2372

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2014/028356

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12Q 1/68 (2014.01)
USPC - 435/6.11
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8) - C12Q 1/68; C40B 20/02, 20/04, 20/06, 20/08, 30/00, 30/10, 40/06, 50/06 (2014.01)
USPC - 435/6.11, 6.12, 6.15; 506/4, 5, 6, 7, 16, 26; 536/23.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
CPC - A61K 38/00; B01J 19/0046, 2219/00585, 2219/00596, 2219/00707, 2219/00722; C12Q 1/6827, 1/6883, 1/689 (2014.06)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Orbit, Google Patents, Google Scholar

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 2009/0298075 A1 (TRAVERS et al) 03 December 2009 (03.12.2009) entire document	1-36, 39-45, 47, 48 --- 37, 38
Y	US 2010/0105052 A1 (DRMANAC et al) 29 April 2010 (29.04.2010) entire document	37, 38
A	US 2012/0071331 A1 (CASBON et al) 22 March 2012 (22.03.2012) entire document	1-45, 47, 48
A	US 2011/0319290 A1 (RAYMOND et al) 29 December 2011 (29.12.2011) entire document	1-45, 47, 48
A	US 2011/0105364 A1 (KURN et al) 05 May 2011 (05.05.2011) entire document	1-45, 47, 48

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
26 June 2014	15 JUL 2014
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (July 2009)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2014/028356

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 46
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

IDT Exhibit 1020.2374

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 4 :		(11) International Publication Number:
C12Q 1/68	A1	WO 89/ 09284
		(43) International Publication Date: 5 October 1989 (05.10.89)

(21) International Application Number: PCT/US89/01025

(22) International Filing Date: 13 March 1989 (13.03.89)

(31) Priority Application Number: 173,127

(32) Priority Date: 24 March 1988 (24.03.88)

(33) Priority Country: US

(71) Applicant: UNIVERSITY OF IOWA RESEARCH FOUNDATION [US/US]; 203 Gilmore Hall, Iowa City, IA 52242 (US).

(72) Inventors: WALDER, Joseph, A. ; WALDER, Roxanne, Y. ; 2107 Slagle Circle, Iowa City, IA 52240 (US).

(74) Agents: SEASE, Edmund, J. et al.; Zarley, McKee, Thomte, Voorhees & Sease, 2400 Ruan Center, Des Moines, IA 50309 (US).

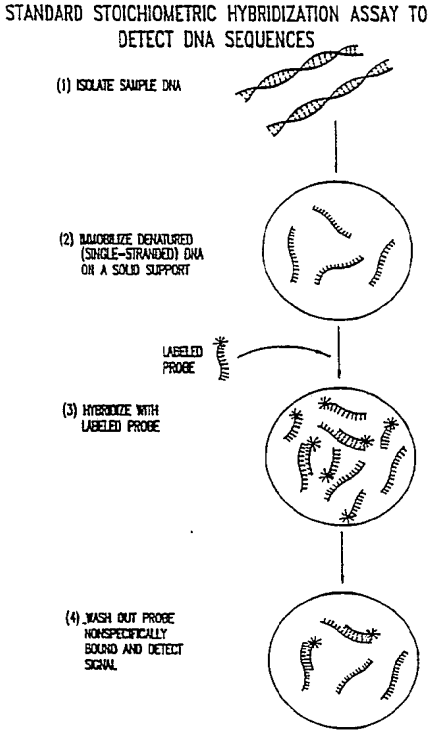
(81) Designated States: AT (European patent), AU, BE (European patent), CH (European patent), DE (European patent), FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent).

Published
With international search report.

(54) Title: CATALYTIC HYBRIDIZATION SYSTEMS FOR THE DETECTION OF NUCLEIC ACID SEQUENCES BASED ON THEIR ACTIVITY AS COFACTORS IN CATALYTIC REACTIONS IN WHICH A COMPLEMENTARY LABELED NUCLEIC ACID PROBE IS CLEAVED

(57) Abstract

A method of detection of nucleic acid (DNA or RNA) target sequence in which such a sequence serves as a cofactor for a catalytic reaction in which a complementary, labeled nucleic acid probe is cleaved such that the target sequence is released intact and can repeatedly recycle through the reaction pathway, thereby providing signal amplification.



IDT Exhibit 1020.2375

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**FR**	France	**ML**	Mali
AU	Australia	**GA**	Gabon	**MR**	Mauritania
BB	Barbados	**GB**	United Kingdom	**MW**	Malawi
BE	Belgium	**HU**	Hungary	**NL**	Netherlands
BG	Bulgaria	**IT**	Italy	**NO**	Norway
BJ	Benin	**JP**	Japan	**RO**	Romania
BR	Brazil	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SN**	Senegal
CH	Switzerland	**LI**	Liechtenstein	**SU**	Soviet Union
CM	Cameroon	**LK**	Sri Lanka	**TD**	Chad
DE	Germany, Federal Republic of	**LU**	Luxembourg	**TG**	Togo
DK	Denmark	**MC**	Monaco	**US**	United States of America
FI	Finland	**MG**	Madagascar		

IDT Exhibit 1020.2376

CATALYTIC HYBRIDIZATION SYSTEMS FOR THE DETECTION OF NUCLEIC ACID SEQUENCES BASED ON THEIR ACTIVITY AS COFACTORS IN CATALYTIC REACTIONS IN WHICH A COMPLEMENTARY LABELED NUCLEIC ACID PROBE IS CLEAVED

SUBSTITUTE SHEET

IDT Exhibit 1020.2377

GRANT REFERENCE

The invention described herein was made in part in course of work under grants from the National Institutes of Health, Grant Nos. HL-33555 and AM-25295.

CROSS-REFERENCE

This is a continuation-in-part application of earlier filed, co-pending application Serial No. 126,564, filed November 30, 1987.

BACKGROUND OF THE INVENTION

1. Field of the Invention

This invention provides a diagnostic assay methodology for the sensitive detection of specific nucleic acid (DNA and RNA) sequences in biological and clinical samples. The invention particularly addresses the problem of detecting low copy numbers of a target sequence with high specificity.

All current methods for the detection of DNA and RNA are based on the stoichiometric (1:1) hybridization of a complementary, labeled nucleic acid probe to the target sequence (see Figure 1). The present invention is a marked departure from this approach: the target sequence serves as a cofactor for a reaction in which the probe is cleaved catalytically. After cleavage of the probe, the target sequence is released intact and can be repeatedly recycled through the reaction leading to a large amplification of the response (see Figure 2). We have termed such reaction schemes catalytic hybridization amplification reactions. In such assays, cleavage of the labeled probe signals the presence of the target sequence. The sensitivity of the method is far greater than existing procedures, all of which are

based on stoichiometric or unitary hybridization schemes in which the target sequence captures one, and only one, molecule of the probe. Several embodiments of the catalytic hybridization amplification method are described below. All operations required for such assays are simple and can be readily carried out in a clinical laboratory. The invention will have applications in the diagnosis of a variety of diseases in man and other species including genetic disorders, infectious diseases, and cancer.

2. Description of Prior Art

It is well known that nucleic acids, i.e., deoxyribonucleic acids (DNA) and ribonucleic acids (RNA) are essential building blocks of all organisms. These are high molecular weight polymers that are made up of many nucleotide units, each such nucleotide unit being composed of a base (a purine or a pyrimidine), a sugar (which is either ribose or deoxyribose) and a molecule of phosphoric acid. DNA contains deoxyribose as the sugar moiety and the bases adenine, guanine, cytosine, and thymine (which may be represented as A, G, C and T, respectively). RNA contains ribose instead of deoxyribose and uracil (U) instead of thymine.

The nucleotide units in DNA and RNA are assembled in definite linear sequences which determine specific biological functions. In bacterial cells, and in all higher species, DNA directs its own replication, and also serves as the template for the synthesis of RNA molecules whose nucleotide sequences carry the information encoded by the DNA. The process of RNA synthesis is termed transcription. RNA molecules serve several different functions within

the cell. Messenger RNAs (mRNAs) direct protein synthesis. Ribosomal RNAs (rRNAs) are important constituents of ribosomes, the organelle within the cell at which mRNAs are read, or translated, and proteins are made.

Collectively, the genetic information of an organism is termed the genome. The genome of bacteria and all higher species is composed of DNA. The genome of viruses may be either DNA or RNA. In any case, the genome of any particular species, whether a virus, bacteria or a higher organism has a characteristic nucleotide sequence which, stated simply, can be viewed as the "fingerprint" of that species. Sequences within a genome that code for proteins, or that are transcribed to form RNAs with specific functions such as ribosomal RNA represent individual genes. Small viruses such as the AIDS virus have as few as 10 genes. The human genome contains approximately 50,000 genes.

According to the well known Watson-Crick model, DNA molecules consist of two polynucleotide strands coiled about a common axis. The resulting double helix is held together by hydrogen bonds between complementary base pairs in each strand. Hydrogen bonds are formed only between adenine (A) and thymine (T) and between guanine (G) and cytosine (C). Hence within the double helix, adenine (A) and thymine (T) are viewed as complementary bases which bind to each other as A-T. The same is true for guanine (G) and cytosine (C) as G-C.

In a single polynucleotide chain, any sequence of nucleotides is possible. However, once the order of bases within one strand of a DNA molecule is specified, the exact sequence of the other strand is simultaneously determined due to the indicated rules

of base pairing. Accordingly, each strand of DNA molecule is the complement of the other. The process by which two complementary DNA strands associate with one another is termed hybridization. The process of strand separation, which is generally accomplished by heating the sample, is termed melting or denaturation of the duplex. Duplexes have a characteristic melting temperature (Tm) depending primarily on length, composition, and salt concentration. The following equation provides a useful approximation of the Tm:

$$Tm=16.6X\log Cs+0.41x(\%G+C)+81.5-820/L \quad (1)$$

where Cs is the salt concentration, (%G+C) is the percent G-C content, and L is the length of the duplex (Schildkraut C. and Lifson, S. (1965) Biopolymers 3, 195-208; and Thomas, C.A. Jr. and Dancis, B.M. (1973) Journal of Molecular Biology 77, 43-55). Mismatches between the two strands of a duplex (e.g. a C-T base pair) can occur, but decrease the stability of the helix significantly and lower the melting temperature.

In addition to DNA:DNA duplexes, DNA:RNA duplexes and RNA:RNA duplexes can also form and occur naturally. DNA:RNA duplexes are formed transiently in the process of transcription in which DNA serves as a template for RNA synthesis. RNA:RNA duplexes occur as the genetic material of certain viruses and are also formed in so-called hairpin loops of ribosomal RNA and transfer RNA. In duplexes involving an RNA strand, uracil (U) pairs with adenine (A). Within a single polynucleotide strand, sequences of both RNA and DNA can also occur. Such RNA-DNA copolymers are formed as intermediates in the replication of DNA. The melting temperature of a DNA:RNA or an RNA:RNA duplex may differ somewhat, depending on solution conditions, from that of a DNA:RNA duplex

of the same sequence, but the equation given above remains a useful approximation of the Tm. For a further discussion of these various polynucleotide structures and their biological functions see "Genes III" by Benjamin Lewin, John Wiley and Sons, 1987, which is incorporated herein by reference.

Recent advances in molecular biology employing recombinant DNA techniques have led to new diagnostic and therapeutic strategies. In the area of DNA probe diagnostics, DNA probes are used to detect the presence of a complementary target nucleic acid sequence in the sample. Applications include the diagnosis of infectious diseases, cancer and genetic disorders. In the case of infectious diseases, the target can be a DNA or RNA sequence unique to a particular bacteria or virus.

In certain cancer cells there are specific gene rearrangements which can be detected, e.g. the translocation of the c-abl oncogene in chronic myelogenous leukemia. Genetic defects can involve large deletions or insertions of DNA, or a single point mutation as in the case of sickle cell disease. In the latter case, it is necessary to detect the alteration of a single base pair in the context of flanking sequences that are identical to those found in the normal gene.

All nucleic acid diagnostic assays involve three fundamental steps:

(1) Isolation and further preparation of DNA or RNA from the sample. When DNA is isolated it is generally sheared by mechanical forces. With gentle manipulation, the largest fragment size may be up to about 100,000 base pairs. DNA may also be purposefully cut with sequence specific endonucleases termed restriction enzymes to give rise to fragments of dis-

crete sizes. Such is the case in so-called Southern blot analysis in which the fragments are subsequently separated according to size by electrophoresis and then immobilized on a filter such as nitrocellulose or a nylon membrane (Southern, E.M. (1975) Journal of Molecular Biology 98, 503-517).

(2) Hybridization of a complementary, labeled nucleic acid probe to the target sequence. In this process, the sample is denatured, generally by heating, to ensure that the target sequence is single stranded. The labeled probe is then allowed to hybridize to the target. Molecules of the probe that are weakly bound to other sequences within the sample, or the filter on which it is immobilized, are then washed away. To achieve adequate specificity, the melting temperature of the probe:target duplex must be at least 5 to 10 degrees above that of duplexes that may be formed between the probe and other sequences present within the sample with which there is one or more mismatches.

(3) Detection of the amount of the labeled probe hybridized to the DNA or RNA within the sample. Suitable reporter groups with which the probe may be labeled include radioisotopes such as ^{3}H, ^{32}P or ^{121}I, fluorescent dyes such as fluorescene, Texas Red or the phycobiliproteins, or enzyme markers such as alkaline phosphatase, β-galactosidase, glucose oxidase or peroxidases. An example of a diagnostic assay in which the sample nucleic acid is immobilized on a filter is shown in Figure 1 (see Falkow et al. (1982) U.S. Patent No. 4,358,535).

In addition to the filter format, a variety of other strategies for diagnostic hybridization assays have been described, see for example Heller et al. (1983) EPA 0 070 685 and EPA 0 707 687; Rentz, M. and

Kurz, C. (1984) Nucleic Acids Research, 12, 3435-3444; Ruth, J.L. (1984) P.C.T. W084/03285; and Kohne, D. (1984) P.C.T. W084/02721. In each of these systems the reporter group that is to be detected is attached directly to the probe as is shown in Figure 1. Alternatively, an indirect labelling scheme may be employed. In the method described by P. Kourlisky et al. (1979) GB 2 019 408 the probe is labeled with biotin. After the probe is hybridized to the target sequence, a complex of the enzyme β-galactosidase covalently linked to the protein avidin is added to the sample. Avidin binds specifically and with very high affinity to the biotin label attached to the probe. The probe:target duplex is then detected by providing an appropriate substrate for β-galactosidase which is hydrolyzed to give a fluorescent or colored product. Biotin-avidin probe systems have also been described by Ward et al. (1982) EPA 0 063 879, and by Engelhardt et al. (1984) EPA 0 097 373. More recently, Carrico (1987) EPA 0 209 702 have described an indirect labelling strategy in which a probe:target duplex formed between an RNA strand and a DNA strand are detected using an antibody which binds specifically to RNA:DNA duplexes.

All of the above prior art methods involve a stoichiometric or unitary hybridization reaction in which the target sequence binds to or captures one and only one molecule of the probe. The present invention provides a novel and improved hybridization assay method in which the target sequence is able to capture many molecules of the probe in a repeating series of reactions. This is accomplished by cleavage of the probe within the probe:target duplex such that the target sequence is released intact and can repeatedly recycle through the reaction pathway. We

term such methods catalytic hybridization amplification (CHA) reactions. A generalized version of the CHA method is shown in Figure 2. In such assays, the signal that is detected results from the cleavage of the probe. The sensitivity of the method is far greater than that of earlier hybridization schemes, all of which are based on a stoichiometric, i.e. 1:1, association between the target sequence and the probe.

Several methods described previously involve cleavage of the probe used in the hybridization assay, but preclude recycling of the target sequence, the essential element of the present invention. Thus, Ashihara et al. (1985) EPA 0 142 299 describe a method in which the DNA:DNA duplex formed between the probe and the target sequence is cleaved by a restriction enzyme. The restriction enzyme recognizes a specific sequence, generally 4 to 6 base pairs in length, within the duplex and cleaves both the probe and the target strand. Because the target is cleaved, it is impossible for it to recycle through the reaction pathway. A procedure for detecting single base changes, i.e. point mutations, within DNA has been described by Myers et al. (1985) Science 230, 1242-1246 that is based on the cleavage of an RNA probe. A labeled RNA probe is first hybridized to the target DNA. The RNA probe hybridized to the DNA is then subjected to cleavage by the enzyme RNaseA. Excess, unbound molecules of the probe are fully degraded. If the RNA forms a perfect duplex with no base pair mismatches with the DNA, the hybridized RNA strand will not be cut. If there are one or more base pair mismatches, the RNA probe will be cut by the enzyme. Cleavage of the RNA probe at the site of the mismatch(es), which may be assayed by a number of different means, is thus used to detect

the altered DNA sequence. Since RNaseA will cleave free, unhybridized RNA molecules, it is essential for the cleaved fragments to remain hybridized to the target sequence. This makes it impossible for the target DNA sequence to react repeatedly with multiple copies of the probe. An assay having some similar characteristics has recently been described by Duck et al. (1987) EPA 0 227 976. In this case, excess, unhybridized probe is first digested away with an enzyme that will not cleave the probe when bound in the probe:target duplex. An assay is then carried out to detect the remaining molecules of the probe hybridized to the target sequence. Because it is essential for excess, unhybridized probe molecules to be first degraded, the target sequence cannot turn-over and react repeatedly in the reaction as in the present invention.

Hull Vary et al. (1986) EPA 0 200 057 has described a system in which hybridization of the target sequence to the probe displaces a third poly-nucleotide that is bound to the probe. At the end of the hybridization reaction, the displaced polynucle-otide is degraded into mononucleotides which are detected. Again in this system, each molecule of the target sequence hybridizes with one and only one copy of the probe. Recycling of the target sequence cannot occur. A method to enhance the rate of nucle-ic acid hybridization has been described by Zapolski et al. (1985) P.C.T. WO 85/05685. In this procedure, the enzyme RecA and a single stranded DNA binding protein are used to promote the hybridization of the probe to the target sequence. This increases the rate of hybridization, but the reaction is still stoichiometric involving the binding of a single molecule of the probe by each molecule of the target.

Turnover of the target sequence enabling it to capture multiple copies of the probe does not occur as in the present invention.

The prior art presented above represents a number of different strategies for labeling probes, carrying out hybridization reactions, and enhancing signal detection. There are as yet, however, no nucleic acid hybridization assay systems in widespread use. Existing methods lack adequate sensitivity for many clinical applications and often involve complex assay procedures difficult to carry out routinely in a clinical laboratory. All of the prior systems involve a stoichiometric or unitary hybridization of the probe to the target sequence, such that each target molecule can capture one an only one molecule of the probe. The present invention makes a fundamental departure from this concept.

The primary objective of the present invention is to provide an improved hybridization assay method in which the target sequence serves as a catalytic cofactor for the cleavage of a complementary, labeled nucleic acid probe. Recycling of the target through the reaction pathway enables it to capture many molecules of the probe, thereby leading to a large increase in the sensitivity of the assay.

Yet a further objective of the present invention is to provide formats, reagents, and assay conditions for such catalytic hybridization amplification (CHA) reactions useful in diagnostic tests.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic illustration of a standard hybridization assay for the detection of a DNA target sequence based on the formation of a stoichiometric (1:1) complex between the target

sequence and a complementary, labeled nucleic acid probe. In this case the sample DNA is immobilized on a filter support.

Figure 2 depicts the use of the CHA method for the detection of a specific DNA sequence. The target DNA sequence serves as a catalytic cofactor for the cleavage of a complementary, labeled nucleic acid probe. Repeated recycling of the target sequence through the reaction (dashed arrow) leads to a large amplification of the signal and enhanced sensitivity compared to standard assay systems based on a stoichiometric or unitary hybridization of the probe to the target sequence.

Figure 3 illustrates the CHA method in a solid support format in which cleavage of the complementary labeled probe attached to the support is mediated by RNaseH.

Figure 4 illustrates a hybridization detection system in which two CHA reactions are coupled to produce an exponential amplification of signal.

Figure 5 demonstrates the use of a competition assay to detect a specific RNA sequence with the CHA method.

SUMMARY OF THE INVENTION

The present invention of catalytic hybridization amplification (CHA) provides a novel and improved nucleic acid hybridization method useful in diagnostic assays. The target sequence is not detected on the basis of stoichiometric (1:1) hybridization to the probe as in all previous assay systems. The invention concerns the target nucleic acid sequence serving as a highly specific cofactor for a catalytic reaction which leads to the cleavage of a complementary, labeled probe hybridized to the target.

Molecules of the probe not hybridized to the target sequence are not cut. Upon cleavage of the labeled probe in the labeled probe:target duplex, the target sequence is released intact and can repeatedly cycle through the reaction pathway. In such assays, cleavage of the probe generates the signal that is detected. The ability of each molecule of the target to capture multiple copies of the probe leads to a large enhancement in the sensitivity of the method. Several embodiments of the CHA method are provided employing simple assay procedures suitable for use in clinical diagnostic tests. Applications of these methods include, but are not limited to, the diagnosis of infectious diseases, genetic defects and cancer. Both single point mutations as well as longer unique sequences can be detected.

DETAILED DESCRIPTION OF THE INVENTION

The invention of catalytic hybridization amplification is unique in that it represents the first example of a non-stoichiometric hybridization process. In such systems the target sequence (either DNA or RNA) serves as a cofactor in a catalytic reaction in which a complementary, labeled nucleic acid probe is cleaved. A generalized version of the CHA method is shown in Figure 2.

In the first step of the process, the target sequence hybridizes to the probe by the usual rules of base pairing. Once hybridized to the target sequence, the probe is then cleaved into two or more smaller fragments. The cleavage reaction can be catalyzed by an enzyme or occur by other chemical methods. Only molecules of the probe hybridized to the target sequence are cleaved. Excess molecules of the probe not hybridized to the target are not cut.

The smaller fragments of the probe generated by the cleavage reaction form less stable duplexes with the target than does the parent molecule, and under conditions of the CHA reaction dissociate from the target. For dissociation to occur rapidly, the melting temperature of the duplexes formed with the smaller fragments should generally be 10°C or more below the reaction temperature. For short oligonucleotide probes from about 15 to about 35 nucleotides in length, this will occur following a single cleavage near the middle of the probe sequence. For longer probes, which may be up to several thousand nucleotides in length, multiple cleavages must occur, generating fragments shorter than about 35 nucleotides in length, preferably between about 10 and 20 nucleotides long. The target sequence must not be cleaved in the reaction. It is released intact from the probe:target duplex and recycles repeatedly through the reaction pathway. At the completion of the CHA reaction the extent of cleavage of the probe is measured. This generally involves separation of the probe fragment bearing the reporter group label form the intact molecule. It is the cleavage of the probe that signals the presence of the target sequence. Turnover of the target sequence, enabling each molecule to capture and promote the cleavage of many copies of the probe, greatly enhances the sensitivity of the method. Several embodiments of the CHA method will now be described.

In the principal embodiment of this invention, as applied to the detection of a DNA sequence, the cleavage reaction is catalyzed by the enzyme RNaseH, and takes place only when the target DNA sequence is hybridized to a complementary RNA probe, forming an RNA:DNA duplex. RNaseH is highly specific: it will

not cut free molecules of the probe that are not hybridized to the target DNA sequence, nor does it cleave the target DNA sequence. Hence, upon cleavage of the probe, the target sequence is released intact and can repeatedly cycle through the reaction pathway as outlined in Figure 2. A detailed description of the components (probes, labels, etc.), procedures, and conditions for this embodiment of the CHA method follows.

Probes that can be used in the invention range from oligonucleotides of about 15 to about 50 nucleotides in length to longer polynucleotides which may be up to several thousand nucleotides long. A probe of about 15 residues is the shortest sequence that can be used to selectively hybridize to a complementary target sequence (e.g. a unique viral sequence) against the background of human DNA (see Thomas, C.A., Jr. (1966) Progress in Nucleic Research and Molecular Biology 5, 315). More often, sequences of at least 20 to about 25 nucleotides have to be used to assure a high level of specificity. Short oligonucleotides up to about 25 residues are generally labeled with a single reporter group, either near the 5'- or 3'-end of the strand. Longer sequences may be labeled up to a density of about one reporter group per 15 nucleotide residues. The composition of the probe can be entirely RNA. Alternatively, the probe may contain one or more interspersed sequences not cleavable by RNaseH. Such sequences not cleaved by RNaseH can be composed of DNA, or of RNA in which either the phosphate groups or the sugar moieties are modified so as to prevent cleavage by the enzyme. Appropriate modifications of the phosphate groups include, but are not limited to, alkyl or aryl phosphotriesters, hydrogen phosphonates, alkyl or aryl

phosphonates, alkyl or aryl phosphoramidates, phosphorothioates, or phosphoroselenates. The preferred modification of the ribose sugar to prevent cleavage by RNaseH is the conversion of the 2'-OH group to an alkyl or aromatic ether. To permit recognition by RNaseH, the cleavable RNA sequence(s) must be at least 3 residues in length.

Probes containing interspersed sequences not cleavable by RNaseH focus the activity of the enzyme to specified regions and are, therefore, preferred. If an oligonucleotide 26 resi-dues in length composed entirely of RNA was used as a probe, cleavage at many sites would be nonproductive, that is, would not lead to release of the target sequence. For example, cleavage of the strand between residues 23 and 24 would generate one fragment 23 residues in length and a second fragment of only 3 residues. The trinucleotide would dissociate from the target sequence very readily, but the 23-mer would form a duplex with the target sequence almost as stable as that formed with the parent molecule, and, hence, would not be released. Alternatively, if the oligonucleotide had the sequence $DNA_{11}RNA_4DNA_{11}$ cleavage would be limited to the central four RNA residues. The longest fragment so generated would be 15 residues. The Tm of this fragment would be about 20°C to 25°C below that of the parent 26-mer (see Equation 1) and would readily dissociate from the target sequence under the CHA reaction conditions. Thus, each time the probe was cleaved, the target sequence would be released and able to recycle through the reaction pathway. In the case of the pure RNA oligonucleotide probe two or more cleavages of the probe may be necessary to release the target. For longer probes, the length of the interspersed noncleavable

sequences should be less than about 35 nucleotides long and preferable between 10 and 20 nucleotides in length, to permit rapid release of the fragments generated by RNaseH cleavage from the target sequence.

A mixed DNA-RNA-DNA oligonucleotide probe is particularly useful for the detection of single point mutations, as is required for the diagnosis of certain genetic diseases such as sickle cell disease. Consider a probe that forms a perfect duplex with the normal target sequence, and a duplex with the mutant gene in which there is a single base pair mismatch. The entire probe, including the flanking DNA sequences, provides the specificity for the target sequence normally achieved through hybridization. The greatest difference in the stability of the two helices is obtained with a relatively short oligonucleotide probe in which the mismatch is near the middle of the sequence. Even in this case, however, the melting temperature of the perfect duplex formed with the normal sequence is only about 5°C to 10°C greater than that for the duplex formed with the mutant sequence having the mismatch. Although it is possible to distinguish the normal form the mutant sequence based on this small difference using a standard hybridization format such as that shown in Figure 1, the reliability of the method is not sufficient for routine diagnostic use. The CHA reaction scheme, however, affords a further level of specificity that enables these two sequences to be readily differentiated. As noted above, cleave of an DNA-RNA-DNA probe by RNaseH is limited to the central RNA sequence. If the RNA sequence is four or five residues, a single base pair mismatch disrupts the structure of the RNA:DNA duplex sufficiently to

prevent cleavage of the RNA strand by RNaseH. Thus, the probe would be cleaved when hybridized to the normal target sequence but not when hybridized to the mutant sequence. Only the normal target sequence would serve as a catalytic cofactor for the cleavage of the probe. If the central RNA sequence is between six and eight residues, two mismatches within the RNA:DNA portion of the duplex may be necessary to disrupt the structure sufficiently to prevent cleavage by RNaseH. This can be accomplished using a probe in which there is a single mismatch with the normal sequence and two mismatches with respect to the mutant sequence, e.g.

```
.....dAdCrArUrCrCrGrArUdAdG... probe
                    *
.....dTdGdTdAdTdGdCdTdAdTdC... normal target sequence

.....dAdCrArUrCrCrGrArUdAdG.... probe
                   * *
.....dTdGdTdAdTdGdAdTdAdTdC.... mutant sequence
```

where dN (N = A,C,T or G) is a deoxyribonucleotide (DNA), rN a ribonucleotide (RNA), and * a mismatched base pair.

Probes themselves can be made and modified by both molecular biology and synthetic techniques. Long probes greater than about 60 nucleotides in length are generally synthesized and labeled by enzymatic methods. Long RNA fragments can be synthesized from a DNA template using either SP6 or T7 RNA polymerase. Oligonucleotides are generally synthesized by chemical methods, see "Oligonucleotide Synthesis: A Practical Approach" (Gait, M.J., ed.) IRL Press, Oxford (1984), which is incorporated herein by reference. The phosphoramidite method is the preferred chemistry at present. The synthesis

can be carried out with the growing oligonucleotide chain attached to a solid support, such as polystyrene or glass beads, or free in solution. DNA, RNA and DNA-RNA-DNA oligonucleotides can be prepared by these approaches. DNA, RNA and mixed DNA-RNA molecules can also be further ligated together in any combination using the enzyme T_4 RNA ligase. Any of the above techniques can be employed for the synthesis of probes useful with this invention.

Radioisotopes, fluorophores, lumiphores (chemilumenescent or biolumeniscent substrates), and enzymes can all be used as reporter groups in catalytic hybridization amplification systems. The fact that the system produces a large amplification response before the actual detection procedure is carried out, in most cases, significantly reduces the demand on the sensitivity of the reporter group being used. Reporters can be chosen for convenience in the assay system, or because they are better adapted to the assay format (instrumentation versus manual procedures), as well as for greatest sensitivity.

Radioisotope such as ^{32}P, ^{14}C, ^{125}I, and tritium can be easily incorporated into DNA and RNA probes with molecular biology techniques, either during synthesis or after completion. Probes synthesized chemically can also be labeled with radioisotopes. While radiolabeling has certain advantages for experimental and research purposes, it does not appear to be the appropriate method for new tests and assays in the clinical laboratory, and is therefore not preferred.

Nonisotopic labels of choice include: (1) fluorophores such as fluorescein, Texas Red, Lucifer Yellow, pyrenes, chelated lanthanides, and phycobili-

proteins to name a few; (2) lumiphores such as luminol and derivatives; and (3) enzymes such as alkaline phosphatase, peroxidase, β-galacto-sidase, and luciferases to name a few. The various enzymes can be used to produce color reactions (alkaline phosphatase, peroxidase), fluorescent products (alkaline phosphatase), and luminescence (luciferase).

Fluorophores and lumiphores can be covalently attached to RNA and DNA sequences by a variety of techniques. In the case of a nucleic acid probe in which the 3'-terminal nucleotide is a ribose residue, the ribose diol at the 3'-end of the sequence can be oxidized with periodate to form a reactive dialdehyde. A reductive alkylation reaction can now be carried out where the dialdehyde can form a Schiff's base with a variety of fluorescent derivatives containing a primary amino group (dansylethylenediamine, N-(1-pyrenesulfonyl)ethylenediamine, 5-((2-aminoethyl)thiouridyl)-fluorescein, etc.). Primary amine derivatives of luminol and other lumiphores can also be coupled. The Schiff's base adducts can be reduced to a stable secondary amine linkage using sodium borohydride or sodium cyanoborohydride. Fluorescent derivatives containing hydrazine groups (fluorescein thiosemicarbazide, Texas Red hydrazide, coumarin hydrazide, Lucifer Yellow CH, etc.) can also be coupled to terminal aldehydes produced by periodate oxidation (see Odom O.W., Jr., et al. (1980) Biochemistry 19, 5947-5954).

In the case of synthetic oligonucleotides, an aliphatic primary amino group can be readily incorporated at the 5'-terminal end of the molecule in

several ways.[1] A specially protected phosphoramidite derivative of 5'-amino-5'-deoxythymidine can be incorporated as the final residue (Smith, L.M. et al. (1985) Nucleic Acids Research 13, 2399-2412). Primary amines can also be incorporated into oligonucleotides via special phosphoramidite derivatives, which can also be added at the last step in the synthetic procedure. Two such derivatives are available commercially, one from Applied Biosystems, Inc. (Amino-link 1, Part Number 400498) and one from ChemGenes, Inc. in which the amino group is protected with the monomethoxytrityl group. Once a primary amine is incorporated into the oligonucleotide, a number of amine selective fluorescent derivatives (fluorescein-5-isothiocyanate, Texas Red, succinimdyl-1-pyrene butyrate, etc.) can be easily coupled to the primary amine group under mild conditions. Fluorescent lanthanides (Europium, Terbium, etc.) can be incorporated into oligonucleotides by first reacting the primary amine with diethylenetriaminepentacetic acid anhydride to produce a chelator for binding the lanthanides.

Enzyme labels of particular interest include: alkaline phosphatase, peroxidase, β-galactosidase, glucose oxidase, bacterial (FMN/NADPH) luciferases, and insect (ATP) luciferases to mention a few. Other labels include microperoxidases, functionalized heme derivatives, and other metal chelates with catalytic activity. Enzymes can be incorporated into RNA, DNA and mixed RNA-DNA probes using procedures similar to those discussed above for fluorescent derivatives. As an example, alkaline phosphatase can be incorporated into an oligonucleo-tide probe con-

[1] Oligonucleotides are generally synthesized from the 3'-to the 5'-end of the chain.

taining a 5'-terminal primary amine in the following manner. The primary amine containing oligonucleotide is reacted with an excess of the bifunctional reagent disuccinimidyl suberate (DSS is a highly specific reagent for coupling primary amines under mild conditions) to produce the mono adduct with the oligonucleotide. The mono adduct DSS-oligonucleotide can be easily purified. The mono adduct DSS-oligonucleotide is now reacted with alkaline phosphatase under mild conditions. The final product "alkaline phosphatase-DSS-5'-oligonucleotide" is purified by gel filtration and ion-exchange chromatography. The specifics of this coupling procedure are given in Jablonski, E. et al. (1986) Nucleic Acid Research 14, 6115-6128, which is incorporated by reference.

In this first embodiment of the CHA method, RNaseH is responsible for catalyzing the cleavage of the probe once it hybridizes to the target sequence. The enzyme is highly specific; it is a ribonuclease that cleaves only the RNA strand within an RNA:DNA duplex. The enzyme can be obtained from both eukaryotic and bacterial sources. The E. coli RNaseH enzyme, used in the examples, is a single polypeptide chain of about 17,500 molecular weight. The enzyme is very stable, the gene for the protein has been cloned and sequenced, and overproducing strains have been constructed which produce the enzyme in large amounts (Kanaya, S. and Crouch, R.J. (1983) Journal of Biological Chemistry 258, 1276-1281. For these reasons, its use is generally preferred. For CHA reactions carried out above a temperature of about 50°C, a more thermostable form of the enzyme is desirable. Such variants of the enzyme may be isolated from thermophilic organisms, or produced by

mutagenesis of E. coli RNaseH using recombinant DNA techniques (e.g. by the introduction of a suitably positioned disulfide bridge).

The optimal temperature for carrying out the CHA reaction is generally from about 5°C to about 25°C below the melting temperature of the probe:target duplex. This provides for a rapid rate of hybridization and high degree of specificity for the target sequence. Consider, for example, a DNA-RNA-DNA probe 26 residues in length in which the G plus C fraction is 50%. At a salt concentration of 0.1 M, the predicted melting temperature of the probe:target duplex, according to Equation 1, is about 55°C. The upper limit for the CHA reaction temperature is, therefore, about 50°C. Cleavage of the probe near the middle of the sequence generates two fragments each of which have melting temperatures of about 25°C. As noted above, for these fragments to dissociate rapidly from the target sequence, the CHA reaction temperature should be at least 10°C greater than this value, i.e. equal to or greater than 35°C. Hence, the optimal range for the CHA reaction temperature with this probe is from about 35°C to about 50°C.

The CHA reaction is usually carried out for about 15 minutes to about 1 hour. Generally, each molecule of the target sequence will turnover between 100 and 1000 times in this period, depending on the length and sequence of the probe, and the specific reaction conditions (see Example 1). For each copy of the target sequence present in the test sample 100 to 1000 molecules of the labeled probe will be cleaved by RNaseH. This level of amplification leads to an enormous increase in sensitivity compared to assays based on a stoichiometric hybridization for-

mat. Even higher levels of amplification can be obtained by allowing the CHA reaction to proceed longer.

During the CHA reaction, it is necessary to suppress cleavage of the probe, as well as the target sequence, by nonspecific nucleases. This is important for assays based on standard hybridization formats as well. Such nucleases are generally removed from the sample during the isolation of the DNA by heating or extraction procedures. A number of inhibitors of single-stranded ribonucleases such as vanadate, Inhibit-ACE (5 Prime - 3 Prime, Inc.) and RNAsin, a placental protein, do not affect the activity of RNaseH. In order to further protect the probe from nonspecific degradation, it is desirable to include such inhibitors during the CHA reaction.

Following the CHA reaction, it is necessary to determine the extent of cleavage of the probe. This is usually accomplished by separating the cleavage products bearing the label from the remaining uncleaved molecules of the probe. Methods for such separation include, but are not limited to:

(1) Electrophoresis of the CHA reaction products to separate the cleaved fragments from molecules of the intact probe on the basis of size.

(2) The use of strong acids, like trichloroacetic acid, to selectively precipitate the relatively large uncleaved labeled probes from the smaller cleaved fragments.

(3) The incorporation of a high affinity label into the probe, in addition to the reporter group, to provide for the separation of the uncleaved probes after the CHA reaction, using an affinity support.

(4) The use of a support material containing a complementary polynucleotide sequence to immobilize

the uncleaved labeled probe after the CHA reaction is carried out, leaving the cleaved fragments in solution.

In the first separation method, electrophoresis of the CHA reaction products through a gel matrix is used to separate the cleaved fragments bearing the label from the intact probe. The smaller fragments migrate through the gel more rapidly than the larger intact molecule. For probes up to about 100 nucleotides in length a polyacrylamide gel matrix is generally used. For longer probes, polyacrlamide, agarose or a composite gel made up of these two components is generally preferred. After the separation has been achieved, fragments bearing the reporter group label and the original probe are resolved into separate bands which can be readily localized and quantitated.

The second separation method involving strong acid precipitation is used mainly in those situations in which a long probe (greater than about 50 nucleotides in length) containing either radioisotope or fluorescent labels is cleaved via the CHA process into a number of relatively small labeled fragments (<10 nucleotides in length). Introduction of a strong acid (trichloroacetic, etc.) into the solution causes the intact probe to precipitate, while the smaller cleaved fragments remain in solution. The solution can be centrifuged or filtered to remove the precipitate. The supernatant containing the cleaved labeled fragments can now be quantitated. A more detailed discussion of this method is given in Example 1, below.

The third separation method involves a probe labeled with both a high affinity group and a reporter group, and an affinity support is used to carry

out the separation procedure. The probes used in this case can be either an RNA sequence or contain interspersed sequences not cleaved by RNaseH such as a mixed DNA-RNA-DNA sequence. The probe is designed with an affinity label (such as biotin, avidin, lectins, haptens, or antibodies) incorporated at or near one of the terminal positions of the probe. The reporter group label (enzyme, fluorophore, luminophore or radioisotope) is incorporated at or near the opposite terminal position. (With a long polynucleotide probe, many or all of the interspersed sequences not cleavable by RNaseH can be labeled with the reporter group). The cleavable portion of the probe sequence, thus, lies between the affinity group and the reporter group. A minimum of about 15 nucleotides is necessary between the two labels, with 20 to 50 nucleotides being more ideal. The same methods as discussed above for incorporating reporter group labels into probes are also used for the attachment of affinity labels. In this separation procedure it is also necessary to have support material which contains a corresponding affinity group which binds specifically and strongly to the affinity label incorporated into the probe. For example, if the affinity label biotin is incorporated into the probe, then the appropriate affinity support material should contain avidin or streptavidin. Many support materials such as glass beads, latex beads, cross-linked dextran (Sepharose) beads, etc. containing covalently linked avidin are commercially available.

After the CHA reaction is carried out, the support material is added to the sample. Molecules of the intact probe bind tightly to the support. Cleaved fragments bearing the reporter group but not the affinity label remain in solution and can be quantitated.

The fourth separation procedure is in principle very similar to that just described, but employs a nucleic acid sequence within the probe not cleaved during the CHA reaction as the affinity label. A complementary nucleic acid sequence is attached to a solid support and serves as a catcher sequence. When the cleavage of the probe during the CHA reaction is mediated by RNaseH, the sequence playing the role as the affinity label can be DNA, or an RNA sequence in which the sugar or phosphate groups are modified to prevent cleavage by RNaseH. The affinity label sequence is separated from the reporter group by a portion of the probe cleaved by RNaseH during the CHA reaction, for example

5'-DNA-RNA-DNA-(reporter group)

in which case the DNA sequence at the 5'-end of the probe, or a portion thereof, serves as the affinity label. An affinity label sequence of 20 to 50 nucleotides in length is most ideal, but longer sequences can also be used. Sequences less than 15 nucleotides are least useful. A portion of the affinity label sequence can be complementary to the target sequence or it can be a separate appendage attached to the probe. The latter is preferred because a generalized catcher sequence can then be used. In this case one set of catcher beads can be used with many different target specific probes.

A number of different chemical methods may be used to attach the catcher sequence to a solid support, see for example, Gilham, P.T. (1971) Methods of Enzymology, Vol. 21 (Grossman, L. and Moldave, K., eds.) pp. 191-197, Academic Press, New York; Potuzak, H. and Dean, P.D.G. (1978) Nucleic Acids Research 5, 297-303; Robertson, D.L. and Davidson, N. (1972) Biochemistry 11, 533-537; and Moss, L.G. et al.

(1981) Journal of Biological Chemistry 256, 12655-12658, which are incorporated herein by reference. It is also possible to synthesize the catcher sequence directly on the solid support which later will be used in the assay. As noted above, oligonucleotides are generally synthesized from the 3'- to the 5'-end of the chain. If the synthesis is carried out on a solid support, the first nucleotide from the 3'-terminal end is generally attached to the support via its 3'-hydroxyl group through an ester linkage. Conditions required to remove protecting groups on the nucleotide bases (treatment with concentrated ammonia at 55°C for 6 to 10 hours) cleave this linkage readily. If the oligonucleotide is to remain bound to the support the linkage to the 3'-hydroxyl group must be changed to one not cleaved under these conditions. Suitable functional groups through which the 3'-OH may be attached include ether linkages, phosphate triesters, phosphate diesters and aromatic carbamates. The functional group preferred for simplicity of synthesis is a β-cyanoethyl phosphtriester. The attachment is accomplished by the reaction of a 3'-β-cyanoethyl-N,N-diisopropyl phosphoramidite derivative of the first nucleotide (available commercially for all four bases) to a solid support containing free OH groups. Such supports can be prepared from a variety of materials including glass, polystyrene, latex or cross-linked dextran (Sepharose) beads, cellulose, or nylon or teflon membranes. After the first nucleotide is attached to the support, the remaining sequence is synthesized by the normal phosphoramidite method. At the culmination of the synthesis, the deblocking procedures convert the β-cyanolethyl phosphotriester to a phosphodiester through which the oligonucleotide

remains linked to the support. After a support material is prepared by any of the above procedures, the attached oligonucleotide can be extended, if desired, by a number of different enzymatic methods, thereby yielding longer nucleotide sequences bound to the support. Suitable procedures include, but are not limited to, addition of nucleotide residues to the fragment with DNA polymerase or terminal transferase, and ligation of polynucleotides to the initial sequence attached to the support using DNA or RNA ligase.

In each of the four separation methods just described, the CHA reaction is first carried out in solution. Alternatively, the labeled probe may be attached to a solid support, in which case cleavage of the probe during the CHA reaction releases fragments bearing the reporter group into solution (see Figure 3). The major advantage of this approach is the relative ease of the separation step after the CHA reaction is completed: the solid support material containing remaining molecules of the intact probe are simply removed from the sample by physical means such as filtration or centrifugation; the cleaved fragments bearing the reporter group are left in solution and can be quantitated. Suitable solid support materials include glass, polystyrene, latex or cross-linked dextran (Sepharose) beads, cellulose, or nylon or teflon membranes. Magnetized beads can also be used, in which case the beads are separated from the solution by the application of a magnetic field. The probe can be attached to the solid support either through a covalent or a noncovalent (e.g. biotin-avidin) linkage. A covalent linkage is preferred because of its greater stability. Suitable procedures for covalently attaching nucleotide se-

quences to solid supports by both chemical and enzymatic methods are described in the preceding paragraph.

When using the solid support CHA method, it is desirable to fist cut the sample DNA into smaller fragments so that the target sequence can readily diffuse to the surface of the bead to hybridize with the bound probe. This can be accomplished by shearing the DNA through mechanical forces (e.g. sonication) or by cleaving the DNA at specific sites using a restriction enzyme. Alternatively, a small fragment containing the target sequence can be synthesized with DNA polymerase using the sample DNA as a template. In this procedure an oligonucleotide used as a primer is first hybridized to the sample DNA upstream (i.e. 5') from the target sequence. The primer is then extended across the target sequence (copying the complementary strand) using DNA polymerase. The product:

5'-Primer-Target Sequence

is then brought into the CHA reaction. It is also possible to repeat the reaction with DNA polymerase a number of times to increase the number of copies of the target sequence. Such a procedure is carried out in the so-called polymerase chain reaction (Mullis, K.B. et al. (1986) EPA O 200 362; and Mullis, K.B. (1986) EPA 0 201 184). Since target amplification and the CHA method involve entirely separate steps in the assay procedure the two can be used in combination. This can be useful to detect a nucleic acid sequence present originally in the sample in extremely low copy number. For most diagnostic applications, however, the high degree of sensitivity of the CHA method alone is sufficient.

All of the above discussions on the use of the CHA method have centered around the detection of DNA target sequences. In some cases, it may be advantageous to detect RNA target sequences, such as ribosomal RNA sequences. This can be readily accomplished with any of the above CHA reaction procedures utilizing RNaseH by first synthesizing a complementary DNA (cDNA) copy of the original RNA sequence with the enzyme reverse transcriptase. In this procedure, an oligodeoxynucleotide primer is first hybridized to the sample RNA upstream (i.e. 5') of the target sequence. With the addition of reverse transcriptase and the four deoxynucleotide triphosphates (A, G, T and C), the primer is extended producing a cDNA copy of the target RNA. The product:

5'-Primer-cDNA Target Sequence

now forms the substrate for the CHA reaction. For a more detailed description of protocols for the use of reverse transcriptase see Maniatis, T., Fritsch, E.F., and Sambrook, J., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, New York, pp. 128-132, 1982, which is incorporated herein by reference.

A second embodiment of the CHA method will now be described in which a phosphorothioate derivative of the target sequence is used as the substrate and a complementary labeled DNA probe is cleaved by a restriction enzyme.

Restriction enzymes recognize a specific DNA sequence, generally 4 to 6 nucleotides in length, and cut both strands of the DNA duplex at that site. Several hundred of these enzymes are now known with many different recognition sequences. These enzymes generally cannot be used with the CHA method since

the target sequence, as well as the probe, would be cut, and therefore would be unable to recycle through the reaction pathway. However, if the phosphate groups of one of the two strands of a DNA duplex are phosphorothioates, rather than normal phosphate diesters, many restriction enzymes (e.g. Ava I, Ava II, Ban II, Hind II, Nci I, Pst I, and Pvu I) will cleave only the unmodified strand (Taylor, J.W. et al. (1985) Nucleic Acids Research 13, 8749-8764). The phosphorothioate sequence is not cut and hence can serve as a catalytic cofactor for cleavage of a complementary, labeled probe by a restriction enzyme. The target sequence to be used in such an assay must of course contain the recognition site for the restriction enzyme. The modified DNA sequence is prepared from a DNA template using DNA polymerase. If the sample is RNA, reverse transcriptase is used. These reactions are carried out as described above, except that deoxynucleotide α-thiotriphosphates are used as substrates instead of the normal deoxynucleotide triphosphates. All four of the deoxynucleotide α-thiotriphosphates are available commercially. Once the modified target sequence is prepared, the CHA reaction can be carried out using any of the assay formats described above. The same methods for synthesis and labeling of the probe are also applicable. Modified DNA or RNA sequences can be incorporated within the probe flanking the restriction enzyme site, but the recognition site itself must be composed of DNA.

All of the various embodiments and formats of the CHA method described thus far provide a linear amplification of the signal. Since the concentration of the target sequence remains constant, the velocity of the reaction remains unchanged, and the number of

molecules of the labeled probe that are cleaved increases linearly with time (see Example 1). As mentioned above, the turnover number, i.e. the number of molecules of the probe cleaved per target sequence per hour, generally ranges between 100 and 1000.

CHA reactions can also be carried out in novel manners that produce an exponential amplification, and therefore, offer an even greater sensitivity. In general, an exponential response is achieved by coupling two or more CHA reactions so as to produce additional copies of the target sequence. A number of different exponential CHA formats can be designed. One version of an exponential CHA format is shown in Figure 4. In this case, two sets of probe sequences are immobilized on solid supports. The first probe sequence on support I contains a cleavable sequence (T') complementary to the original target sequence, and a non-cleavable sequence (X). The sequence (X) is complementary to a cleavable sequence (X') immobilized on support II. In the first step of the process, target DNA from the sample hybridizes to the complementary cleavable sequence (T') on support I. Once hybridized to the target sequence, (T') is cut, releasing the sequence (X) from the support. (X) then diffuses from the support and hybridizes with the cleavable sequence (X') on support II. This cannot occur until (X) is released from the support. Sequences (X) and (X') while immobilized on two separate macroscopic supports cannot hybridize with one another. Once the duplex between (X) and (X') is formed, (X') is cleaved. This releases a second copy of the target sequence which can in turn react with support I and further propagate the reaction. Thus, one cycle of the CHA reaction yields two copies

of the target sequence; the second cycle yields four copies of the target sequence; the third cycle yields eight copies; etc. In the configuration shown, one reporter group label is released from support II for each molecule of target that is generated in the reaction. It can be shown that the amount of the free label released from support II is given by the following formula:

$$L = (A/2) \times (e^{kt} + e^{-kt}) - A \qquad (2)$$

where A is the initial number of copies of the target present in the sample, k is the square root of the product of the turnover numbers for the two separate CHA reactions, and t is time. The first term of the expression ($A/2 \times e^{kt}$) rapidly becomes dominant. The amount of free label increases exponentially with time, conferring an extremely high degree of sensitivity to the assay. If the turnover number for each of the CHA reactions is only 1 per minute, a single copy of the target sequence will cause 2.4×10^8 labels to be released into solution within 20 minutes. This number can be easily detected with simple fluorescent or enzyme labeling groups. Turnover numbers much greater than 1 per minute can be achieved (see Example 1). Coupled CHA assay systems, thus, provide the capability for single copy target detection with a reaction time of only a few minutes.

Finally, we describe a competition assay based on the CHA method to detect an RNA target sequence. In this case, both a DNA molecule and a labeled probe are added to the sample to be analyzed (Figure 5). The probe sequence (A') may be composed entirely of RNA or contain interspersed sequences that cannot be cleaved by RNaseH. The DNA fragment contains within it a sequence (A) complementary to the labeled probe. In the example shown, the central portion of the

probe is RNA, and it is flanked on either side by DNA sequences. The RNA:DNA duplex formed when (A) and (A') hybridize together forms a substrate for RNaseH. RNaseH can cleave the labeled probe, releasing the DNA fragment intact which serves as a catalyst to further propagate the reaction. This would occur if the target RNA sequence were not present within the sample (Figure 5b). If the target RNA sequence were present, this reaction would be inhibited (Figure 5a). The target sequence contains the sequence (A'), as well as an adjacent sequence (B') further complementary to the DNA fragment. The DNA fragment thus forms a more stable duplex with the target sequence than with the probe. This enables the target molecule to compete very effectively for the binding of the DNA fragment, and thereby blocks the cleavage of the labeled probe by RNaseH (Figure 5a). This competitive inhibition of the cleavage of the labeled probe is the basis for the assay. The sensitivity of the method may be further enhanced by preventing the RNA-DNA duplex formed with the target sequence from being degraded by RNaseH. This may be accomplished by suitably modifying the DNA fragment. We have found that a variety of modifications of the phosphate backbone of the DNA strand will prevent cleavage of the complementary RNA strand by RNaseH. Such modifications include, but are not limited to, substitution of the normal phosphate diesters with alkyl or aryl phosphotriesters, or with alkyl or aryl phosphonates. These modifications would be incorporated into the DNA molecule used in the assay throughout region B, and in all of region A except for that portion which hybridizes to the central RNA sequence of the probe. If this unmodified sequence within region A is precisely complementary to the RNA

sequence within the probe, but contains one or more mismatches with respect to the target sequence, only the probe will be cleaved by RNaseH. The mismatch(es) within the duplex formed between the DNA strand and the target molecule disrupt the helix, thereby, preventing cleavage of the target strand within this region by RNaseH. Cleavage of the target strand elsewhere is prevented by the modifications introduced into the phosphate backbone of the DNA molecule. The mismatch(es) would decrease the stability of the duplex formed between the target sequence and the DNA molecule somewhat, but, due to the additional sequence complementary in region B, the stability of the duplex would remain much greater than that formed with the labeled probe.

The following examples are offered to further illustrate, but not limit, the process, product and techniques of the invention and serve to establish the unique features of the invention which form the basis of its utility.

EXAMPLE 1
Demonstration of Catalytic Hybridization Amplification Mediated by RNaseH

The following example demonstrates the principle of the catalytic hybridization amplification method. The target sequence in this case was oligodeoxythymidylic acid of 12 to 18 nucleotide residues in length (oligo dT_{12-18}, purchased from Pharmacia). The molecule serving as the probe was polyriboadenosine labeled with tritium (Amersham). The specific activity of the poly rA used in the study was 574 mCi/nucleotide residue; the chain length varied from about 40 to 140 residues. In the presence of the target sequence, RNaseH catalyzes the cleavage of the

labeled poly rA probe. Oligo dT is then released from the duplex intact and recycles repeatedly through the reaction as outlined in Figure 2.

Cleavage of the poly rA probe catalyzed by RNaseH was compared under three conditions: in the absence of oligo dT; with an equivalent concentration of oligo dT (based on nucleotide residues); and in the presence of 1/1000 the concentration of oligo dT. Each reaction was conducted in a final volume of 100 ul containing 44 picamoles of poly rA (by nucleotide units), 50 mM Tris-HCl buffer pH 8.0, 20 mM KCl, 9 mM $MgCl_2$, 1 mM 2-mercaptoethanol and 25 ug of bovine serum albumin. The reaction was initiated by the addition of 2.3 units of E. coli RNaseH (Bethesda Research Laboratories) and was maintained at 30°C. Aliquots of the reaction mixture were removed at 30 minutes and added to 0.2 ml of a carrier DNA solution (0.5 mg/ml of salmon sperm DNA, 0.1 M sodium pyrophosphate and 1 mM ethylenediamine tetraacetic acid) and 0.3 ml of 10% trichloroacetic acid. Trichloroacetic acid causes the precipitation of higher molecular weight nucleic acids over about 10 nucleotides in length. Smaller cleavage products, i.e. mononucleotides and short oligonucleotides, remain in solution. After incubation of the samples on ice for 20 minutes, the precipitated nucleic acids were pelleted by centrifugation at 16,000 X g for 15 minutes. The supernatants containing the cleaved fragments were carefully removed, mixed with scintillation cocktail and counted for radioactivity.

In the control sample without oligo dT, less than 0.5% of the poly rA was cleaved in the course of 90 minutes. In the presence of an equivalent concentration of oligo dT nearly 100% of the poly rA was cleaved to trichloroacetic acid soluble fragments

within 30 minutes. Cleavage of poly rA in the presence of 1/1000 the amount of oligo dT was linear over 90 minutes, at which time 21% of the poly rA was cleaved to trichloroacetic acid soluble fragments. Since oligo dT was present in a concentration sufficient to hybridize to only 0.1% of the poly rA sequence, each molecule of the oligo dT must have recycled through the reaction, on average, at least 210 times (2.3 times/minute). Since cleavage of the poly rA strand is in some cases nonproductive, i.e. does not lead to the release of labeled fragments small enough to remain in solution in the presence of trichloroacetic acid, the true turnover number must be even higher. In diagnostic tests based on the CHA method, turnover numbers exceeding 100/hour provide a great enhancement in sensitivity compared to assays based on a stoichiometric hybridization format.

EXAMPLE 2

Site-Specific Cleavage of β-Globin mRNA by RNaseH in the Presence of a Complementary DNA Target Sequence

The following is an important example in which the specificity of strand cleavage by RNaseH is clearly demonstrated. RNaseH was shown to cleave β-Globin mRNA at one unique site when a mixture of both α- and β-globin mRNAs was hybridized to a short 25-residue DNA target sequence complementary to a portion of the β-globin mRNA.

The oligodeoxynucleotide 5'-TGTCCAAGTGATTCAG-GCCATCGTT (CODMB-25) was synthesized by the phosphoramidite method using a Beckman automated DNA synthesizer. It is complementary to the nucleotide sequence of mouse β-globin mRNA from residue 269 to 293. Mouse globin mRNA was isolated from reticulo-

cytes obtained from mice rendered anemic by treatment with phenylhydrazine as described by Goosens and Kan (Methods in Enzymology (Antonini, E., Rossi-Bernardi, L. and Chiancone, E., eds.) Vol. 76, pp. 805-817, 1981). Two micrograms of the RNA was added to 1.0 nanomoles of CODMB-25 in a final volume of 25 microliters containing 10 mM Tris-HCl buffer at pH 7.5, 5 mM $MgCl_2$, 25 mM NaCl and 1 mM dithiothreitol. The concentration of the oligonucleotide is approximately 100-fold greater than the mRNA on a molar basis. The reaction was initiated by the addition of 2 units of E. coli RNaseH, and allowed to proceed for 30 minutes at 37°C. The reaction was stopped by extraction of the sample with a 1:1 mixture (volume: volume) of chloroform and phenol. The RNA species remaining after digestion with RNaseH were then isolated by precipitation with ethanol and analyzed on Northern blots.

The RNA samples were reacted with 1 M glyoxal in a 1:1 volume/volume mixture of 10 mM sodium phosphate buffer at pH 6.5 and dimethylsulfoxide for 1 hour at 50°C, and then electrophoresed on a 1.5% agarose gel. The RNAs were transferred from the gel to a Gene Screen Plus filter (DuPont-NEN) by capillary blotting according to the method described by the manufacturer. Prehybridization of the filter was done in 1% sodium dodecylsulfate, 1 M NaCl, 10% dextran sulfate, 50 mM sodium phosphate pH 6.5 for 2 hours at 35°C. The filter was then hybridized with either an α- or β-globin specific oligonucleotide probe. The probes were 25 residues in length and hybridize to the very 5'-end of the respective RNAs. Hybridization was conducted in 5 ml of the prehybridization buffer containing 200 microgram/ml denatured salmon sperm DNA plus 2 X 10^6 counts/min of the probe,

radiolabeled with ^{32}P at the 5'-end of the molecule, for 24 hours at 35°C. After the hybridization, the filter was washed in the following order: once in 2X SSC (1X SSC = 0.15 M NaCl, 0.015 M sodium citrate) for 5 minutes at room temperature, twice in 2X SSC plus 1% sodium dodecylsulfate for 30 minutes at 35°C, and once in 0.1X SSC for 5 minutes at room temperature. The filter was blotted dry and exposed to Kodak XAR-5 film for autoradiography.

In the absence of oligonucleotide, no cleavage of either α- or β-globin mRNA occurred during the reaction with RNaseH. The α- or β-globin specific probes each hybridized to a single band on the Northern blots corresponding to the full length mRNA. The reaction in the presence of CODMB-25 resulted in essentially complete cleavage of the β-globin mRNA. The size of the cleavage product detected with the 5'-β-globin specific probe was exactly that predicted for scission of the mRNA at the site of the RNA:DNA duplex, i.e. 280 +/- 20 nucleotides. No other cleavage products were observed. In addition, no cleavage of the α-globin mRNA was evident using the α-globin specific oligonucleotide probe. RNaseH, thus, has the requisite specificity for use in diagnostic tests based on the CHA method. An RNA sequence used as probe will be cleaved by the enzyme if, and only if, it encounters the complementary target DNA sequence within the sample; and free single stranded molecules of the probe not hybridized to the target sequence will not be cut.

EXAMPLE 3

Detection of Cytomegalovirus Using a CHA Method

Cytomegalovirus (CMV) is a large double-stranded DNA virus. Infection with CMV is clinically very important in immunocompromised patients. It is the most prevalent serious infectious disease complication following renal transplantation.

The following illustrates the use of the CHA method to detect CMV. The probe used is a CMV specific sequence 25 nucleotides in length derived from the major immediate early gene:

5'-TCTTGGCAGAGGACTCCATCGTGTC

The central six residues are RNA; the flanking sequences are composed of DNA. The probe is linked to latex beads at the 3'-end of the sequence. Alkaline phosphatase is attached at the 5'-terminal position. DNA is isolated from blood, urine and sputum samples obtained from the patient. The assay for CMV in each of these samples is carried out by the CHA method shown in Figure 3. Cleavage of the labeled probe is mediated by E. coli RNaseH (5 units/100 microliters). The CHA reaction is carried out for a period of 45 minutes at a temperature of about 40°C. RNasin (50 units/100 microliters) is added to the reaction mixture to suppress nonspecific cleavage of the probe. If CMV is present within the sample, the target sequence will hybridize to the probe, and the probe will be cleaved by RNaseH. Repeated cycling of the target sequence through the reaction pathway enables each copy of the target sequence to capture and promote the cleavage of many molecules of the labeled probe. Following the CHA reaction, the extent of cleavage of the probe is quantitated by measuring the amount of the reporter group, i.e.

alkaline phosphatase, released into solution. The beads are first removed from the sample by filtration or centrifugation. The alkaline phosphatase label on the probe fragments in solution is then measured using either a colorimetric or fluorescent producing reagent detection system. The enzyme can be assayed spectrophotometrically at 405 nm by following the hydrolysis of the substrate p-nitrophenyl phosphate to the colored product p-nitrophenol at pH 8.5. Alternatively, the enzyme can be measured fluorometrically using the substrate 4-methylumbelliferyl phosphate. The fluorescent excitation and emission wavelengths used for the assay are 363 nm and 447 nm, respectively. The final color or fluorescent signal provides a measure of the amount of CMV present in the original sample.

In summary, the results of the studies presented demonstrate the utility of the CHA method as a means for the detection of specific nucleic acid sequences. In such assay systems, the target sequence serves as a cofactor for a catalytic reaction in which a complementary, labeled nucleic acid probe is cleaved. Each molecule of the target sequence is thus able to capture and promote the cleavage of multiple copies of the probe. This results in a large increase in the level of sensitivity of the method compared to current diagnostic tests, all of which are based on a stiochiometric hybridization reaction in which the target sequence is able to bind one, and only one, molecule of the probe. As in the examples described, cleavage of the probe must occur only when hybridized to the target sequence. The target sequence is then released from the duplex intact and repeatedly recycles through the reaction pathway. Formats are provided for use of the CHA

method in both direct and competition assays. In general, a detection system based on a single CHA reaction provides a linear amplification of signal with time. By coupling two or more CHA reactions together in an appropriate manner, an exponential amplification of signal can be achieved, giving rise to an even greater level of sensitivity.

It therefore can be seen that the invention accomplishes all of the objectives heretofore stated.

What is claimed is:

1.

A method of detecting nucleic acid sequences in which said sequences serve as a cofactor for a catalytic reaction in which a complementary labeled nucleic acid probe is cleaved and detected, said method comprising:

hybridizing of a target sequence to a labeled nucleic acid probe to provide a probe:target sequence duplex;

cleaving of the labeled probe within the probe:target sequence duplex to release the target sequence intact;

recycling of the target sequence repeatedly through the reaction pathway; and

detecting the extent of cleavage of the labeled probe.

2.

The method of claim 1 wherein said detection method is used for the diagnosis of a disease or genetic trait in man or other species.

3.

The method of claim 1 wherein said detection method is used to identify a nucleic acid sequence within a system used in an industrial or agricultural process.

4.

The method of claim 1 wherein said nucleic acid probe is labeled with a radioactive, fluorescent, electrochemical, luminescent or enzyme marker, or other reporter group to permit its detection.

5.

The method of claim 1 wherein the extent of cleavage of the probe is determined by electrophoret-

ic separation of the smaller cleavage fragments carrying the labeling group from the remaining molecules of the intact probe.

6.

The method of claim 1 wherein the extent of cleavage of the probe is determined by using strong acids such as trichloroacetic acid to selectively precipitate the larger uncleaved molecules of the probe, leaving the smaller cleaved fragments bearing the label in solution.

7.

The method of claim 1 wherein the probe is modified with both a high affinity label and a reporter group, and in which a solid support to which a corresponding affinity group is attached is used to adsorb molecules of the intact probe, leaving the cleaved fragments bearing the label in solution, said method comprising the additional steps of:

adding of solid support material to the sample following the hybridizing and cleaving reaction;

associating of the high affinity label attached to the probe with the corresponding affinity group on the solid support;

separating of the solid support from the solution containing the cleaved fragments of the probe bearing the reporter group; and

detection of the reporter groups in solution.

8.

The method of claim 7 in which said solid support is composed of latex, polystyrene, cross-linked dextran or glass beads, cellulose, or a teflon or nylon membrane.

9.

The method of claim 7 in which said solid support is magnetized and the solid support is sepa-

rated from the bulk solution by the application of a magnetic field.

10.

The method of claim 7 wherein said affinity label is a nucleic acid sequence not cleaved in the hybridization and cleavage reaction and the corresponding affinity group attached to the solid support is a complementary nucleic acid sequence.

11.

The method of claim 1 wherein said labeled probe is attached to a solid support during the hybridization and cleavage reaction and cleaved fragments bearing the reporter group are released into solution.

12.

The method of claim 11 in which said solid support is composed of latex, polystyrene, cross-linked dextran or glass beads, cellulose, or a teflon or nylon membrance.

13.

The method of claim 11 in which said solid support is magnetized and the solid support is separated from the bulk solution by the application of a magnetic field.

14.

The method of claim 1 for detection of a DNA sequence in which cleavage of the labeled probe is mediated by an enzyme or other catalyst with RNaseH activity.

15.

The method of claim 14 in which nonspecific cleavage of the labeled probe is suppressed by the use of single-stranded ribonuclease inhibitors such as vanadate, RNAsin, or Inhibit-ACE.

16.

The method of claim 14 wherein the labeled probe is cleaved by E. coli RNaseH.

17.

The method of claim 14 wherein the labeled probe is cleaved by a heat stable form of RNaseH.

18.

The method of claim 17 in which said heat stable RNaseH is from a thermophilic organism or is developed by site-directed mutagenesis from a non-heat stable form of RNaseH.

19.

The method of claim 14 wherein the labeled probe is an oligonucleotide from about 15 to about 50 nucleotides in length.

20.

The method of claim 19 wherein said oligonucleotide is composed entirely of RNA.

21.

The method of claim 19 wherein said oligonucleotide contains one or more interspersed sequences that are not cleavable by RNaseH.

22.

The method of claim 19 wherein the RNA sequences within the probe cleavable by RNaseH vary from about 3 to about 15 nucleotides in length.

23.

The method off claim 19 wherein an oligonucleotide composed of a DNA-RNA-DNA sequence is used as the probe.

24.

The method of claim 23 wherein said DNA-RNA-DNA probe is used to detect a point mutation such as occurs in sickle cell disease.

25.

An oligonucleotide 15 to 50 nucleotides in length containing an RNA sequence and one or more additional sequences not cleavable by RNaseH.

26.

The oligonucleotide in claim 25 in which said RNA sequence is from about 3 to about 15 nucleotides in length.

27.

The oligonucleotide of claim 26 labeled at or near the 5'- or 3'-end of the chain with a radioactive, fluorescent, electrochemical, luminescent or enzyme marker, or other reporter group.

28.

The oligonucleotide of claim 27 further modified with a high affinity group at or near the opposite end of the chain from which said reporter group is attached.

29.

The oligonucleotide of claim 27 covalently linked to a solid support at or near the opposite end of the chain from which said reporter group is attached.

30.

The oligonucleotide of claim 29 wherein said solid support is composed of latex, polystyrene, cross-linked dextran or glass beads, cellulose, or a teflon or nylon membrane.

31.

The oligonucleotide of claim 29 wherein said solid support is magnetized.

32.

The oligonucleotide of claim 25 in which said RNA sequence is from about 4 to about 8 nucleotides in length and contains one or more mismatches with respect to a target DNA sequence.

33.

The oligonucleotide of claim 25 in which said additional sequences are composed of RNA modified either at the phosphate or sugar groups to prevent cleavage by RNaseH.

34.

The oligonucleotide of claim 33 wherein said phosphate groups are modified as either alkyl or aryl phosphotriesters, hydrogen phosphonates, alkyl or aryl phosphonates, alkyl or aryl phosphoramidates, phosphorothioates, or phosphoroselenates.

35.

The oligonucleotide of claim 33 wherein said sugar groups are modified by conversion of the 2'-OH group of the ribose ring to an alkyl or aromatic ether.

36.

The oligonucleotide of claim 25 in which said additional sequences are composed of DNA.

37.

The method of claim 14 wherein the labeled probe is a polynucleotide from about 50 to several thousand nucleotides in length.

38.

The method of claim 37 wherein said polynucleotide is composed entirely of RNA.

39.

The method of claim 37 wherein said polynucleotide contains one or more interspersed sequences that are not cleavable by RNaseH.

40.

The method of claim 37 wherein the RNA sequences within the probe cleavable by RNaseH are 3 nucleotides in length or greater.

41.

A polynucleotide from about 50 to several thousand nucleotides in length composed of RNA plus one or more interspersed sequences that are not cleavable by RNaseH.

42.

The polynucleotide of claim 41 in which the stretches of RNA sequence are 3 nucleotides in length or greater.

43.

The polynucleotide of claim 42 in which one or more of the interspersed sequences not cleavable by RNaseH is labeled with a radioactive, fluorescent, electrochemical, luminescent or enzyme marker, or other reporter group.

44.

The polynucleotide of claim 43 further modified with a high affinity group at or near the 5'- or 3'-end of the chain.

45.

The polynucleotide of claim 43 covalently linked to a solid support at or near the 5'- or 3'-end of the chain.

46.

The polynucleotide of claim 45 wherein said solid support is composed of latex, polystyrene, cross-linked dextran or glass beads, cellulose, or a teflon or nylon membrane.

47.

The polynucleotide of claim 45 wherein said solid support is magnetized.

48.

The polynucleotide of claim 41 in which said interspersed sequences are composed of RNA modified either at the sugar or phosphate groups to prevent cleavage by RNaseH.

49.

The polynucleotide of claim 48 wherein said phosphate groups are modified as either alkyl or aryl phosphotriesters, hydrogen phosphonates, alkyl or aryl phosphonates, alkyl or aryl phosphoramidates, phosphorothioates, or phosphoroselenates.

50.

The polynucleotide of claim 48 wherein said sugar groups are modified by conversion of the 2'-OH group of the ribose ring to an alkyl or aromatic ether.

51.

The polynucleotide of claim 41 in which said interspersed sequences are composed of DNA.

52.

The method of claim 1 for detection of an RNA sequence in which the sample RNA is first used as template for the synthesis of a complementary DNA copy of the target sequence with reverse transcriptase, said method comprising:

hybridizing of a DNA primer to the sample RNA 5' to the target sequence;

extending of the primer with reverse transcriptase across the target sequence to produce a complementary DNA copy of the target sequence; and

detecting of the complementary DNA copy with a CHA method.

53.

The method of claim 52 in which the hybridization and cleavage method is based on cleavage of a labeled probe by an enzyme or other catalyst with RHaseH activity.

54.

The method of claim 52 in which a phosphorothioate derivative of the complementary DNA

copy of the target sequence is synthesized with reverse transcriptase using deoxynucleotide α-thio-triphosphates as substrates for the reaction such that the phosphate groups of the DNA strand synthesized are phosphorotioates rather than the naturally occurring phosphodiesters.

55.

The method of claim 54 in which the hybridization and cleavage method is based on cleavage of a labeled probe by an enzyme or other catalyst with RNaseH activity.

56.

The method of claim 54 in which the hybridization and cleavage method is based on cleavage of a labeled DNA probe by a restriction enzyme, said method comprising:

hybridizing of the phosphorothioate derivative synthesized to a complementary labeled DNA probe;

cleaving of the labeled probe within the probe:phosphorotioate DNA duplex by a restriction enzyme at its recognition site to release the phosphorothioate strand intact;

recycling of the phosphorothioate strand repeatedly through the reaction pathway; and

detecting of the extent of cleavage of the probe following the reaction.

57.

The method of claim 1 for detecting of a DNA sequence in which a copy of the target sequence is first synthesized with DNA polymerase using the sample DNA as a template, said method comprising:

denaturing the sample DNA;

hybridizing of a DNA primer to the sample DNA 5' to the target sequence;

extension of the primer with DNA polymerase across the target sequence to produce a copy of the complementary DNA strand;

detecting of the DNA copy with a CHA method.

58.

The method of claim 57 in which the CHA method is based on cleavage of a labeled probe by an enzyme or other catalyst with RNaseH activity.

59.

The method of claim 57 in which a phosphorothioate derivative of the copy of the target sequence is synthesized with DNA polymerase using deoxynucleotide α-thiotriphosphates as substrates for the reaction such that the phosphate groups of the DNA strand synthesized are phosphorothioates rather than the naturally occurring phosphodiesters.

60.

The method of claim 59 in which the hybridization and cleavage method is based on cleavage of a labeled probe by an enzyme or other catalyst with RNaseH activity.

61.

The method of claim 59 in which the hybridization and cleavage method is based on cleavage of a labeled DNA probe by a restriction enzyme, said method comprising:

hybridization of the phosphorothioate derivative synthesized to a complementary labeled DNA probe;

cleavage of the labeled probe within the probe:phosphorothioate DNA duplex by a restriction enzyme at its recognition site;

release of the phosphorothioate strand intact;

recycling of the phosphorothioate strand repeatedly through the reaction pathway; and

detection of the extent of cleavage of the probe following the reaction.

62.

The method of claim 1 in which two or more hybridization and cleavage reactions are coupled to achieve an exponential amplification of signal.

63.

The method of claim 62 in which one or more of the hybridization and cleavage reactions are based on cleavage of a labeled probe by an enzyme or other catalyst with RNaseH activity.

64.

The method of claim 62 in which one or more of the hybridization and cleavage reactions are based on the use of a phosphorothioate derivative as the substrate and a complementary DNA sequence is cleaved by a restriction enzyme at its recognition site.

65.

A method for detection of an RNA sequence using a competition assay, said method comprising:
hybridization of the sample RNA to a DNA molecule complementary to the target sequence;
addition of a labeled RNA probe complementary to said DNA molecule plus an enzyme or other catalyst with RNaseH activity to the test sample;
incubation of the test sample to allow the hybridizing and cleaving reaction to proceed; and
measuring of the extent of cleavage of the RNA probe.

66.

The method of claim 65 in which the labeled probe contains one or more interspersed sequences not cleavable by RNaseH.

67.

The method of claim 65 wherein said DNA molecule contains one segment complementary to both the target

RNA sequence and the labeled probe and a second segment complementary to just the target sequence.

68.

The method of claim 67 wherein the DNA molecule contains a sequence at least 4 nucleotides in length in which the phosphate groups are unmodified, and the remaining phosphate groups are modified so as to prevent cleavage of the complementary target and probe RNA sequences by RNaseH.

69.

The method of claim 68 wherein the phosphate groups are modified as alkyl or aryl phosphotriesters, alkyl or aryl phosphonates or alkyl or aryl phosphoramidates.

70.

The method of claim 65 in which said DNA molecule is complementary to the labeled probe but forms a duplex with the target RNA sequence with one or more mismatches, such that cleavage of the target RNA sequence by RNaseH is inhibited.

STANDARD STOICHIOMETRIC HYBRIDIZATION ASSAY TO DETECT DNA SEQUENCES

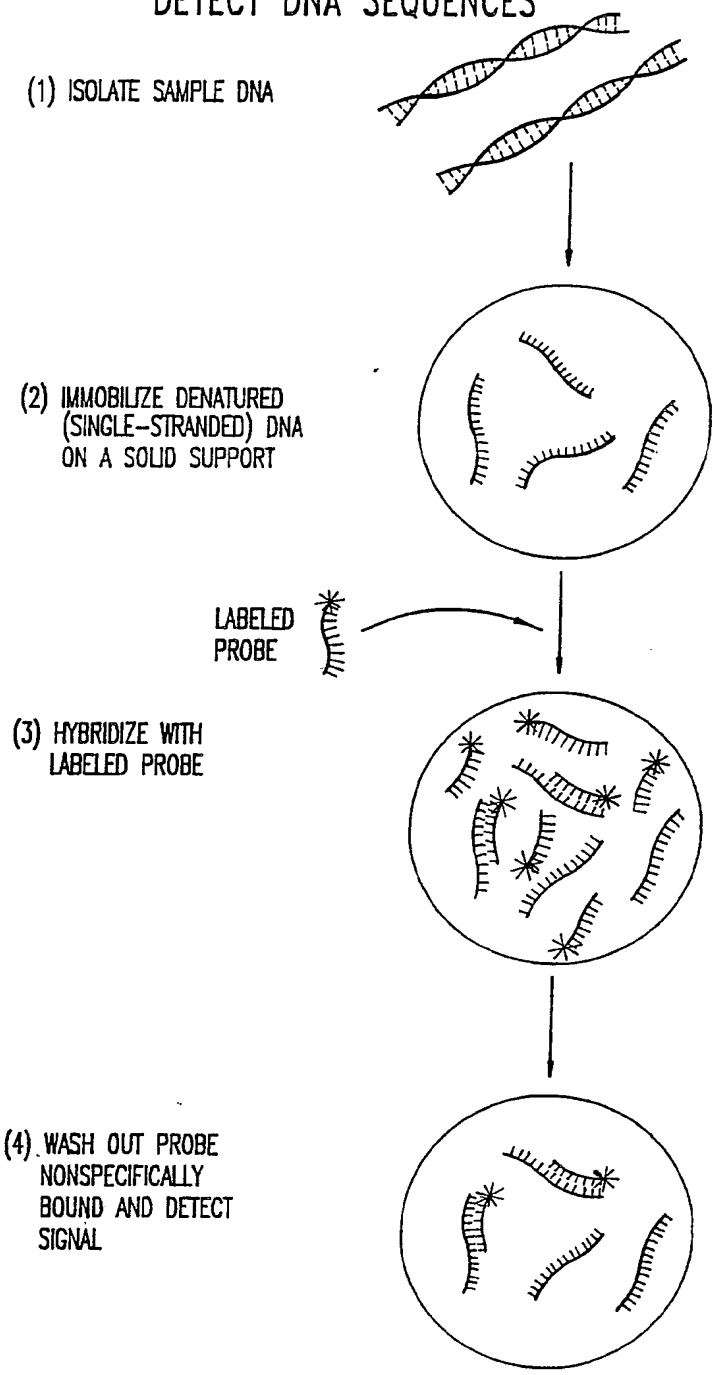


FIG.1

IDT Exhibit 1020.2432

DETECTION OF A SPECIFIC DNA SEQUENCE USING THE CHA METHOD

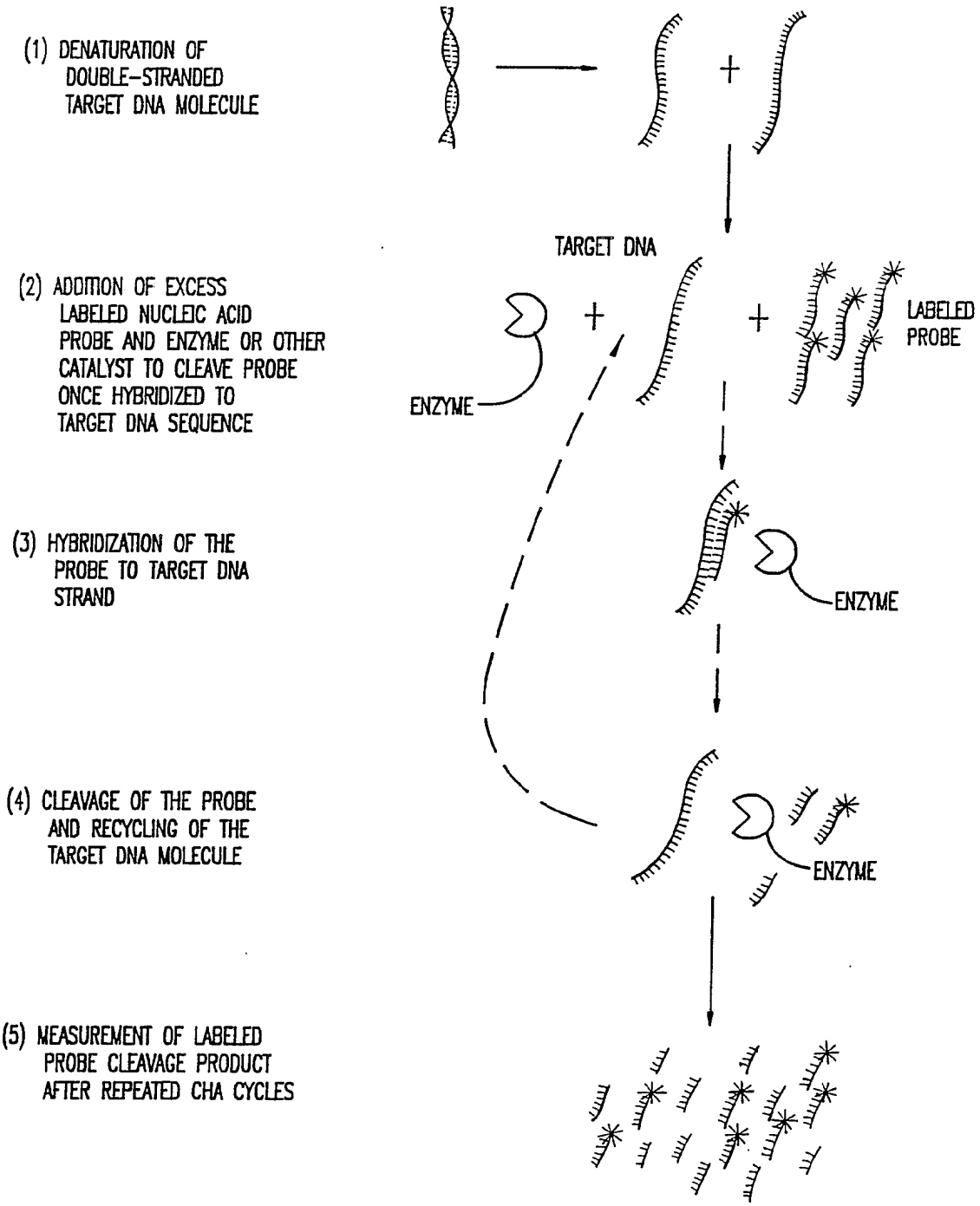


FIG.2

IDT Exhibit 1020.2433

CATALYTIC HYBRIDIZATION AMPLIFICATION IN SOLID PHASE FORMAT

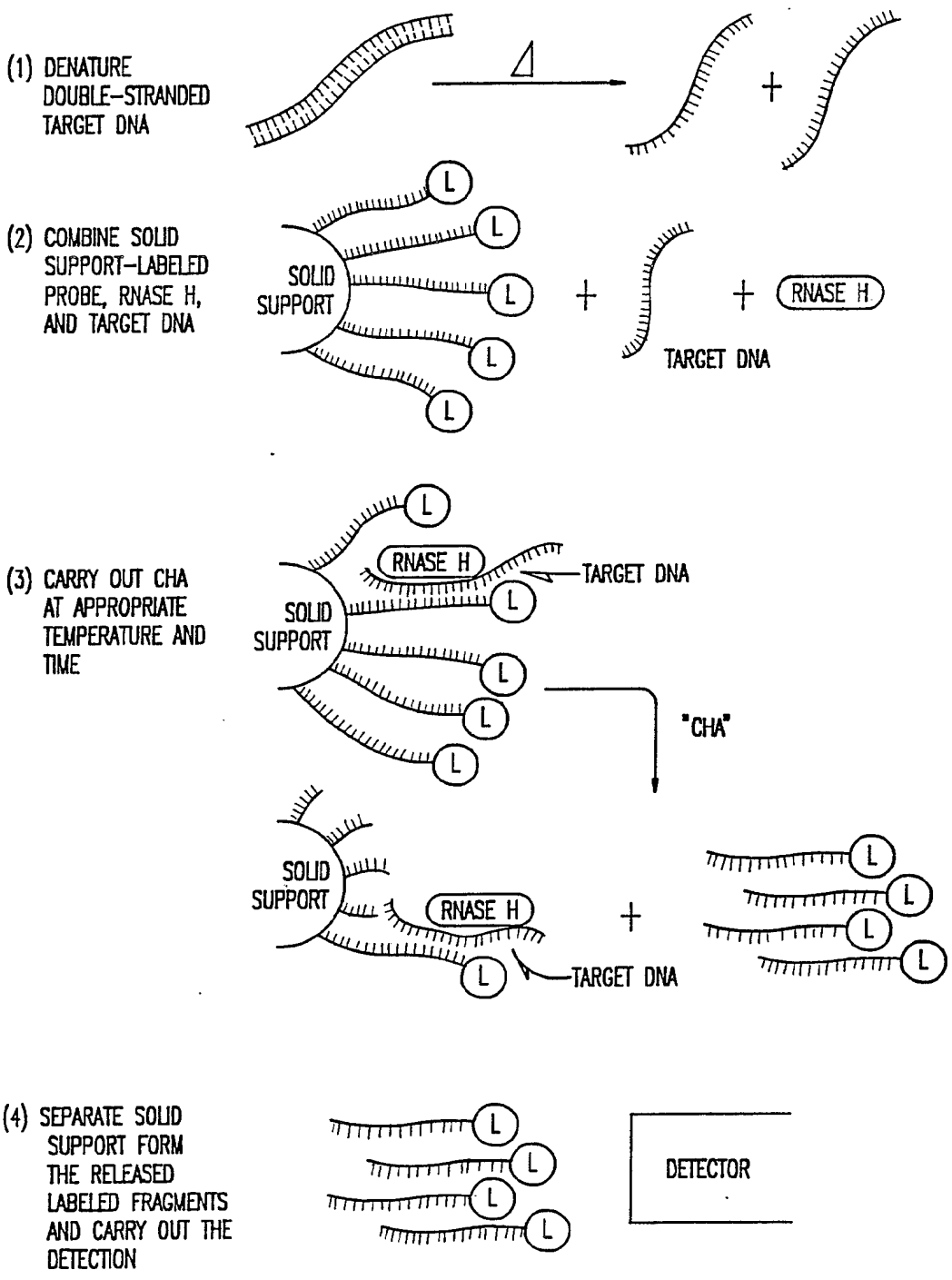
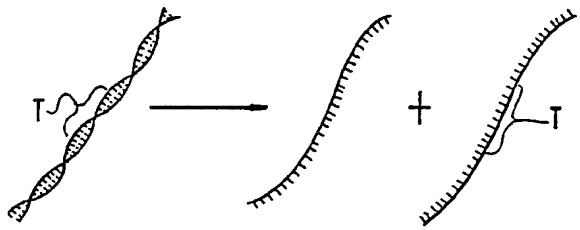


FIG.3

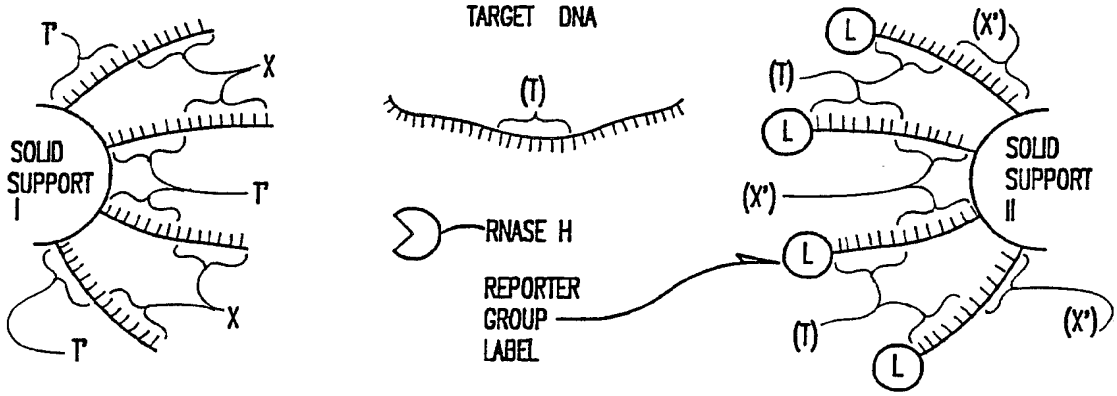
IDT Exhibit 1020.2434

EXPONENTIAL CATALYTIC HYBRIDIZATION AMPLIFICATION SYSTEM

(1) DENATURATION OF SAMPLE DNA TO FORM SINGLE-STRANDED SPECIES.
UNIQUE TARGET SEQUENCE IS REPRESENTED BY (T).



(2) ADDITION OF TARGET DNA AND RNASE H TO SOLID SUPPORT SYSTEM I AND II.
SOLID SUPPORT I CONTAINS PROBE WITH COMPLEMENTARY CLEAVABLE (T') AND (X) SEQUENCE.
SOLID SUPPORT II CONTAINS COMPLEMENTARY CLEAVABLE (X') AND LABELED (T) SEQUENCE.



(3) TARGET DNA SEQUENCE (T) HYBRIDIZES TO COMPLEMENTARY CLEAVABLE (T') ON SOLID SUPPORT I.

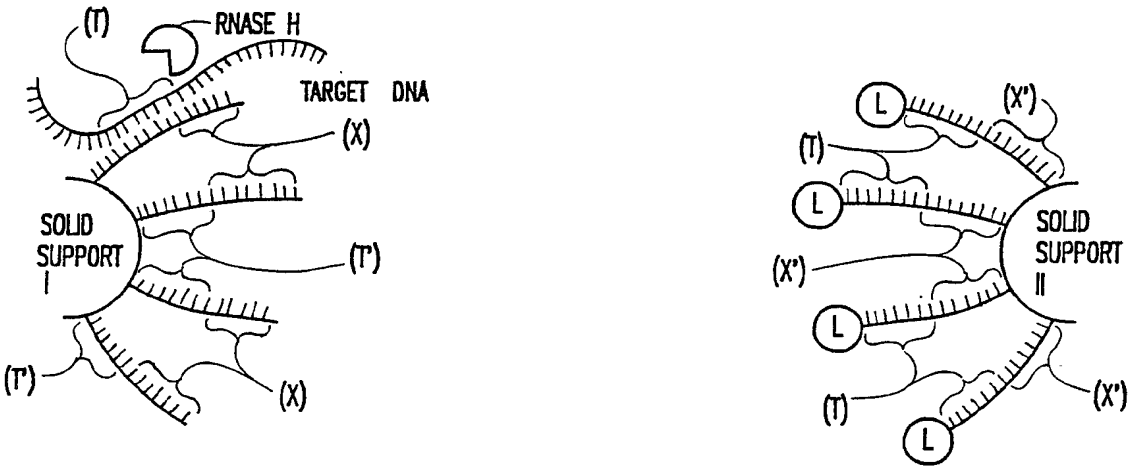
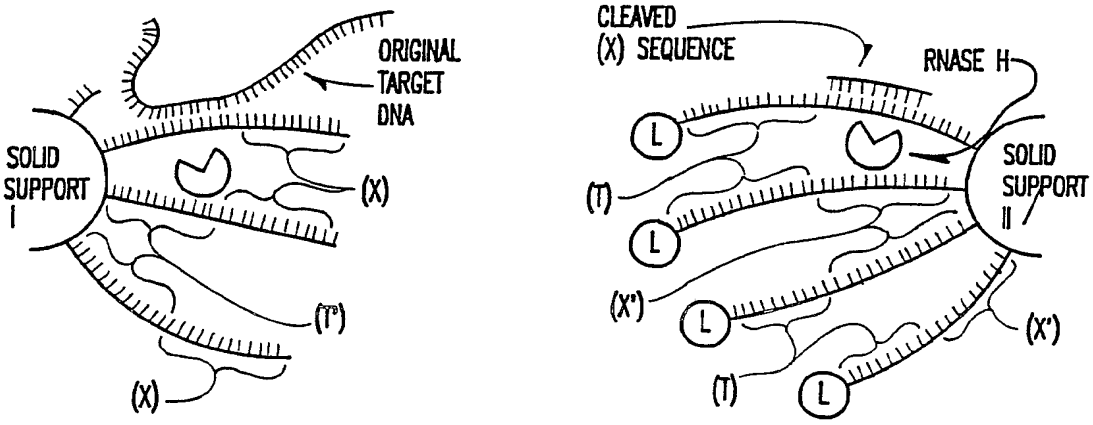


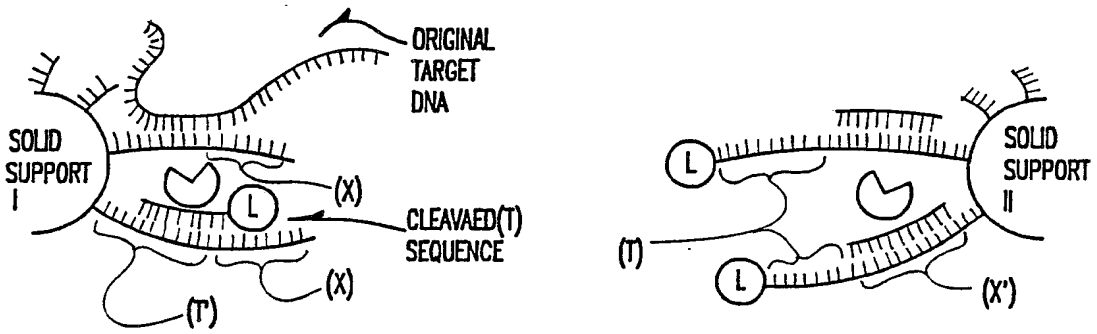
FIG.4A

IDT Exhibit 1020.2435

(4) CLEAVAGE OF (T') RELEASES SEQUENCE (X) WHICH DIFFUSES TO SOLID SUPPORT II, HYBRIDIZES TO COMPLEMENTARY CLEAVABLE SEQUENCE (X') AND RELEASES A LABELED (T) SEQUENCE. ORIGINAL TARGET SEQUENCE (T) CONTINUES TO RELEASE MORE (X) SEQUENCES.



(5) LABELED (T) SEQUENCE DIFFUSES BACK TO SOLID SUPPORT I, HYBRIDIZES WITH CLEAVABLE (T'), AND RELEASES MORE (X) SEQUENCES.



(6) CHA PROCESS CONTINUES BETWEEN THE TWO SOLID SUPPORT SYSTEMS WITH THE EXPONENTIAL RELEASE FROM SOLID SUPPORT II OF LABELED PROBES FOR SUBSEQUENT DETECTION.

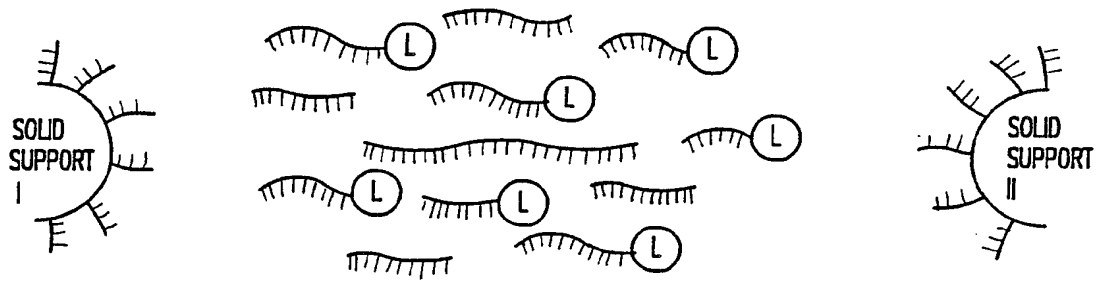


FIG.4B

SUBSTITUTE SHEET

DETECTION OF SPECIFIC RNA SEQUENCES USING A COMPETITIVE CHA ASSAY

ADD COMPLEMENTARY DNA MOLECULE TO TEST SAMPLE OF RNA, HYBRIDIZE THEN ADD LABELED RNA PROBE AND RNASE H:

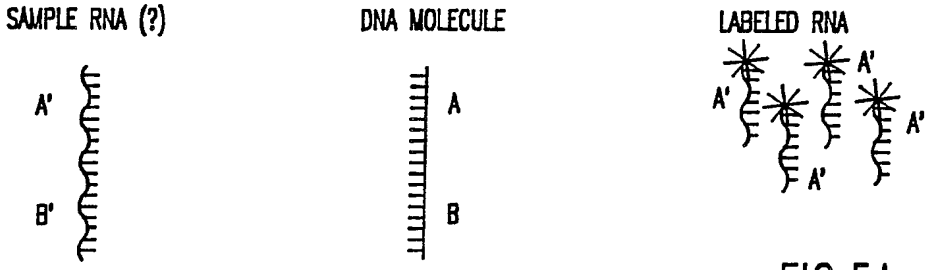
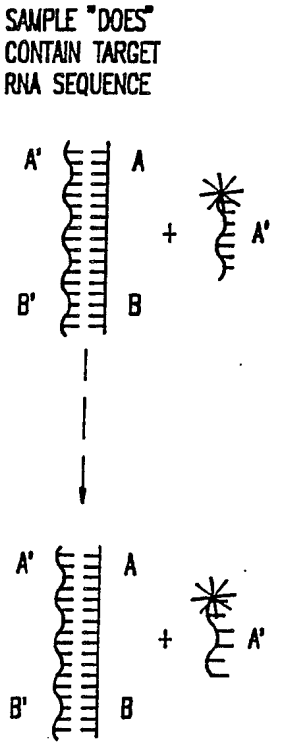
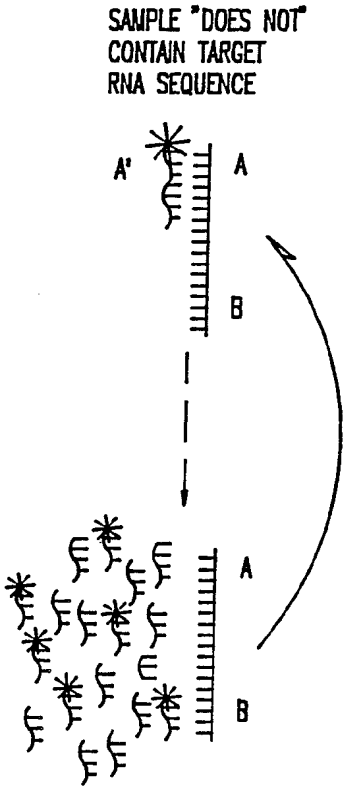


FIG.5A



NO LABELED FRAGMENTS RELEASED

FIG.5B



LABELED FRAGMENTS RELEASED

FIG.5C

IDT Exhibit 1020.2437

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US89/01025

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) [6]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC(4): C12Q 1/68
U.S. CL.: 435/6

II. FIELDS SEARCHED

Minimum Documentation Searched [7]

Classification System	Classification Symbols
U.S.	435/6; 536/27; 935/78, 86, 2 436/501,504, 172, 171

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched [8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT [9]

Category *	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No. [13]
Y	US, A, 4,683,195, Mullis et al, 28 July 1987.	1-70
Y	US, A, 4,683,202, Mullis, 28 July 1987.	1-70
Y,P	US, A, 4,800,159, Mullis et al., 24 January 1989.	1-70

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
30 May 1989	26 JUN 1989
International Searching Authority	**Signature of Authorized Officer**
ISA/US	AMELIA BURGRESS YARBROUGH

Form PCT/ISA/210 (second sheet) (Rev.11-87)

Espacenet

Bibliographic data: WO2010085715 (A1) — 2010-07-29

METHOD FOR ENRICHMENT OF SELECTED RNA MOLECULES

Inventor(s): RASHTCHIAN AYOUB [US] ± (RASHTCHIAN, AYOUB)

Applicant(s): QUANTA BIOSCIENCES [US]; RASHTCHIAN AYOUB [US] ± (QUANTA BIOSCIENCES, ; RASHTCHIAN, AYOUB)

Classification: - international: ***C07H21/04; C12P19/34; C12Q1/68***
- cooperative: **C12Q1/6806**; **C12Q1/6848**; C12Q2521/107; C12Q2521/327; C12Q2525/101; C12Q2525/207 more

Application number: WO2010US21897 20100122 Global Dossier

Priority number (s): US20090146647P 20090122

Also published as: CA2750547 (A1) EP2389454 (A1) EP2389454 (A4) EP2389454 (B1) US2010297709 (A1) more

Abstract of WO2010085715 (A1)

Improved methods of studying RNA molecules are provided. In particular, methods of treating mixtures of RNA molecules so as to enrich the mixture for a desired type of RNA molecule are provided. For example, the methods permit depletion of mRNA from complex mixtures to facilitate study of microRNAs in the mixture.

IDT Exhibit 1020.2439

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 July 2010 (29.07.2010)

PCT

(10) International Publication Number
WO 2010/085715 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01) *C07H 21/04* (2006.01)
C12P 19/34 (2006.01)

(21) International Application Number:
PCT/US2010/021897

(22) International Filing Date:
22 January 2010 (22.01.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/146,647 22 January 2009 (22.01.2009) US

(71) Applicant *(for all designated States except US)*: **QUANTA BIOSCIENCES** [US/US]; 202 Perry Parkway, Suite 1, Gaithersburg, MD 20877 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **RASHTCHIAN, Ayoub** [US/US]; 202 Perry Parkway, Suite 1, Gaithersburg, MD 20877 (US).

(74) Agent: **BOOTH, Paul, M.**; Perkins Coie LLP, 607 Fourteenth Street NW, Suite 800, Washington, DC 20005 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2010/085715 A1

(54) **Title**: METHOD FOR ENRICHMENT OF SELECTED RNA MOLECULES

(57) **Abstract**: Improved methods of studying RNA molecules are provided. In particular, methods of treating mixtures of RNA molecules so as to enrich the mixture for a desired type of RNA molecule are provided. For example, the methods permit depletion of mRNA from complex mixtures to facilitate study of microRNAs in the mixture.

IDT Exhibit 1020.2440

Methods for enrichment of selected RNA molecules

FIELD

The methods described below are in the fields of molecular and cellular biology. The methods permit enrichment of RNA molecules from, for example, complex mixtures of RNA molecules.

BACKGROUND

Within a given cell, tissue or organism, there exist many mRNA species, each encoding a separate and specific protein. This fact provides a powerful tool to investigators interested in studying genetic expression in a tissue or cell. mRNA molecules may be isolated and further manipulated by various techniques, thereby allowing the elucidation of the full functional genetic content of a cell, tissue or organism. The identity and levels of specific mRNAs present in a particular sample provides clues to the biology of the particular tissue or sample being studied. Therefore, the detection, analysis, transcription, and amplification of RNAs are important procedures in modern molecular biology.

In addition to mRNA, cells contain a variety of noncoding RNAs, including components of the machinery of gene expression, such as tRNAs and rRNAs and regulatory RNAs that influence the expression of other genes. Noncoding RNAs are diverse and a significant fraction of the genes of all organisms do not encode proteins. One class of small noncoding RNAs – the microRNAs or miRNAs – has recently been recognized to be quite numerous and phylogenetically extensive. MicroRNA genes produce tiny transcripts of about 22 nucleotides in length and function as antisense regulators of other RNAs.

A common approach to the study of abundance level or presence or absence of a particular RNA is the production of complementary DNA (cDNA). In this technique, the

IDT Exhibit 1020.2441

RNA molecules from an organism are isolated from an extract of the cells or tissues of the organism. From these purified RNA molecules, cDNA copies may be made using the enzyme reverse transcriptase (RT) or DNA polymerases having RT activity, which results in the production of single-stranded cDNA molecules. The term "reverse transcriptase" describes a class of polymerases characterized as RNA dependent DNA polymerases. All known reverse transcriptases require a primer to initiate synthesis of a first strand cDNA transcript from an RNA template.

One commonly used reverse transcriptase (RT) is Avian myoblastosis virus (AMV) reverse transcriptase (Verma, *Biochem. Biophys. Acta* 473:1(1977); Berger *et al.*, *Biochemistry* 22:2365 2372 (1983)). This enzyme has 5'-3' RNA directed DNA polymerase activity, 5'-3' DNA directed DNA polymerase activity, and RNase H activity. RNase H is a processive 5' and 3' ribonuclease specific for the RNA strand for RNA DNA hybrids (Perbal, A Practical Guide to Molecular Cloning, New York: Wiley & Sons (1984)). Errors in transcription cannot be corrected by reverse transcriptase because known viral reverse transcriptases lack the 3'-5' exonuclease activity necessary for proofreading (Saunders and Saunders, Microbial Genetics Applied to Biotechnology, London: Croom Helm (1987)). Another common reverse transcriptase is from Moloney murine leukemia virus (M-MLV). *See, e.g.*, Gerard, *DNA* 5:271 279 (1986) and Kotewicz *et al.*, *Gene* 35:249 258 (1985). M-MLV reverse transcriptase substantially lacking in RNase H activity has been described. *See, e.g.*, U.S. Patent No. 5,244,797.

A common technique uses first-strand cDNA as a template for amplification by the polymerase chain reaction, PCR. This method, often referred to as RNA PCR or reverse transcriptase PCR (RT-PCR), exploits the high sensitivity and specificity of the PCR process and is widely used for detection and quantification of RNA. Recently, the ability to measure the kinetics of a PCR reaction by on-line detection in combination with these RT-PCR

IDT Exhibit 1020.2442

techniques has enabled accurate and precise measurement of RNA sequences with high sensitivity. This has become possible by detecting the RT-PCR product through fluorescence monitoring and measurement of PCR product during the amplification process by fluorescent dual-labeled hybridization probe technologies, such as the "TaqMan" 5' fluorogenic nuclease assay described by Holland *et al.* (*Proc. Natl. Acad. Sci. U.S.A.* 88, 7276 (1991)), Gibson *et al.* (*Genome Res.* 6, 99 (1996)), and Heid *et al.* (*Genome Res.* 6, 986 (1996)); or "Molecular Beacons" (Tyagi *et al.*, *Nature Biotechnology* 14, 303 (1996)). Nazarenko *et al.* (*Nucleic. Acids Res.* 25, 2516 (1997)) have described use of dual-labeled hairpin primers, as well as recent modifications utilizing primers labeled with only a single fluorophore (Nazarenko *et al*,. 30, No. 9 **e37** (2002)). Another method is the addition of double-strand DNA-specific fluorescent dyes to the reaction such as: ethidium bromide (Higuchi *et al.*, *Biotechnology* (1992) and Higuchi *et al.*, *Biotechnology* 11, 102610, 413 (1993)), YO-PRO-1 (Ishiguro *et al.*, *Anal. Biochem.* 229, 207 (1995)), or SYBR Green I (Wittwer *et al.*, *Biotechniques* 22,130 (1997)). The concept of combining amplification with product analysis has become known as "real time" PCR, also referred to as quantitative PCR, or qPCR.

In qPCR using a double-strand specific fluorescent dye the fluorescent signal generated at each cycle of PCR is proportional to the amount of PCR product. A plot of fluorescence versus cycle number is used to describe the kinetics of amplification and a fluorescence threshold level is used to define a fractional cycle number related to initial template concentration. Specifically, the log of the initial template concentration is inversely proportional to the fractional cycle number (threshold cycle, or Ct), defined as the intersection of the fluorescence versus cycle number curve with the fluorescence threshold. Higher amounts of starting template results in PCR detection at a lower Ct value, whereas lower amounts require a greater number of PCR cycles to achieve an equivalent fluorescent threshold (Ct) and are detected at higher Ct values. Typically, the setting of this fluorescence

IDT Exhibit 1020.2443

threshold is defined as a level that represents a statistically significant increase over background fluorescent noise. Since this occurs at an early stage in the PCR process when critical substrates are not limiting, quantification of starting template occurs over a broad dynamic range with high accuracy, precision, and sensitivity. However, real-time PCR quantification of mRNA is still bounded by limitations of the process of reverse transcription.

The RT-PCR procedure, carried out as either an end-point or real-time assay, involves two separate molecular syntheses: (i) the synthesis of cDNA from an RNA template; and (ii) the replication of the newly synthesized cDNA through PCR amplification. In the so called "uncoupled" RT-PCR procedure (e.g., two step RT-PCR), reverse transcription is performed as an independent step using the optimal buffer condition for reverse transcriptase activity. Following cDNA synthesis, the reaction is diluted to decrease $MgCl_2$, and deoxyribonucleoside triphosphate (dNTP) concentrations to conditions optimal for Taq DNA Polymerase activity, and PCR is carried out according to standard conditions (see U.S. Patent Nos. 4,683,195 and 4,683,202). By contrast, "coupled" RT PCR methods use a common or compromise buffer for reverse transcriptase and Taq DNA Polymerase activities. In one version, the annealing of reverse primer is a separate step preceding the addition of enzymes, which are then added to the single reaction vessel. In another version, the reverse transcriptase activity is a component of the thermostable Tth DNA polymerase. Annealing and cDNA synthesis are performed in the presence of Mn^{++} then PCR is carried out in the presence of Mg^{++} after the removal of Mn^{++} by a chelating agent. Finally, the "continuous" method (e.g., one step RT-PCR) integrates the three RT-PCR steps into a single continuous reaction that avoids the opening of the reaction tube for component or enzyme addition. Continuous RT-PCR has been described as a single enzyme system using the reverse transcriptase activity of thermostable Taq DNA Polymerase and Tth polymerase and as a two

IDT Exhibit 1020.2444

enzyme system using AMV RT and Taq DNA Polymerase wherein the initial 65°C RNA denaturation step was omitted.

SUMMARY

Improved methods for synthesizing cDNA and for detection and/or quantification of RNA are provided. More specifically, the methods increase both the efficiency of cDNA synthesis and the sensitivity and accuracy of quantification of gene expression. The methods are useful in gene discovery, genomic research, diagnostics and identification of differentially expressed genes and identification of genes of importance to disease.

In one aspect, there is provided a method for selective enrichment of one or more RNA sequences in a complex sample comprising (a).reverse transcribing at least one RNA molecule to produce an RNA:cDNA hybrid molecule, using at least one reverse transcriptase and an exo-sample nucleotide that is incorporated into the cDNA; (b) incubating the resulting cDNA:RNA hybrid with Rnase H and an enzyme that specifically digests the cDNA sequence containing exo-sample nucleotide residues. The exo-sample nucleotide may be a substrate for an enzyme, for example, the exo-sample nucleotide used may be dUTP and the enzyme used for digestion of the cDNA is uracil-N-glycosylase. Other exo-sample nucleotide residues may be, for example, bromodeoxyuridine, 7-methylguanine, 5,6-dihyro-5,6 dihydroxydeoxythymidine, or 3-methyldeoxadenosine.

In step (a) above, the primer used for reverse transcription may contain one or more exo-sample nucleotide residues, which may be the same or different. The reverse transcriptase may have Rnase H activity, and in this case no additional Rnase H need be added in step (b), although the skilled artisan will recognize that additional RNAse H optionally may be added. When the RT is Rnase H deficient or has reduced Rnase H activity such that the RNA strand of the hybrid is not degraded, then an RNAse H is added in step (b).

In some embodiments, more than one reverse transcriptase may be used in step (a).

IDT Exhibit 1020.2445

In particular embodiments, the primer used for cDNA synthesis is oligo dU or oligo dT and the complex sample contains mRNA. The primer may also be a sequence-specific primer. The primermay be, for example, 6-12, 12-18, 18-25, 25-30, 30-40, or 40-100 bases long. An oligo dU or oligo dT may contain about 6-30 residues, for example. The oligo (dT) may be an anchored oligo(dT). In specific embodiments the 3' nucleotide of the anchored oligo(dT) may be dA, dC or dG, or the the 3' end of the anchored oligo(dT) may be dAdA, dAdG, dAdC, dGdG, dGdA, dGdC, dCdC, dCdA, or dCdG. The oligo dT may contain at least one one exo sample nucleotide, for example a dU residue.

The RT may be a viral reverse transcriptase. Suitable RTs include, but are not limited to: AMV RT, RSV RT, MMLV RT, HlV RT, EIAV RT, RAV2 RT, TTH DNA polymerase, C.hydrogenoformans DNA polymerase, Sensiscript™, Omniscript™, SUPERSCRIPT I® RT, SUPERSCRIPT II® RT, SUPERSCRIPT III® RT THERMOSCRIPT® RT, and MMLV and Rnase H deficient mutants of these enzymes.

In any of the methods described above, any free exo-sample nucleotide remaining after step (a) may be degraded with a reagent that is suitable for selective degradation of the exo-sample nucleotide, such as an enzyme. For example, when the exo-sample nucleotide is dUTP, the dUTP remaining after step (a) may be degraded with dUTPase.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows a qPCR amplification profile of various reactions.

Figure 2 shows the melting profile of the reactions in Figure 1. A single amplification product (melting at 73 C, shown by brace) was observed from a sample containing a mimic RNA, while HeLa RNA without any mimic RNA (baseline) produced no amplification products.

Figure 3 shows a qPCR amplification profile.

IDT Exhibit 1020.2446

Figure 4 shows the melting profiles of the reactions in Figure 3. Three amplification products were observed from samples containing mimic RNA: specific product melting at 73 C; and 2 other non specific products melting at 77 C and 84 C (square brackets0. The HeLa RNA without any mimic RNA (curves shown by brace) produced the 2 non-specific amplification products melting at 77 C and 84 C.

DETAILED DESCRIPTION

When RNA species are being analyzed , complexity of the sample and the total number of different RNA species present in the sample, can have profound effects on the sensitivity and accuracy of detection and/ or quantification. This problem is especially true for RNA analytes that are present in low concentration and where the sample contains a significant amount of background RNA. This problem can be further exacerbated if there is any sequence homology between the analyte RNA and other RNAs present in the sample. This is a constant problem in dealing with detection and quantification of rarely expressed genes, regulatory RNAs such as microRNAs and detection of viral RNAs in complex biological specimens.

Although there are many RNA fractionation and purification methods and commercial products available for RNA purification, they generally require large amounts of starting material and are based on the physical characteristics of different RNAs rather than their sequence or type. For example, size fractionation has been used to attempt to purify microRNAs

Methods for selective degradation of specific RNA species in a complex sample are provided. The methods are very flexible and can be modulated to deplete or effect removal of a few species of RNA based on their sequence or of an entire class of RNAs depending on the application of the method and the desired outcome.

IDT Exhibit 1020.2447

The methods enzymatically form RNA:DNA hybrids from RNA molecules present in a mixture. One or more modified nucleotides are used in this first step that make the cDNA recognizable or degradable in a specific manner. RNAse H activity, which degrades RNA only in an RNA;DNA hybrid is used to degrade the RNA portion of the hybrid molecules. The methods therefore selectively remove unwanted RNA molecules from complex mixtures/samples and improves detection or accuracy of quantification for desired RNA analytes.

In one aspect, a complex RNA sample containing both poly A-tailed species and non poly-A molecules, can be treated using the methods described herein so that the poly-A containing RNA can be selectively digested and therefore removed as a potentially confounding species for further analyses. The resulting sample is therefore enriched for non-poly-A RNAs. In this case, reverse transcription uses a primer that hybridizes to the poly A tail of the RNA. Such cDNA synthesis using an "exo sample" nucleotide generates a cDNA that is recognizable based on the use of a particular "exo sample" nucleotide. In the meantime non-poly A RNA of the sample remains as single stranded RNA and is not converted into cDNA. cDNA synthesis using poly A RNA produces a RNA: DNA hybrid. These hybrids are substrates for RNAse H, which digests the RNA strand of these molecules. Moreover, since the cDNA strand was made using exo sample nucleotides it can also be digested specifically using an appropriate enzyme, making it impossible for the resulting product to be amplified.

Exo-sample nucleotides

Many different exo-sample nucleotides that can be incorporate into cDNA and that permit subsequence specific degradation reactions can be used and are available commercially and/or have been described in the literature. Alternatively, depending on the nature of the exo sample nucleotide used, the cDNA can be removed physically using an

IDT Exhibit 1020.2448

appropriate binding partner. For example, a biotin-containing nucleotide can be used and the cDNA removed using streptavidin. More generally, exo-sample nucleotides containing different antigenic moieties have been linked to nucleoside triphosphates and can be incorporated into DNA by DNA polymerases including reverse transcriptases. The cDNA molecule may then be selectively removed using an antibody or other binding molecule that recognizes the antigenic moiety. Examples of such nucleotides are Fluorescent moieties such as fluorescene, Texas Red, Cy3, Cy5 and many others that are commercially available from variety of sources (Invitrogen/Life Technologies, Carlsbad, CA; Amersham/GE Life Sciences; Sigma, St. Louis, MO, and the like). Other moieties that have been used include antigenic moieties such as digoxigenin, biotin, and others (see Sigma and Roche Applied Sciences for commercial sources). It will be apparent to those skilled in the art that other non-natural or modified nucleotide can be used with the methods described herein.

"Exo-sample nucleotide", as used herein, refers to a nucleotide which is generally not found in a sequence of DNA. See US Patent No. 5,137,814. For most DNA samples, deoxyuridine is an example of an exo-sample nucleotide. Although the triphosphate form of deoxyuridine, dUTP, is present in living organisms as a metabolic intermediate, it is rarely incorporated into DNA. When dUTP is incorporated into DNA, the resulting deoxyuridine is promptly removed in vivo by normal processes, e.g. processes involving the enzyme uracil DNA glycosylase (UDG) (U.S. Pat. No. 4,873,192; Duncan, The Enzymes XIV: 565-586 (1981), both references herein incorporated by reference in their entirety). Thus, deoxyuridine occurs rarely or never in natural DNA. It is recognized that some organisms may naturally incorporate deoxyuridine into DNA. For nucleic acid samples of those organisms, deoxyuridine would not be considered an exo-sample nucleotide. Examples of other exo-sample nucleotides include bromodeoxyuridine, 7-methylguanine, 5,6-dihyro-5,6 dihydroxydeoxythymidine, 3-methyldeoxadenosine, etc. (see, Duncan, *supra*).

IDT Exhibit 1020.2449

Other exo-sample nucleotides will be evident to those in the art. The presence of deoxyuridine, or any other exo-sample nucleotide, may be determined readily using methods well known to the art. Other exo-sample nucleotides may be envisioned. Numerous DNA glycosylases are known to the art. An exo-sample nucleotide which may be incorporated into DNA during DNA or cDNA synthesis and a DNA glycosylase that acts on it may be used in the methods described herein. For example, bromodeoxyuridine (BdUR) is well known in the art to be incorporated into DNA. DNA containing BdUR may be degraded on exposure to light under appropriate conditions.

The deoxyribonucleoside triphosphate dUTP exemplifies an exo-sample nucleotide which may be conveniently incorporated into an enzymatic cDNA synthesis and /or DNA amplification procedures, thereby resulting in deoxyuridine-containing DNA. The DNA products of such a reaction will normally contain many uracil bases. Treatment of DNA containing uracil bases with uracil DNA glycosylase (UDG) results in cleavage of the glycosidic bond between the deoxyribose of the DNA sugar-phosphate backbone and the uracil base. The loss of the uracil creates an apyrimidinic site in the DNA, which blocks DNA polymerase from using the DNA strand as a template for the synthesis of a complementary DNA strand. The presence of substantial numbers of apyrimidinic sites in each DNA target molecule interferes with amplification procedures which use DNA polymerase to synthesize copies of target DNA. It is also known that apyrimidinic sites are susceptible to breakage upon heating or exposure to basic pH.

The removal of uracil or other exo-sample nucleotides from DNA by action of DNA glycosylases results in DNA that contain abasic sites. This term has been used in the literature as a generic way of describing such nucleic acids molecules. Other methods may be used for degrading cDNA molecules or otherwise making them unsuitable for further enzymatic reactions such as PCR amplification or other methods of amplification and

IDT Exhibit 1020.2450

detection or quantification. It is known in the art that there are specific enzymes that cleave abasic sites and can be used as part of the methods described herein. These enzymes are part of the repair mechanism of practically all organisms and many such enzymes have been isolated and characterized. These enzymes are refered to as AP endonucleases(Apurinic/apyrimidinic endonuclease). There are four types of AP endonucleases that have been classified according to their sites of incision. Class I and class II AP endonucleases incise DNA at the phosphate groups 3′ and 5′ to the baseless site leaving 3′-OH and 5′-phosphate termini. Class III and class IV AP endonucleases also cleave DNA at the phosphate groups 3′ and 5′to the baseless site, but they generate a 3′-phosphate and a 5′-OH. Human AP Endonuclease (APE1), like most AP endoucleases, is of class II and requires an Mg^{2+} in its active site in order to carry out its role in base excision repair.

"Uracil DNA glycosylase" (UDG), a term of art, refers to an enzyme which cleaves the glycosidic bond between the base uracil and the sugar deoxyribose, only when the monomeric nucleotide dUTP is incorporated into a DNA molecule, resulting in incorporation of a deoxyuridine moiety (Duncan (1981) in The Enzymes 14:565, ed.: Boyer P). The enzyme does not act upon free dUTP, free deoxyuridine, or RNA (Duncan, supra).

The use of exo-sample nucleotides to remove potential contaminants from PCR amplified samples has been described by Longo *et al* (*Gene* 93:125-128 (1990), U.S. Pat. No. 5,035,966), herein incorporated by reference in their entireties. Longo describes the use of either dU-containing oligonucleotides or dUTP in the PCR-directed amplification of a target sequence.

Primers

For removal of mRNA that contains a polyA tail, oligo dT or oligo dU primers can be used for the cDNA synthesis step. For removal of other RNA molecules, sequence-specific primers can be used. The length of oligo dT or oligo dU or other primers to be used for

IDT Exhibit 1020.2451

cDNA synthesis can vary from 6 bases to 100 bases or even longer. Most primers generally used are in range of 10 to 40 bases long, although the skilled worker will recognize that any length may be used that permits the desired priming to occur. Other types of primers with different composition can be used in place of oligo dT. Examples of such compositions include, but are not limited to, oligo dT where the 3' base is A, or C, or G. Alternatively, two bases at the 3' end can be variable and can be any combination of A, C, or G. Other sequences or moieties that can base pair with poly A sequences of mRNA can also be used. An example, without limitation, is deoxyuridine, (dU). Other exo-sample nucleotides can be used in the primer sequence. The primer(s) used for cDNA synthesis can also contain ribonucleotide residues and be RNA primers or can contain both ribonucleotide and deoxyribonucleotide residues.

The primer used for cDNA synthesis can be any sequence depending on the purpose of the assays and can contain one or more exo-sample nucleotides.

The methods described above can be used for depletion or removal of mRNA species from a complex sample and therefore to provide effective enrichment of non-mRNA species in the sample. The non-mRNA species can be any RNA that does not contain a poly A tail. In a specific example, total RNA from a sample is mixed with a cDNA primer that contains dU residues, instead of dT, *i.e.* oligo dU. The oligo dU primer anneals to the polyA tail of the mRNA species present in a sample and a suitable reverse transcriptase is used for cDNA synthesis in the presence of dNTPs containing dUTP. The cDNA produced contains dU residues substituted for dT, and therefore is a substrate for the enzyme uracil-N-glycosylase or uracil DNA glycosylase. The resulting cDNA:mRNA hybrid is incubated with Rnase H to digest/remove the mRNA strands that are present in the RNA:DNA hybrids. The cDNA and the oligo dU primer in the reaction can be digested with UDG. These enzymatic degradation reactions can be sequential in any order, or simultaneous. This degradation process will

IDT Exhibit 1020.2452

specifically digest and deplete or remove mRNA species and the corresponding cDNA, while leaving intact the other RNA species that did anneal to the primer and make cDNA.

The resulting RNA in the sample is therefore enriched and less complex than the starting total RNA in the sample and can be used for a variety of analysis techniques, such as PCR amplification or hybridization. Many different amplification methods have been described in the past few years including, but not limited to, the ligase chain reaction, NASBA, LAT (Ligation Activated Transcription amplification); TAS (Transcription Activated Amplification) and numerous other methods. The method of enriching RNA molecules of interest prior to amplification or analysis can be used as part of any of these or other methods. Some of the above-mentioned amplification methods are isothermal while others require thermocycling, but both types may be used in conjunction with the methods described herein.

Reverse transcriptases

Reverse transcriptases that may be used in the methods described herein nclude, but are not limited to, retroviral reverse transcriptase, retrotransposon reverse transcriptase, hepatitis B reverse transcriptase, cauliflower mosaic virus reverse transcriptase, bacterial reverse transcriptase, Tth DNA polymerase, Taq DNA polymerase, Tne DNA polymerase (WO 96/10640 and WO 97/09451), Tma DNA polymerase (U.S. Pat. No. 5,374,553) and mutants, variants or derivatives thereof (see e.g., WO 97/09451 and WO 98/47912). Some RTs have reduced, substantially reduced or eliminated RNase H activity. An enzyme "substantially reduced in RNase H activity" as that term is used herein has less than about 20%, more preferably less than about 15%, 10% or 5%, and most preferably less than about 2%, of the RNase H activity of the corresponding wild type or RNase H+ enzyme such as wild type Moloney Murine Leukemia Virus (M-MLV), Avian Myeloblastosis Virus (AMV) or Rous Sarcoma Virus (RSV) reverse transcriptases. The RNase H activity of any enzyme

IDT Exhibit 1020.2453

may be determined by a variety of assays, such as those described, for example, in U.S. Pat. No. 5,244,797, in Kotewicz *et al. Nucl. Acids Res.* 16:265 (1988) and in Gerard *et al.* (1992) *FOCUS* 14:91. Other RTs that may be used include, but are not limited to, M-MLV H$^-$ reverse transcriptase, RSV H$^-$ reverse transcriptase, AMV H$^-$ reverse transcriptase, RAV (rous-associated virus) H$^-$reverse transcriptase, MAV (myeloblastosis-associated virus) H$^-$ reverse transcriptase and HIV H$^-$ reverse transcriptase (see U.S. Pat. No. 5,244,797 and WO 98/47912).

Many types and mutants of reverse transcriptases are available commercially or have been described in the literature. Commercially available RTs include AMV RT, RSV RT, MMLV RT, HIV RT, EIAV RT, RAV2 RT, THERMOSCRIPT® RT, ASLV, Foamy Virus family of reverse transcriptases and Rnase H$^-$ mutants thereof, SUPERSCRIPT III® RT ,SUPERSCRIPT II® RT, and SUPERSCRIPT® RT. Other examples include Sensiscript™ and OmniScript™ from Qiagen Corp. (Germantown, MD)

Enzymes with lowered RNaseH activity or virtually no RNaseH activity are suitable for use in the methods described herein. Therefore, depending on the reverse transcriptase being used, additional RNAse H addition may or may not be necessary and all variations are within the scope of these methods. Rnase H is also a ubiquitous enzyme in many organisms and bacteria and is available from many sources. The most commonly used is RNASE H from E. coli, however, there are many sources for this type of enzyme (Schultz *et al. Virus Research*, 134(1-2): 86-103(2008). Cerritelli *et al. FEBS J.*276(6):1494-505 (2009); Tadokoro *et al., FEBS J.* 276(6):1482-93 (2009). Other enzymes and proteins with similar activity that may not be referred to as RnaseH also may be used.

Other suitable RTs include DNA polymerases and thermostable DNA polymerases having natural reverse transcriptase activity or that demonstrate RT activity under certain conditions. Examples are Tth DNA polymerase, DNA Polymerase I or Klenow fragment of

IDT Exhibit 1020.2454

DNA Polymerase I. All such enzymes are included without limitation in the scope of the present invention. Use of thermostable enzymes may present certain advantages in some situations, such as being able to control the stringency of annealing of primer or primers to RNA in the reverse transcription step. Under higher temperature the annealing and reaction may be more specific and advantageous for certain applications. Reaction conditions and specificity can be modified, as known in the art, by physical (temperature and/or time) or chemical means such as addition of compounds that affect the specificity of primer annealing and priming/extension. Many such compounds or additives are know in the art, glycerol, DMSO, Betain, Proline, being a small example. PCT/US99/05538 describes many compounds that affect the annealing and secondary structure of nucleic acids and may be used as part of the methods described herein.

Single Strand binding proteins (SSB) are known from a variety of organisms, including thermo-tolerant, thermophilic or psycrophilic organism, and can be used to improve specificity of primer annealing and extension during enzymatic reactions (see United States Patent 5,605,824). US Application 20070059713 describes fusion proteins between polymerases and SSBs that can be used in the present methods. All such protein mixtures or fusions, including fusions to reverse transcriptases can be used in the present methods.

Depending on the specific requirements of the downstream processes for analysis of the RNA other modification or additions can be made to the methods of the invention. For example, one improvement would be to additionally treat the mixture with the enzyme dUTPase to remove the dUTPs in the mixture. This may be useful in subsequent cDNA synthesis and amplification of the analyte RNAs in the sample. Moreover, the resulting RNAs can be further purified using physical separation method to remove the low molecular weight digested DNA and RNA, and this may provide further advantages for some applications. If a different exo-sample nucleotide is used and there is a desire to remove or

IDT Exhibit 1020.2455

eliminate the extra nucleotides an appropriate enzyme or method would be necessary. Suitable exo-sample nucleotides and corresponding enzymes are known in the art and will be evident to those skilled in the art.

Uses of the methods

The methods described herein are broadly applicable to a variety of RNA species and there is no limitation on the kind or sequence of the process used for cDNA synthesis. For example, one skilled on the art will appreciate that the methods described herein can be applied for enrichment of RNA sequences for detection of viral RNAs. Viral RNAs are generally not polyadenylated (i.e. retroviral sequences, HIV, SIV, etc). Detection of HIV is often carried out in very complex samples containing large quantity of cellular RNA. The presence of such large quantities of unrelated background sample RNA interferes with the sensitivity and accuracy of detection. The methods of invention described herein will improve detection of rare viral sequences.

As described above the sequence of the primer for cDNA synthesis can be any sequence. For example, the primer can be complementary to rRNA, and tRNA that represents the majority of cellular RNA species. Use of this method can provide an easy method for enrichment of mRNA species in a total RNA sample. It will become evident to those skilled on the art that a combination of primers containing dU residues can be used to deplete or digest a number of species of condensed RNA from complex samples.

The methods described herein also can be used to digest RNA species that are abundant. For example a number of housekeeping genes such as Actin and GAPDH are present in high concentration in the cellular RNA, and can be selectively removed using primers specific to the sequence in the methods of the invention. These primers can be combined with primers for rRNA and tRNA to enrich the desired RNA species of the sample.

IDT Exhibit 1020.2456

It will become evident to those skilled on the art that a wide range of variations and applications of the methods described herein are possible.

Although the foregoing refers to particular preferred embodiments, it will be understood that the methods described herein are not so limited. It will occur to those of ordinary skill in the art that various modifications may be made to the disclosed embodiments and that such modifications are intended to be within the scope of the present methods.

Example 1.

This example demonstrates the difficulties of attempting to detect and quantify a microRNA in a complex sample of total cellular RNA. HeLa cell total RNA was used as a source of complex RNA containing many different RNA species, however, any RNA can be used. Since HeLa cells and other biologically derived RNAs contain various micro RNAs (miRNAs, or MIRS) and the exact quantity of each MIR is not known, we synthesized a microRNA mimic that normally does not exist in biological samples. The pure RNA was quantified accurately in its pure form and was used in conjunction with HeLa RNA or in pure form. This system provided a way to distinguish qPCR signal generated from specific amplification of the mimic RNA versus the non-specific signal generated from the complex HeLa RNA. The mimic RNA had the sequence 5' GGAUGGCGGAGCGGUCU 3' (SEQ ID NO: 1)

Enrichment protocol: For enrichment of desired RNA species (in this case micro RNAs and other RNAs without a poly A tail) complex samples were subjected to the following enzymatic enrichment protocol. Each 10 μL reaction contained:

- 2 μL of 5X cDNA synthesis buffer (qScript Flex cDNA kit , cat # 95049 from Quanta BioSciences, Inc. Gaithersburg, MD 20877).

IDT Exhibit 1020.2457

- 0.4 µL dNTPs where TTP had been replaced with dUTP or a mixture of TTP and dUTP was used.
- 1 µL qScript RT (MMLV RT with normal RNAse H activity)
- 1 µL oligo dU 20 (2µM final concentration)
- 5 µL sample RNA

The reaction was mixed and incubated at 37° C for cDNA synthesis for 15 minutes. Following cDNA synthesis samples were incubated with 1 unit of UDG (E. coli UDG) at 42° C for 10 min. The control samples were incubated without addition of UDG. Following digestion of dU containing DNA with UDG the reactions were incubated at 95° C for 5 min to inactivate the enzymes.

Analysis of microRNAs by qPCR: The following procedure was used for tailing of non polyA RNAs (including microRNAs) with Poly A Polymerase prior to conversion into cDNA which is subsequently amplified by qPCR. Each 15 uL reaction contained:

- 3 µL of 5X cDNA synthesis buffer
- 0.4 µL ATP (10mM stock)
- 1-2 units of Poly A polymerase (Epicenter Biotechnologies or New England BioLabs)
- 11.5 µL RNA sample and water

The reactions were mixed and incubated at 37° C for 10 min followed by heating at 95° C for 5 Min. After Poly A tailing reagents for cDNA synthesis were added as follows:

- 1 µL of 5x cDNA synthesis buffer
- 1 µL of AD1 dT adapter primer (10 µM stock)
- 0.5 µL of qScript reverse transcriptase
- 2.5 uL of water to bring the volume to 20 uL

The sequence of AD1 dT adapter primer is:

IDT Exhibit 1020.2458

5' GGTTAACCCAGAATGGCGGTAAGGGTGTGGTAGGCGAGTCA$(T)_{20}$ 3' (SEQ ID NO:2)

The reaction was mixed and incubated at 42° C for 15 min for cDNA synthesis, followed by 95° C for 5 min. Generally 2 μL of this reaction was used for qPCR using Perfecta SYBR green qPCR SuperMix (Quanta BioSciences, Gaithersburg, MD). The primers for amplification were used according to the instructions of the SYBR green Supermix and had the sequences:

AD1: 5' GGTTAACCCAGAATGGCGGTA 3' (SEQ ID NO: 3)

Smir-1: 5' GGATGGCGGAGCGGTCT 3' (SEQ ID NO: 4)

Results: The above enrichment protocol was applied to HeLa RNA and was demonstrated to be useful in reduction or elimination of the non-specific products that arise from non specific amplification of sequences not related to microRNAs. We started with 100 ng of HeLa RNA in 2 duplicate samples. Two of the samples were treated with the enrichment protocol above prior to tailing with PolyA polymerase and cDNA synthesis, while the control samples were not treated with the enrichment protocol. When the resulting cDNAs were subjected to amplification with the Smir1 and AD1 primers, a significant amount of a non-specific amplification product with a melting temperature of 84° C was observed from the control HeLa cDNA. When the cDNA from the enriched samples was analyzed, as expected, no amplification products were observed. As a positive control we also used the synthetic mimic as template for tailing and cDNA synthesis and observed the specific amplification product melting at 73° C. When the mimic RNA was added to HeLa RNA and was tested without enrichment, it produced 2 products melting at 73° C and 84° C which represented the specific and non-specific products, respectively. When the mimic RNA was added to HeLa RNA and was enriched, it only produced the specific product melting at 72°C.

IDT Exhibit 1020.2459

Example 2:

Enrichment protocol: For enrichment of desired RNA species (in this case micro RNAs and other RNAs without a poly A tail) complex samples were subjected to the following enzymatic enrichment protocol. Each 10 µL reaction contained:

- 2 µL of 5X cDNA synthesis buffer(qScript Select cDNA synthesis kit , cat # .....from Quanta BioSciences, Inc. Gaithersburg, MD 20877).
- 0.4 µL dNTp's where TTP had been replaced with dUTP (25mM each dATP,dUTP,dGTP,dCTP).
- 1µL qScript RT (MMLV RT with normal RNAse H activity)
- 1µL oligo dT 20 (2uM final concentration)
- 5 µL sample RNA

The reaction was mixed and incubated at 37° C for cDNA synthesis for 15 minutes. Following cDNA synthesis samples were incubated with 0.5 units of UDG (E. coli UDG) at 42° C for 20 min. The control samples were incubated without addition of UDG. Following digestion of dU containing DNA with UDG the reactions were incubated at 95° C for 5 min to inactivate the enzymes.

Analysis of microRNAs by qPCR: The following procedure was used for tailing of non polyA RNAs (including microRNAs) with Poly A Polymerase and conversion into cDNA which is subsequently amplified by qPCR. A 10 µL reaction mix was prepared as shown below and was added to 10 L RNA sample from enrichment step above, followed by incubation at 37° C for 30 min. After tailing and cDNA synthesis, the reaction was heated to 95° C for 5 min and then was diluted with addition of 80 µL TE buffer.

IDT Exhibit 1020.2460

- 2 μL of 5X cDNA synthesis buffer
- 0.8 μL ATP (2.5 mM stock)
- 1 μL of AD1 dT adapter primer (10 μM stock)
- 1uL dNTP (10mM each dATP,dTTP,dGTP,dCTP)
- 1.8 units of Poly A polymerase (Epicenter Biotechnologies or New England BioLabs)
- 1.4 μl qScript RT (with Rnase inhibitor)
- Water to bring the volume to 10 μL

5 μL of the diluted cDNA mix was used for qPCR using Perfecta SYBR green qPCR SuperMix (Quanta BioSciences). The primers used for amplification were AD1 and the Smir-1 according to the instructions of the SYBR Green Supermix.

Results: In this experiment 5 duplicate samples were prepared all of which had 100 ng of HeLa RNA and varying amounts of mimic RNA ranging from 0 to 10^5, or 10^6, or 10^7 or 10^8 copies. One set of these samples were enriched as described in protocol for example 2 while the other set served as control without addition of UDG. The results are shown in Figures 1-4.

Figure 1 demonstrates the qPCR amplification profile and Figure 2 shows the melting profile of the reactions in Figure 1. As can be seen when the RNA was treated with UDG as in the enrichment protocol, a single amplification product (melting at 73 C, shown by brace) was observed from the sample containing mimic RNA, and the HeLa RNA without any mimic RNA (baseline) produced no amplification products.

Figure 3 demonstrates the qPCR amplification profiles and Figure 4 shows the melting profiles of the reactions in Figure 3. As can be seen when the control RNA which was not treated with UDG as in the enrichment protocol, 3 amplification products were observed from sample containing mimic RNA: specific product melting at 73 C; and 2 other

IDT Exhibit 1020.2461

non specific products melting at 77 C and 84 C (square brackets0. The HeLa RNA without any mimic RNA (curves shown by brace) produced the 2 non-specific amplification products melting at 77 C and 84 C.

IDT Exhibit 1020.2462

Claims:

What is claimed is:

1. A method for selective enrichment of one or more RNA sequences in a complex sample comprising;

a. reverse transcribing at least one RNA molecule to produce an RNA:cDNA hybrid molecule, using at least one reverse transcriptase and an exo-sample nucleotide that is incorporated into said cDNA.

b. incubating the resulting cDNA:RNA hybrid with Rnase H and an enzyme that specifically digests the cDNA sequence containing exo-sample nucleotide residues.

2. The method of claim 1, wherein said exo-sample nucleotide is a substrate for an enzyme.

3. The method of claim 1, wherein step (a) uses at least one primer comprising at least one exo-sample nucleotide.

4. The method of claim 1, where in the reverse transcriptase has Rnase H activity, and no additional Rnase H is added.

5. The method of claim 1, where in the RT is Rnase H deficient or has reduced Rnase H activity.

6. The method of claim 1 or 5, further comprising in step (b) adding Rnase H

7. The method according to any of claims 1 – 6, where more than one reverse transcriptase is used in step (a).

8. The method according to any of claims 1 – 7, wherein the exo-sample nucleotide used is dUTP and the enzyme used for digestion of cDNA is uracil-N-glycosylase.

9. The method according to any of claims 1-8 wherein the exo-sample nucleotide residues are chosen from the group consisting of: bromodeoxyuridine, 7-methylguanine, 5,6-dihyro-5,6 dihydroxydeoxythymidine, and 3-methyldeoxadenosine,

IDT Exhibit 1020.2463

10. The method according to any of claims 1 – 7, wherein the primer used for cDNA synthesis is oligo dU and said complex sample comprises mRNA.

11. The method according to any of claims 1 – 7, wherein the primer used for cDNA synthesis contains an exo-sample nucleotide.

12. The method according to any of claims 1-7, wherein more than 1 primer is used for cDNA synthesis.

13. The method according to claim 12, wherein the primers anneal to target abundant RNA species within the complex RNA mixture.

14. .The method according to any of claims 1-7, wherein said reverse transcription is carried out in the presence of oligo(dU) consisting essentially of between about 6 and about 30 residues.

15. The method according to any of claims 1-7, wherein the primer for said cDNA synthesis is 6-12, 12-18, 18-25, 25-30, 30-40, or 40-100 bases long.

16. The method according to any of claims 1-7, wherein said reverse transcription is carried out in the presence of oligo dT that is 6-12 or 12-18 or 18-25 or 25-30 or 30-40 or40-100 bases long.

17. The method according to any preceding claim, wherein said at least one reverse transcriptase is a viral reverse transcriptase.

18. The method according to any of claims 1-16, wherein said reverse transcriptase is selected from the group consisting of AMV RT, RSV RT, MMLV RT, HlV RT, EIAV RT, RAV2 RT, TTH DNA polymerase, C.hydrogenoformans DNA polymerase, Sensiscript™, Omniscript™, SUPERSCRIPT I® RT, SUPERSCRIPT II® RT, SUPERSCRIPT III® RT THERMOSCRIPT® RT, MMLV and Rnase H deficient mutants thereof.

19. The method according to any preceding claim, wherein said reverse transcription is carried out in the presence of an anchored oligo(dT).

IDT Exhibit 1020.2464

20. The method according to claim 19, wherein the 3' nucleotide of said anchored oligo(dT) is dA.

21. The method according to claim 19, wherein the 3' nucleotide of said anchored oligo(dT) is dC.

22. The method according to claim 19, wherein the 3' nucleotide of said anchored oligo(dT) is dG.

23. The method according to claim 19, wherein the 3' end of said anchored oligo(dT) is selected from the group consisting of dAdA, dAdG, dAdC, dGdG, dGdA, dGdC, dCdC, dCdA, and dCdG.

24. The method according to any of claims 1-23 , wherein said reverse transcription is carried out in the presence of an oligo(dT) which comprises at least one dU residue.

25. The method according to claim 24 wherein said oligo dT contains at least one exo sample nucleotide residue.

26. The method according to any preceding claim, further comprising degrading any free exo-sample nucleotide remaining after step (a) with an enzyme.

27. The method according to claim 26, wherein said exo-sample nucleotide is dUTP, and said dUTP is degraded with dUTPase.

IDT Exhibit 1020.2465

FIGURE 1

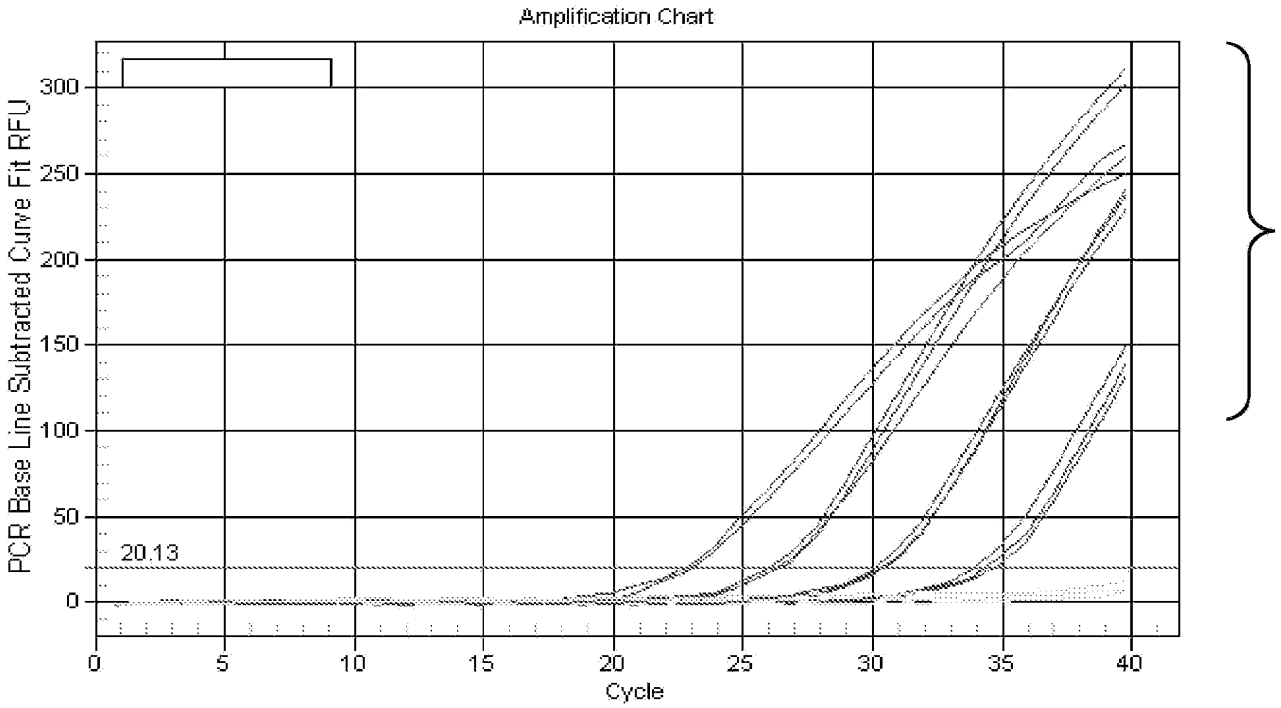
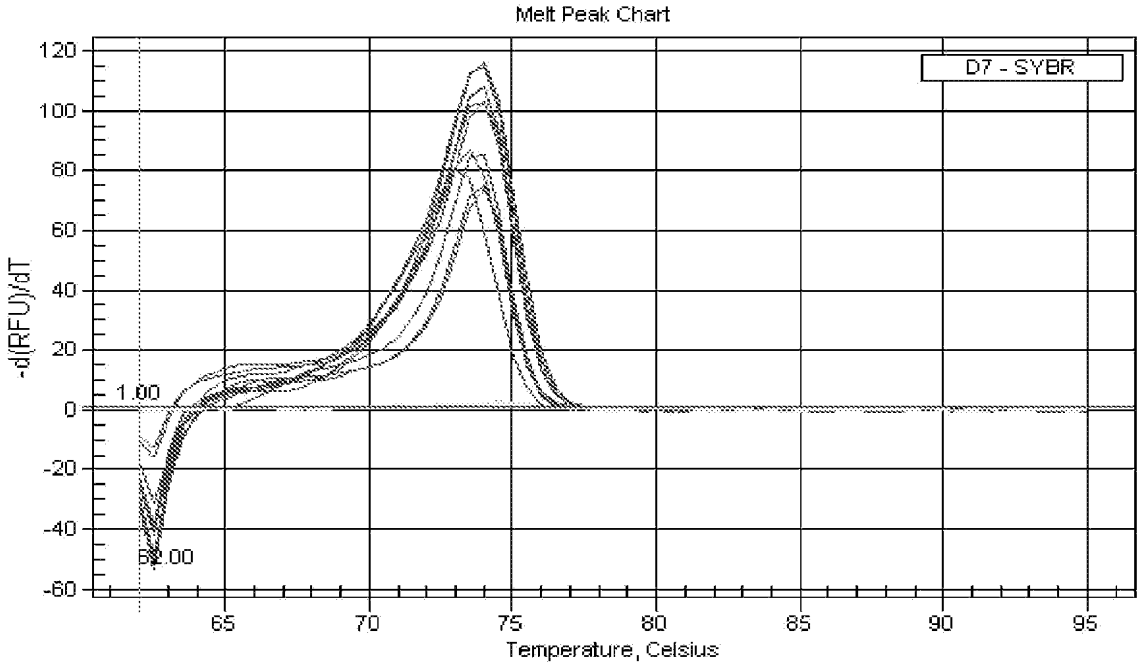


FIGURE 2



IDT Exhibit 1020.2466

FIGURE 3

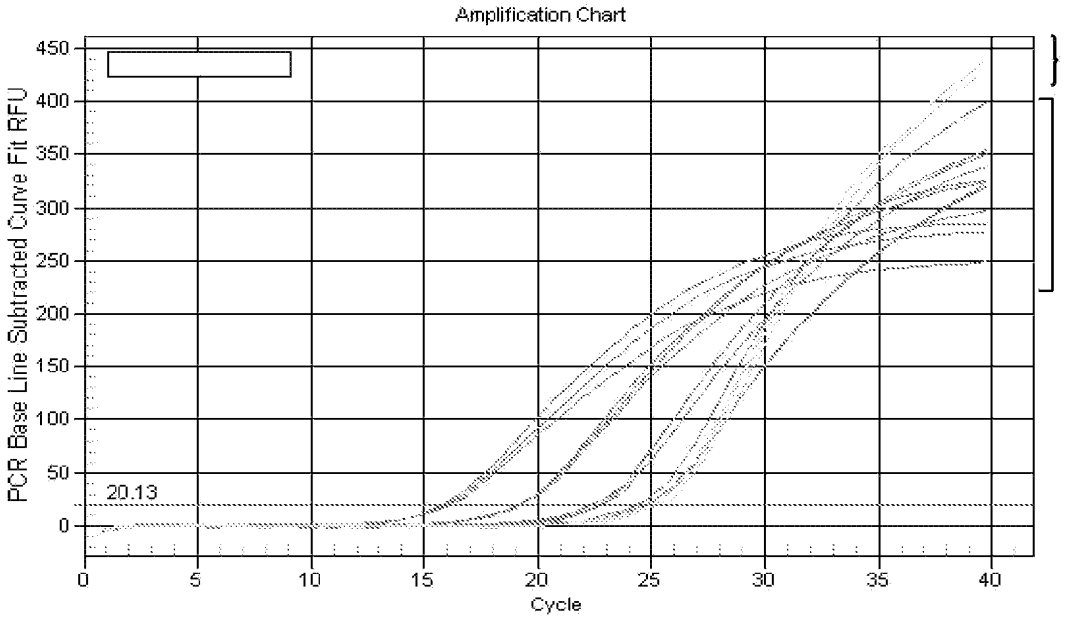
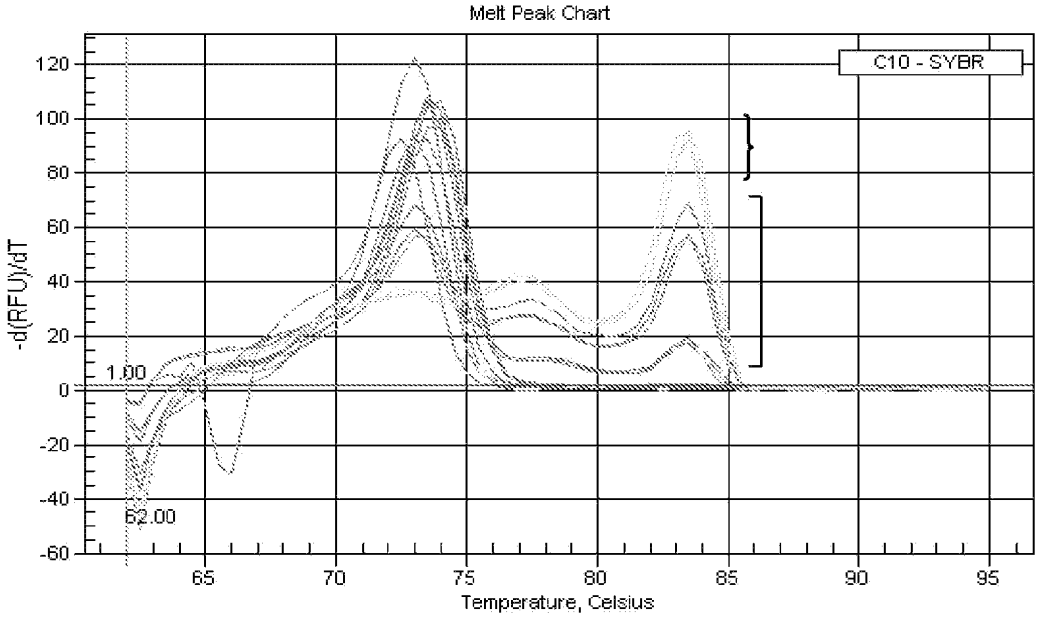


FIGURE 4



IDT Exhibit 1020.2467

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 10/21897

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12Q 1/68, C12P 19/34; C07H 21/04 (2010.01)
USPC - 435/6, 435/91.51, 435/91.53, 536/23.1
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
USPC:435/6, 435/91.51, 435/91.53, 536/23.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PubWEST DB=PGPB,USPT,USOC,EPAB,JPAB; PLUR=NO; OP=ADJ, Google Scholar, Google Patent
Search Terms Used: exo-sample, nucleotide, RnaseH, reverse transcriptase

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 20030228620 A1 (Du Breuil Lastrucci) 11 December 2003 (11.12.2003) entire document, esp: abstract, para [0024], [0025], [0026], [0040], [0116], [0117], [0119].	1-6
Y	US 6,518,026 B2 (Hartley) 11 February 2003 (11.02.2003) entire document, esp col 12 ln 17-34.	1-6
A	US 5,229,283 A (Berninger) 20 July 1993 (20.07.1993) entire document.	1-6
A	US 5,334,515 A (Rashtchian, et al.) 02 August 1994 (02.08.1994) entire document.	1-6
A	US 6,063,608 A (Kotewicz, et al.) 16 May 2000 (16.05.2000) entire document.	1-6

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
06 April 2010 (06.04.2010)	22 APR 2010
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (July 2009)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 10/21897

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 7-27
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

IDT Exhibit 1020.2469

Espacenet

Bibliographic data: WO2010129937 (A2) — 2010-11-11

METHODS FOR DETECTING GENETIC VARIATIONS IN DNA SAMPLES

Inventor(s): RAYMOND CHRISTOPHER K [US]; MAGNUS JILL F [US] ± (RAYMOND, CHRISTOPHER, K, ; MAGNUS, JILL, F)

Applicant(s): LIFE TECHNOLOGIES CORP [US]; RAYMOND CHRISTOPHER K [US]; MAGNUS JILL F [US] ± (LIFE TECHNOLOGIES CORPORATION, ; RAYMOND, CHRISTOPHER, K, ; MAGNUS, JILL, F)

Classification: **- international:** ***C12N15/11; C12Q1/68***
- cooperative: **C12Q1/686** more

Application number: WO2010US34145 20100507 Global Dossier

Priority number (s): US20090241352P 20090910 ; US20090176806P 20090508

Also published as: WO2010129937 (A3) US2011003301 (A1)

Abstract of WO2010129937 (A2)

The invention provides methods, compositions and kits for detecting genetic variation in a DNA sample at one or more polymorphic loci of interest. In some embodiments, the invention provides methods, compositions, and kits for determining the nucleotide present at a single nucleotide variant position of interest in a test sample.

IDT Exhibit 1020.2470

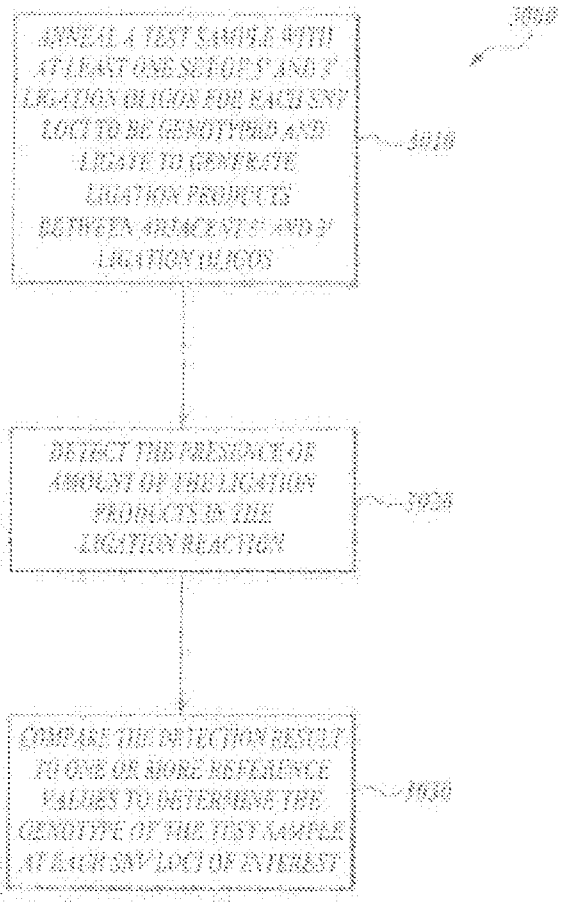


Fig. 9

IDT Exhibit 1020.2471

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 November 2010 (11.11.2010)

PCT

(10) International Publication Number
WO 2010/129937 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01) *C12N 15/11* (2006.01)

(21) International Application Number:
PCT/US2010/034145

(22) International Filing Date:
7 May 2010 (07.05.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/176,806 8 May 2009 (08.05.2009) US
61/241,352 10 September 2009 (10.09.2009) US

(71) Applicant *(for all designated States except US)*: **LIFE TECHNOLOGIES CORPORATION** [US/US]; 5791 Van Allen Way, Carlsbad, CA 92008 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **RAYMOND, Christopher, K.** [US/US]; 2626 NE 86th Street, Seattle, WA 98115 (US). **MAGNUS, Jill, F.** [US/US]; 11528 NE 39th Avenue, Seattle, WA 98125 (US).

(74) Agent: **GERMAN, Roberta**; Life Technologies Corporation, c/o Intellevate Corporation, P.O. Box 52050, Minneapolis, MN 55402 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO,

[Continued on next page]

(54) Title: METHODS FOR DETECTING GENETIC VARIATIONS IN DNA SAMPLES

(57) Abstract: The invention provides methods, compositions and kits for detecting genetic variation in a DNA sample at one or more polymorphic loci of interest. In some embodiments, the invention provides methods, compositions, and kits for determining the nucleotide present at a single nucleotide variant position of interest in a test sample.

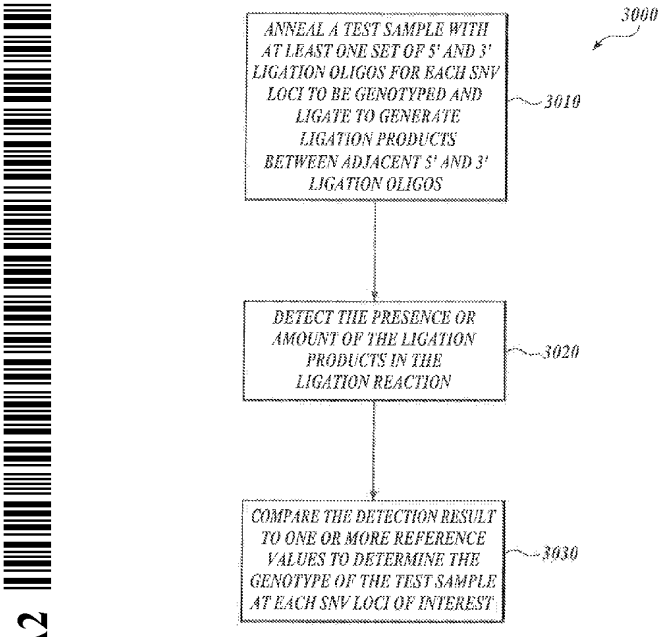


Fig. 9.

WO 2010/129937 A2

DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

METHODS FOR DETECTING GENETIC VARIATIONS IN DNA SAMPLES

This application claims the filing date benefit of U.S. Provisional Application Nos.: 61/176,806, filed on May 8, 2009, and 61/241,352, filed on September 10, 2009. The contents of each foregoing patent applications are incorporated by reference in their entirety.

Throughout this application various publications are referenced. The disclosures of these publications in their entireties are hereby incorporated by reference into this application in order to more fully describe the state of the art to which this invention pertains.

BACKGROUND

Molecular profiling will be a key technology in achieving personalized medicine, such as personalized oncology health therapies. The genomes of all mammalian subjects, including humans, undergo spontaneous mutations during the course of evolution. The majority of such mutations create polymorphisms, such that the mutated sequence and the initial sequence co-exist in the species population. The majority of DNA base differences are functionally inconsequential because they do not affect the amino acid sequence of encoded proteins and/or they do not affect the expression levels of the encoded proteins. However, some polymorphisms that lie within genes or their promoters do have a phenotypic effect, such as physical appearance, disease susceptibility, disease resistance, and responsiveness to drug treatments. Single nucleotide polymorphisms (SNPs) represent the most frequent type of human population DNA variation. Other forms of variation include copy number variation (CNVs), as well as short tandem repeats (e.g., microsatellites), long tandem repeats (e.g., minisatellite), and other insertions and deletions.

The study of complex genomes, and in particular, the search for the genetic basis of disease in humans, requires genotyping on a massive scale, which is demanding in terms of cost, time, and labor. Such costly demands are even greater when the methodology employed involves serial analysis of individual DNA samples, i.e., separate reactions for individual samples. Resequencing of polymorphic areas in the genome that are linked to disease development will contribute greatly to the understanding of diseases, such as cancer, and therapeutic development. While high-throughput sequencing platforms (e.g., a flow cell for massively parallel sequencing) provide a vast quantity of data with regard to disease-associated patterns of genetic variation on a genome-wide scale, this capability comes at the cost of a higher error rate than has been associated with traditional DNA sequencing platforms. Therefore, follow-on validation of primary

sequencing results is often carried out using labor intensive technologies such as locus-by-locus PCR and capillary resequencing in order to validate potential mutations, such as single nucleotide variants (SNVs).

Thus, there is a need for accurate, high-throughput, and cost-effective methods for high-throughput genotyping of target regions of the genome and/or transcriptome for pharmacogenetics applications, genetic disease association studies, and for validation of cell mutations detected in sequencing.

SUMMARY

This summary is provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. This summary is not intended to identify key features of the claimed subject matter, nor is it intended to be used as an aid in determining the scope of the claimed subject matter.

In one aspect, the present invention provides a method of determining the genotype of a test sample at one or more polymorphic loci of interest, the method comprising: (a) contacting in a reaction mixture, a test sample comprising one or more polymorphic loci of interest within one or more target nucleic acid region(s) of interest with one or more set(s) of query oligonucleotides, wherein each set of query oligonucleotides comprises: (i) at least one 5' ligation oligonucleotide comprising, from the 5' to 3' end, a first PCR primer binding region, a target-specific binding region selected to hybridize 5' of a polymorphic locus of interest, and a 3' region chosen to hybridize to either a consensus or variant nucleotide sequence at the polymorphic locus of interest; and (ii) a phosphorylated 3' ligation oligonucleotide comprising, from the 5' to 3' end, a target-specific binding region selected to hybridize 3' of the polymorphic locus of interest and a second PCR primer binding region, under conditions that allow hybridization between the query oligonucleotides and the target nucleic acid region(s) of interest; (b) contacting the reaction mixture of step (a) with DNA ligase under conditions suitable to ligate the 5' ligation oligonucleotides having a 3' region that hybridizes to the nucleotide sequence present at the polymorphic locus of interest in the test sample and the adjacent 3' phosphorylated ligation oligonucleotides, thereby generating a plurality of ligation products indicative of the genotype of the test sample at the one or more polymorphic loci of interest; and (c) measuring the amount of the ligation products in the reaction mixture of step (b). In some embodiments, the one or more polymorphic loci of interest comprise one or more SNV position(s) of interest. In some

embodiments, the test sample comprising one or more polymorphic loci of interest within one or more target nucleic acid region(s) of interest is contacted with a thermostable DNA ligase and one or more set(s) of query oligonucleotides.

In another aspect, the present invention provides a method of genotyping a test sample at one or more single nucleotide variant(s) (SNVs) position(s) of interest, the method comprising: (a) for each SNV position of interest, contacting in three separate reaction mixtures: (i) a synthetic template comprising the target region of interest having a consensus nucleotide at the SNV position of interest; (ii) a synthetic template comprising the target region of interest having a variant nucleotide at the SNV position of interest; and (iii) a test sample comprising the target region of interest comprising the SNV position of interest to be genotyped; with one or more set(s) of SNV query oligonucleotides, each set comprising: (i) a pair of allele-specific 5' ligation oligonucleotides, the pair comprising a first 5' ligation oligonucleotide comprising, from the 5' to 3' end, a first PCR primer binding region, a target-specific binding region selected to hybridize 5' of the SNV nucleotide position of interest, and a 3' region chosen to hybridize to the consensus nucleotide sequence at the SNV position of interest and a second 5' ligation oligonucleotide comprising, from the 5' to 3' end, a first PCR primer binding region, a target-specific binding region selected to hybridize 5' of the SNV nucleotide position of interest, and a 3' region chosen to hybridize to the variant nucleotide sequence at the SNV position of interest; and (ii) a phosphorylated 3' ligation oligonucleotide comprising from the 5' to 3' end, a target-specific binding region selected to hybridize 3' of the SNV position of interest and a second PCR primer binding region, under conditions that allow hybridization between the SNV query oligonucleotides and the nucleic acid target regions of interest; (b) contacting the three separate reaction mixtures of step (a) with DNA ligase under conditions suitable to ligate the 5' ligation oligonucleotides having a 3' region that hybridizes to the nucleotide sequence present at the SNV nucleotide position of interest in the synthetic templates and test samples and the adjacent 3' phosphorylated ligation oligonucleotides, thereby generating three separate ligation mixtures; and (c) measuring the amount of the ligation products in each of the three ligation mixtures of step (b). In some embodiments, the synthetic template comprising the target region of interest having a consensus nucleotide at the SNV position of interest, the synthetic template comprising the target region of interest having a variant nucleotide at the SNV position of interest, and the test sample comprising the target nucleic acid region(s) of interest comprising the SNV position of interest to be genotyped, are separately contacted with a thermostable DNA ligase and the one or more set(s) of

query oligonucleotides. In some embodiments, step (c) comprises amplification of the ligation products with a plurality of detection primer pairs, each pair comprising a forward PCR primer that binds to the first PCR primer binding region in the 5' ligation oligonucleotide and a reverse PCR primer that binds to the second PCR primer binding region in the 3' ligation oligonucleotide.

In another aspect, the present invention provides a method of producing a multi-well container comprising a matrix of detection primer pairs for decoding a multiplexed assay, the method comprising: (a) designing a plurality of detection primer pairs, each pair comprising a forward primer and a reverse primer for amplifying a target nucleic acid molecule of interest comprising a 5' primer binding region and a 3' primer binding region, wherein each forward primer comprises a 5' region that hybridizes to the 5' primer binding region of the target nucleic acid molecule of interest and a 3' region selected to avoid primer-dimer formation with the reverse primer; and wherein each reverse primer comprises a 5' region that hybridizes to the 3' primer binding region of the target nucleic acid molecule of interest and a 3' region selected to avoid primer-dimer formation with the forward PCR primer; and (b) dispensing each of the plurality of detection primer pairs into a well in a multi-well container comprising an ordered array of wells arranged in a matrix comprising a plurality of perpendicular rows distributed along the vertical axis of the container and a plurality of columns distributed along the longitudinal axis of the container, such that each well in the matrix is positionally addressable.

In another aspect, the present invention provides a kit for genotyping a test sample at one or more polymorphic loci of interest, the kit comprising at least one set of query oligonucleotides for genotyping a polymorphic loci of interest, the set comprising: (i) at least one 5' ligation oligonucleotide comprising, from the 5' to 3' end, a first PCR primer binding region, a target-specific binding region selected to hybridize 5' of the polymorphic loci of interest, and a 3' region chosen to hybridize to either a consensus or variant nucleotide sequence at the polymorphic loci of interest; and (ii) a phosphorylated 3' ligation oligonucleotide comprising from the 5' to 3' end, a target-specific binding region selected to hybridize 3' of the polymorphic loci of interest and a second PCR primer binding region.

The methods and kits of the invention can be used to genotype a haploid or diploid test sample for the presence or absence of one or more genetic variations, such as an insertion of one or more nucleotides, a deletion of one or more nucleotides, one or more single nucleotide variants (SNVs), one or more duplications, one or more inversions, one or more translocations, one or more repeat sequence expansions or contractions (i.e., changes in microsatellite sequences) at one

or more polymorphic loci of interest within a target region of interest. The multi-well containers (e.g., assay plates) of the present invention can be used to measure the presence or amount of one or more target nucleic acid molecules of interest, such as ligation products generated from a multiplexed ligation-dependent genotyping assay according to various embodiments of the methods of the invention.

DESCRIPTION OF THE DRAWINGS

The foregoing aspects and many of the attendant advantages of this invention will become more readily appreciated as the same become better understood by reference to the following detailed description, when taken in conjunction with the accompanying drawings, wherein:

FIGURE 1 illustrates a method for determining the genotype of a test sample at a single nucleotide variant (SNV) position of interest using a single allele-specific 5' ligation oligonucleotide, in accordance with an embodiment of the methods of the invention;

FIGURE 2 illustrates a method for determining the genotype of a test sample at a single nucleotide variant (SNV) position of interest using a pair of 5' allele-specific ligation oligonucleotides, in accordance with an embodiment of the methods of the invention, as described in Examples 1 and 3;

FIGURE 3 illustrates representative allele-specific 5' ligation oligonucleotides, 3' ligation oligonucleotides, and detection PCR primers for use in the methods, multi-well containers and kits of the invention;

FIGURE 4 illustrates exemplary reagents for use in a multiplexed ligation-dependent genotyping assay for genotyping a plurality of SNV positions of interest, wherein each assay for each target region of interest (e.g., Gene 1) is carried out with a pair of allele-specific 5' ligation oligos (300, 400), and a common 3' ligation oligo (500), and the quantitative PCR detection assay is carried out using corresponding detection PCR primer pairs (600, 700) disposed in a multi-well assay plate at discrete well locations (e.g., plate location A1, B1);

FIGURE 5A shows a perspective view of a representative multi-well container of the present invention comprising pairs of detection PCR primers arranged in a matrix for decoding a multiplexed assay comprising query oligonucleotides having regions complementary to the detection PCR primer pairs;

FIGURE 5B shows a portion of a transverse cross-section of the representative multi-well container shown in FIGURE 5A;

FIGURE 6 illustrates the decoding results obtained after carrying out a quantitative PCR assay in three separate, identical assay plates comprising detection PCR primer pairs arranged in a matrix, wherein (A) shows the assay results from a ligation mixture comprising a plurality of consensus synthetic templates and a pool of SNV query ligation oligos, (B) shows the assay results from a ligation mixture comprising a plurality of variant synthetic templates and the pool of SNV query ligation oligos, and (C) shows the assay results from a ligation mixture comprising a test sample and the pool of SNV query ligation oligos, wherein each well (826) on the assay plate (800) contains a unique pair of PCR primers corresponding to a set of target-specific SNV query oligonucleotides, such that adjacent wells (e.g., A1 and B1) provide the qPCR results for a pair of alleles at a particular SNV position of interest, as further illustrated in FIGURE 4 and described in Examples 2 and 3;

FIGURE 7 illustrates a method of enriching a population of DNA molecules for target regions of interest using capture probes (1200), in accordance with an embodiment of the methods of the invention, as described in Example 3;

FIGURE 8 is a flow chart showing the steps of a method for enriching a population of DNA molecules for target regions of interest with solution-based capture using target capture probes (1200), in accordance with an embodiment of the methods of the invention;

FIGURE 9 is a flow chart showing the steps of a ligation-dependent genotyping assay in accordance with various embodiments of the invention; and

FIGURE 10 is a flow chart showing the steps of a multiplexed ligation-dependent genotyping assay for simultaneously genotyping a plurality of SNV positions of interest, in accordance with various embodiments of the invention.

DETAILED DESCRIPTION

This section presents a detailed description of the many different aspects and embodiments that are representative of the inventions disclosed herein. This description is by way of several exemplary illustrations of varying detail and specificity. Other features and advantages of these embodiments are apparent from the additional descriptions provided herein, including the different examples. The provided examples illustrate different components and methodology useful in practicing various embodiments of the invention. The examples are not intended to limit the claimed invention. Based on the present disclosure, the ordinary skilled artisan can identify and employ other components and methodology useful for practicing the present invention.

I. Definitions

Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by one of ordinary skill in the art to which this invention belongs. Practitioners are particularly directed to Sambrook, et al., *Molecular Cloning: A Laboratory Manual*, 2d ed., Cold Spring Harbor Press, Plainsview, New York (1989); and Ausubel, et al., *Current Protocols in Molecular Biology* (Supplement 47), John Wiley & Sons, New York (1999), for definitions and terms of the art.

It is contemplated that the use of the term "**about**" in the context of the present invention is to connote inherent problems with precise measurement of a specific element, characteristic, or other trait. Thus, the term "about," as used herein in the context of the claimed invention, simply refers to an amount or measurement that takes into account single or collective calibration and other standardized errors generally associated with determining that amount or measurement. For example, a concentration of "about" 100 mM of Tris can encompass an amount of 100 mM ± .5 mM, if .5 mM represents the collective error bars in arriving at that concentration. Thus, any measurement or amount referred to in this application can be used with the term "about," if that measurement or amount is susceptible to errors associated with calibration or measuring equipment, such as a scale, pipetteman, pipette, graduated cylinder, etc.

The use of the words "**a**" or "**an**" when used in conjunction with the term "comprising" in the claims and/or the specification may mean "one," but it is also consistent with the meaning of "one or more," "at least one," and "one or more than one."

As used herein, the term "**nucleic acid molecule**" encompasses both deoxyribonucleotides and ribonucleotides and refers to a polymeric form of nucleotides including two or more nucleotide monomers. The nucleotides can be naturally occurring, artificial, and/or modified nucleotides.

As used herein, the term **"oligonucleotide"** refers to a single-stranded multimer of nucleotides of from about 10 to 200 nucleotides that is usually synthetic.

As used herein, an "**isolated nucleic acid**" is a nucleic acid molecule that exists in a physical form that is non-identical to any nucleic acid molecule of identical sequence as found in nature; "isolated" does not require, although it does not prohibit, that the nucleic acid so described has itself been physically removed from its native environment. For example, a nucleic acid can be said to be "isolated" when it includes nucleotides and/or internucleoside bonds not found in nature. When, instead, composed of natural nucleosides in phosphodiester linkage, a nucleic acid

can be said to be "isolated" when it exists at a purity not found in nature, where purity can be adjudged with respect to the presence of nucleic acids of other sequences, with respect to the presence of proteins, with respect to the presence of lipids, or with respect to the presence of any other component of a biological cell, or when the nucleic acid lacks a sequence that flanks an otherwise identical sequence in an organism's genome, or when the nucleic acid possesses a sequence not identically present in nature. As so defined, "isolated nucleic acid" includes nucleic acids integrated into a host cell chromosome at a heterologous site, recombinant fusions of a native fragment to a heterologous sequence, recombinant vectors present as episomes, or as integrated into a host cell chromosome.

As used herein, "**subject**" refers to an organism or to a cell sample, tissue sample, or organ sample derived therefrom, including, for example, cultured cell line, biopsy, blood sample, or fluid sample containing a cell. For example, an organism may be an animal, including but not limited to, an animal such as a cow, a pig, a mouse, a rat, a chicken, a cat, a dog, etc., and is usually a mammal, such as a human.

As used herein, the term "**specifically bind**" refers to two components (e.g., target-specific binding region and target) that are bound (e.g., hybridized, annealed, complexed) to one another sufficiently that the intended capture and enrichment steps can be conducted. As used herein, the term "specific" refers to the selective binding of two components (e.g., target-specific binding region and target) and not generally to other components unintended for binding to the subject components.

As used herein, the term "**high stringency hybridization conditions**" means any condition in which hybridization will occur when there is at least 95%, preferably about 97% to 100% nucleotide complementarity (identity) between the nucleic acid sequences of the nucleic acid molecule and its binding partner. However, depending upon the desired purpose, the hybridization conditions may be "medium stringency hybridization," which can be selected that require less complementarity, such as from about 50% to about 90% (e.g., 60%, 70%, 80%, 85%). The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm of Karlin and Altschul (*Proc. Natl. Acad. Sci. USA 87*:2264-2268 (1990)), modified as in Karlin and Altschul (*Proc. Natl. Acad. Sci. USA 90*:5873-5877 (1993)). Such an algorithm is incorporated into the NBLAST and XBLAST programs of Altschul et al. (*J. Mol. Biol. 215*:403-410 (1990)).

As used herein, the term "**complementary**" refers to nucleic acid sequences that are capable of base-pairing according to the standard Watson-Crick complementary rules. That is, the larger purines will base pair with the smaller pyrimidines to form combinations of guanine paired with cytosine (G:C) and adenine paired with either thymine (A:T) in the case of DNA, or adenine paired with uracil (A:U) in the case of RNA.

As used herein, the term "**target nucleotide**" refers to a nucleic acid molecule or polynucleotide in a starting population of nucleic acid molecules having a target sequence whose presence and/or amount and/or nucleotide sequence is desired to be determined and which has an affinity for a given ligation oligonucleotide. Examples of targets include regions of genomic DNA, PCR amplified products derived from RNA or DNA, DNA derived from RNA or DNA, ESTs, cDNA, and mutations, variants or modifications thereof.

As used herein, the term "**target sequence**" refers generally to a nucleic acid sequence on a single strand of nucleic acid. The target sequence may be a portion of a gene, a regulatory sequence, genomic DNA, cDNA, RNA including mRNA and rRNA, or others. The target sequence may be a target sequence from a sample, or a secondary target, such as a product of an amplification reaction.

As used herein, the term "**predetermined nucleic acid sequence**" means that the nucleic acid sequence of a nucleic acid probe is known and was chosen before synthesis of the nucleic acid molecule in accordance with the invention disclosed herein.

As used herein, the term "**essentially identical**" as applied to synthesized and/or amplified nucleic acid molecules refers to nucleic acid molecules that are designed to have identical nucleic acid sequences, but that may occasionally contain minor sequence variations in comparison to a desired sequence due to base changes introduced during the nucleic acid molecule synthesis process, amplification process, or due to other processes in the method. As used herein, essentially identical nucleic acid molecules are at least 95% identical to the desired sequence, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99% identical, or absolutely identical, to the desired sequence.

As used herein, the term "**resequencing**" refers to a technique that determines the sequence of a genome of an organism using a reference sequence that has already been determined. It should be understood that resequencing may be performed on both the entire genome/transcriptome of an organism or a portion of the genome/transcriptome large enough to include the genetic change of the organism as a result of selection. Resequencing may be carried out

using various sequencing methods, such as any sequencing platform amenable to producing DNA sequencing reads that can be aligned back to a reference genome, and is typically based on highly parallel technologies such as, for example, dideoxy "Sanger" sequencing, pyrosequencing on beads (e.g., as described in U.S. Patent No. 7211390, assigned to 454 Life Sciences Corporation, Brandord, CT), ligation based sequencing on beads (e.g., Applied Biosystems Inc./Invitrogen), sequencing on glass slides (e.g., Illumina Genome Analyzer System, based on technology described in WO 98/44151 (Mayer, P., and Farinelli, L.)), microarrays, or fluorescently labeled micro-beads.

As used herein, the term "**polymorphism**" refers to the occurrence of two or more genetically determined alternative sequences or alleles in a population. A "**polymorphic locus**" refers to the locus at which **genetic variation** occurs. A polymorphic locus can include any type of genetic variation, such as an insertion of one or more nucleotides, a deletion of one or more nucleotides, one or more single nucleotide variants (SNVs), one or more duplications, one or more inversions, one or more translocations, one or more repeat sequence expansions or contractions (i.e., changes in microsatellite sequences). In some embodiments, a polymorphic locus can be as small as one base pair (single nucleotide variant (SNV), which encompasses a single nucleotide polymorphism (SNP)). The first identified allele of a polymorphic locus is arbitrarily designated as the "**consensus**" allele and the other allele is designated as the "**variant**" (also sometimes referred as a "mutant") allele. Typically, a polymorphic locus has at least two alleles, each occurring at a frequency of greater than 1% of a selected population.

The allele occurring most frequently in a selected population is sometimes referred to as the "wild-type" or "consensus" allele. Diploid organisms may be homozygous or heterozygous for the variant allele. The variant allele may or may not produce an observable physical or biochemical characteristic ("phenotype") in an individual carrying the variant allele. For example, a variant allele may alter the enzymatic activity of a protein encoded by a gene of interest.

As used herein, the term "**genetic variation**" refers to genotypic differences among individuals in a population, at one or more polymorphic loci, and includes an insertion of one or more nucleotides, a deletion of one or more nucleotides, one or more single nucleotide sequence variations (SNVs), such as SNPs, copy number variation, such as one or more duplications, sequence rearrangements, such as one or more inversions, one or more translocations, or one or more repeat sequence expansions or contractions (i.e., changes in microsatellite sequences) at one or more polymorphic loci of interest within a target region of interest as compared to known reference sequences.

As used herein, the term "**single nucleotide variant**" or "**SNV**" refers to a DNA base within an established nucleotide sequence that differs from the known reference sequences. SNVs may be found within a patient sample (e.g., a tumor), they may or may not be present in unperturbed populations, and they include naturally occurring single nucleotide polymorphisms, also referred to as "SNPs".

As used herein, the term "**single nucleotide polymorphism**" or "**SNP**" refers to a single nucleotide position in a genomic sequence for which two or more alternative alleles are present at an appreciable frequency (e.g., at least 1%) in a population of organisms.

As used herein, the term "**genotype**" broadly refers to the genetic composition of an organism, including, for example, whether a diploid organism is heterozygous or homozygous for one or more single nucleotide variant alleles (SNVs) at a position of interest.

As used herein, the term "**haplotype**" refers to the identity of the nucleotide(s) that are present at a polymorphic position in the genome of a cell. For example, if the haplotype is bivariant (e.g., "A" and C," then the haplotypes are AA, CC and AC).

II. Aspects and Embodiments of the Invention

In accordance with the foregoing, in one aspect, the invention provides a method of determining the genotype of a test sample at one or more polymorphic loci of interest, the method comprising: (a) contacting in a reaction mixture, a test sample comprising one or more polymorphic loci of interest within one or more target nucleic acid region(s) of interest with one or more set(s) of query oligonucleotides, wherein each set of query oligonucleotides comprises: (i) at least one 5' ligation oligonucleotide comprising, from the 5' to 3' end, a first PCR primer binding region, a target-specific binding region selected to hybridize 5' of a polymorphic locus of interest, and a 3' region chosen to hybridize to either a consensus or variant nucleotide sequence at the polymorphic locus of interest, and (ii) a phosphorylated 3' ligation oligonucleotide comprising, from the 5' to 3' end, a target-specific binding region selected to hybridize 3' of the polymorphic locus of interest and a second PCR primer binding region, under conditions that allow hybridization between the query oligonucleotides and the target nucleic acid region(s) of interest; (b) contacting the reaction mixture of step (a) with DNA ligase under conditions suitable to ligate the 5' ligation oligonucleotides having a 3' region that hybridizes to the nucleotide sequence present at the polymorphic locus of interest in the test sample and the adjacent 3' phosphorylated ligation oligonucleotides, thereby generating a plurality of ligation products indicative of the

genotype of the test sample at the one or more polymorphic loci of interest; and (c) measuring the amount of the ligation products in the reaction mixture of step (b).

In some embodiments of the method, the hybridization and ligation steps are combined (i.e., coupled), wherein a test sample comprising one or more polymorphic loci of interest within one or more target nucleic acid region(s) of interest is contacted with a thermostable DNA ligase and one or more set(s) of query oligonucleotides. In other embodiments of the method, the hybridization and ligation reactions are carried out sequentially under separate reaction conditions (i.e., uncoupled), and may utilize either thermostable or non-thermostable DNA ligase.

The methods described herein may be used to detect any type of genetic variation, such as an insertion of one or more nucleotides, a deletion of one or more nucleotides, one or more single nucleotide sequence variations (SNVs), such as SNPs, copy number variation, such as one or more duplications, sequence rearrangements, such as one or more inversions, one or more translocations, or one or more repeat sequence expansions or contractions (i.e., changes in microsatellite sequences) at the polymorphic loci of interest in either a haploid or diploid sample of interest as compared to known reference sequences.

In some embodiments, the genetic variation detected is a single nucleotide variation (SNV) at an SNV position of interest. As described in Examples 1 and 3, it has been determined that the sensitivity of the assay methods described herein for a single mismatch adjacent the ligation site can be used to distinguish between two sequences that differ only with respect to the single nucleotide at the SNV position of interest. Accordingly, it will be understood by those of skill in the art that the methods described and demonstrated herein for use in detecting single nucleotide variations can also be used to detect larger regions of genetic variation, such as insertions, deletions, and sequence rearrangements in a haploid or diploid sample of interest. It will therefore be understood by those of skill in the art that while the descriptions herein of methods, kits, and compositions are described with reference to the detection of single nucleotide variants (SNVs), the methods, compositions, and kits are not intended to be limited to detection of SNVs, and are generally applicable to the detection of any type of genetic variation at one or more polymorphic loci of interest. Non-limiting examples of polymorphic loci of interest that may be detected using the methods described herein include nucleotide insertions, deletions, duplications, inversions, translocations, and changes in microsatellite sequences (i.e., sequence expansions and contractions) wherein the methods described herein are suitable for detecting various types of DNA rearrangements in addition to detecting changes in a nucleotide base sequence.

FIGURE 1 illustrates an embodiment of this aspect of the method of the invention. As shown in FIGURE 1 at Step A, an assay is carried out using a set of query oligonucleotides (e.g., SNV query oligonucleotides) comprising two oligos per polymorphic locus of interest (e.g., SNV position of interest): an allele specific 5' ligation oligo and a common phosphorylated 3' ligation oligo, based on the following steps. As shown in Step A, the test diploid genome 10 has a first allele with an "A" nucleotide at the SNV position of interest 100, and a second allele with a "G" nucleotide at the SNV position of interest 100. At Step A, an allele-specific 5' ligation oligo 300 and 3' ligation oligo 500 are each hybridized to the target region of the test diploid genome 10 that contains the SNV position of interest 100 under conditions that allow hybridization between the SNV query oligos and the target nucleic acid region of interest. As shown in FIGURE 1, the 5' ligation oligo 300 comprises, from the 5' to 3' end, a first PCR primer binding region 302, a target-specific binding region 304 selected to hybridize immediately 5' of the SNV position of interest 100, and a 3' region 306 (shown as comprising nucleotide "T") that is complementary to the wild-type sequence "A" in the test genome (and, therefore, not complementary to the SNV sequence "G").

As further shown in FIGURE 1, the phosphorylated (P) 3' ligation oligo 500 comprises, from the 5' to 3' end, a target-specific binding region 504 selected to hybridize immediately 3' of the SNV position of interest 100, and a region 502 at the 3' end that contains a second PCR primer binding region.

As shown in FIGURE 1, Step A, a DNA ligase enzyme 50 is contacted with the annealed mixture. For example, a thermostable DNA ligase enzyme may be present in the annealed mixture, or a non-thermostable DNA ligase enzyme may be added after the mixture is annealed. As further shown in FIGURE 1, Step B, as a result of the ligation reaction, the adjacent oligos 300 and 500 that are annealed to the test genome with an "A" at SNV position 100, a ligation product 200 is formed that is indicative of the genotype of the test sample, with a "T" at SNV position 100, flanked by 5' primer binding region 302 and 3' primer binding region 502. In contrast, the oligo 300 that is annealed to the test genome with a "G" at SNV position 100, resulting in a mismatch, does not form a ligation product with the adjacent oligo 500, due to the mismatch.

As shown in FIGURE 1, Step C, the ligation product formed in Step B may be assayed by a quantitative PCR (qPCR) assay using a forward PCR primer 600 that binds to the 5' primer binding region 302 and a reverse PCR primer 700 that binds to the 3' primer binding region 502 on

the ligation product 200. However, as illustrated in the representative graph shown in Step C, it is more difficult to distinguish between a homozygote (e.g., AA or GG) and a heterozygote (e.g., AG) using the two oligo per SNV assay approach, with one 5' ligation oligo specific for a given allele in a diploid organism. In haploid organisms, the expected genotype will only be consensus or variant, and not potentially a heterozygous blend of the two as is found in a diploid organism such as a human.

In another embodiment of the method, each set of query oligonucleotides (e.g., SNV query oligonucleotides) according to step (a) comprises a pair of allele-specific 5' ligation oligonucleotides for each SNV position of interest, the pair comprising a first 5' ligation oligonucleotide comprising a 3' region chosen to hybridize to the consensus nucleotide sequence at the SNV position of interest and a second 5' ligation oligonucleotide comprising a 3' region chosen to hybridize to the variant nucleotide sequence at the SNV position of interest. In accordance with this embodiment, as shown in FIGURE 2, an assay is carried out using a set of SNV query oligonucleotides comprising three primers per SNV position of interest: an allele specific 5' ligation oligo that binds to the consensus sequence at the SNV position of interest 100, an allele specific 5' ligation oligo that binds to the variant sequence at the SNV position of interest 100, and a common phosphorylated 3' ligation oligo. As shown in FIGURE 2, Step A, the test diploid genome 10 has a first allele with an "A" nucleotide at the SNV position of interest 100, and a second allele with a "G" nucleotide at the SNV position of interest 100. At Step A, the set of SNV oligonucleotides comprising the three ligation oligos: a 5' ligation oligo 300 that binds to the consensus sequence "A" at position 100, a 5' ligation oligo 400 that binds to the variant sequence "G" at position 100, and a common 3' ligation primer 500 are annealed to the region of the test diploid genome 10 that contains the SNV position 100.

The 5' ligation oligo 300 comprises, from the 5' to 3' end, a first PCR primer binding region 302, a target-specific binding region 304 selected to hybridize immediately 5' of the SNV position of interest 100, and a 3' region 306 (shown as comprising nucleotide "T") that is complementary to the wild-type sequence "A" in the test genome (and, therefore, not complementary to the SNV sequence "G").

The 5' ligation oligo 400 comprises, from the 5' to 3' end, a first PCR primer binding region 402, a target-specific binding region 404 selected to hybridize immediately 5' of the SNV position of interest 100, and a 3' region 406 (shown as comprising nucleotide "C") that is

complementary to the variant sequence "G" in the test genome (and, therefore, not complementary to the wild-type sequence "A").

As further shown in FIGURE 2, Step A, the phosphorylated (P) 3' common ligation oligo 500 comprises, from the 5' to 3' end, a target-specific binding region 504 selected to hybridize immediately 3' of the SNV position of interest 100, and a region 502 at the 3' end that contains a second PCR primer binding region.

As shown in FIGURE 2, Step A, ligase enzyme 50 is either present in the annealed mixture, or added to the annealed mixture, and as shown in FIGURE 2, Step B, as a result of the ligation reaction, a ligation product 200 is formed by ligating the oligo 300 that is annealed to the test genome with an "A" at SNV position 100 and the adjacent common oligo 500, with a "T" at SNV position 100, flanked by 5' primer binding region 302 and 3' primer binding region 502. In contrast, the oligo 300 that is annealed to the test genome with a "G" at SNV position 100, resulting in a mismatch, does not form a ligation product with the adjacent oligo 500, due to the mismatch. As further shown in FIGURE 2, Step B, a ligation product 250 is formed by ligating the oligo 400 that is annealed to the test genome with a "G" at position 100 and the adjacent common oligo 500, with a "C" at SNV position 100, flanked by 5' primer binding region 402 and 3' primer binding region 502.

As shown in FIGURE 2, Step C, the amount of the ligation products 200 and 250 formed in Step B may be measured by performing an assay, such as a quantitative PCR (qPCR) assay using a first set of primers: forward PCR primer 600 that has a region 602 that binds to the 5' primer binding region 302 and a reverse PCR primer 700 that binds to the 3' primer binding region 502 on the ligation product 200, and a second set of primers: forward PCR primer 600' that has a region 602 that binds to the 5' primer binding region 402 and a reverse PCR primer 700 that binds to the 3' primer binding region 502 on the ligation product 250. As illustrated in the representative graph shown in Step C, the use of the three query oligos per SNV assay allows a read-out of both alleles at a polymorphic site in a diploid sample. Therefore, homozygotes (e.g., AA or GG) are read out in one or the other PCR assays, while heterozygotes (e.g., AG), are read out in both PCR assays, thereby allowing for unambiguous identification of homozygotes and heterozygotes.

Query Oligonucleotides

As shown in FIGURE 3, the query oligonucleotides for use in the various embodiments of the ligation-dependent genotyping methods described herein (e.g., SNV query oligonucleotides),

include a 5' ligation oligonucleotide (300), a variant 5' ligation oligonucleotide (400), and a common phosphorylated 3' ligation oligonucleotide (500).

5' ligation oligonucleotides (300, 400)

As shown in FIGURE 3A, the 5' ligation consensus oligo 300 comprises, from the 5' to 3' end, a first PCR primer binding region 302, a target-specific binding region 304 selected to hybridize to the target nucleic acid region immediately 5' of the SNV position of interest 100, and a 3' region 306 that is complementary to the wild-type (i.e., consensus) sequence at the SNV position of interest.

As shown in FIGURE 3B, the 5' ligation variant oligonucleotide 400 comprises, from the 5' to 3' end, a first PCR primer binding region 402, a target-specific binding region 404 selected to hybridize immediately 5' of the SNV position of interest 100, and a 3' region 406 that is complementary to the variant (e.g., mutant) sequence at the SNV position of interest.

The length of each 5' ligation oligo (300, 400) is typically at least 40 nucleotides, such as at least 45 nucleotides, at least 50 nucleotides, at least 55 nucleotides, at least 60 nucleotides, at least 65 nucleotides, at least 70 nucleotides, up to a maximum length of about 200 nucleotides. In some embodiments, the 5' ligation oligos are each from about 45 nucleotides to about 70 nucleotides in length.

The target-specific binding region 304, 404, selected to hybridize immediately 5' of the SNV position of interest 100, is typically at least 10 nucleotides in length, such as at least 15 nucleotides, at least 20 nucleotides, at least 25 nucleotides, at least 30 nucleotides, at least 35 nucleotides, at least 40 nucleotides, at least 45 nucleotides, at least 50 nucleotides, up to 150 nucleotides in length. In some embodiments, the target-specific binding region 304, 404 is from about 20 to 30 nucleotides in length. Typically, the target-specific binding region 304, 404 is designed to have a sequence that is complementary, or substantially complementary, to the nucleic acid sequence contained in a region of interest immediately 5' of an SNV position of interest 100. In one embodiment, the target-specific binding region (304, 404) comprises a sequence that is 100% complementary to the target region 5' of the SNV position of interest.

In another embodiment, the target-specific binding region (304, 404) comprises a first region comprising the 20 nucleotides 5' (upstream) to the SNV position of interest that is 100% complementary to the target region, and a second region comprising from 21 nucleotides 5' (further upstream) of the SNV position of interest to the 5' end of the target-specific region (304, 404), wherein the second region comprises a sequence that is substantially complementary (i.e., at

least 90% identical, at least 95% identical, at least 96% identical, at least 97% identical, at least 98% identical or at least 99% identical) to the target region 5' of the SNV position of interest.

One of skill in the art can use art-recognized methods to determine the features of a target-specific binding region (304, 404) that will hybridize to the target region 5' of the SNV position of interest with minimal non-specific hybridization. For example, one of skill can determine experimentally the features such as length, base composition, and degree of complementarity that will enable a nucleic acid molecule (e.g., the target-specific binding region of a ligation oligo) to specifically hybridize to another nucleic acid molecule (e.g., the nucleic acid target) under conditions of selected stringency, while minimizing non-specific hybridization to other substances or molecules. The target-specific binding region may be designed to take into account genomic features of the target region, such as genetic variation (other than at the SNV position of interest), G:C content, predicted oligo Tm, and the like.

As shown in FIGURES 3A and 3B, the 5' ligation oligos further comprise a region 306, 406 at the 3' end of the oligo that has a sequence selected to hybridize to either the consensus (306) or variant (406) nucleotide present at the SNV position of interest. Typically, the region 306, 406 is a single nucleotide in length located at the 3' end of the ligation oligo 300, 400.

In some embodiments, the region 306, 406 is larger than a single nucleotide in length (e.g., from 2 nt to 1000 nt, 10,000 nt, 100,000 nt or larger), and is selected to detect a genetic variation, such as an insertion, a deletion, or a rearrangement (e.g., inversion, translocation) in the nucleotide sequence at the polymorphic position of interest, such as an SNV position of interest. It is noted that the methods described herein for detecting genetic variation at an SNV position of interest are not limited by the size of the polymorphic locus of interest, and may be used, for example, to detect the presence or absence of a rearrangement, such as a translocation event, between chromosomes in a haploid or diploid sample. In such embodiments, all that is required is the precise knowledge of the nucleotide sequence of the translocation break points.

3' ligation oligonucleotides (500)

As shown in FIGURES 3A and 3B, the phosphorylated (P) 3' common ligation oligo 500 comprises, from the 5' to 3' end, a target-specific binding region 504 selected to hybridize immediately 3' of the SNV position of interest 100, and a region 502 at the 3' end that contains a second PCR primer binding region. In operation, the 3' ligation primer 500 is typically phosphorylated at the 5' end prior to annealing to the test genome 10.

The length of each 3' ligation oligo (500) is typically at least 40 nucleotides, such as at least 45 nucleotides, at least 50 nucleotides, at least 55 nucleotides, at least 60 nucleotides, at least 65 nucleotides, at least 70 nucleotides, up to a maximum length of about 200 nucleotides. In some embodiments, the 3' ligation oligos are each from about 45 nucleotides to about 70 nucleotides in length.

The target-specific binding region 504, selected to hybridize to the target nucleic acid region starting at the nucleotide position immediately 3' of the SNV position of interest 100, is typically at least 10 nucleotides in length, such as at least 15 nucleotides, at least 20 nucleotides, at least 25 nucleotides, at least 30 nucleotides, at least 35 nucleotides, at least 40 nucleotides, at least 45 nucleotides, at least 50 nucleotides, up to 150 nucleotides in length. In some embodiments, the target-specific binding region 504 is from about 20 to 30 nucleotides in length. The target-specific binding region 504 is designed to have a sequence that is complementary, or substantially complementary, to the nucleic acid sequence contained in a region of interest immediately 3' of an SNV position of interest 100. In one embodiment, the target-specific binding region (504) comprises a sequence that is 100% complementary to the target region 3' of the SNV position of interest. In another embodiment, the target-specific binding region (504) comprises a sequence that is substantially complementary (i.e., at least 90% identical, at least 95% identical, at least 96% identical, at least 97% identical, at least 98% identical or at least 99% identical) to the target region 5' of the SNV position of interest.

In another embodiment, the target-specific binding region (504) comprises a first region comprising the 20 nucleotides downstream (3') of the SNV position of interest that is 100% complementary to the target region, and a second region comprising from 21 nucleotides further 3' of the SNV position of interest to the 3' end of the target-specific region (504), wherein the second region comprises a sequence that is substantially complementary (i.e., at least 90% identical, at least 95% identical, at least 96% identical, at least 97% identical, at least 98% identical, or at least 99% identical) to the target region 3' of the SNV position of interest.

The 5' ligation oligonucleotides (300) and variant 5' ligation oligonucleotides (400), each include PCR primer binding regions 302, 402 (also referred to as "primer tails") located at the 5' end of the oligos, for binding to forward PCR primers for use in a quantitative PCR assay. Similarly, the 3' ligation oligonucleotides (500) each include a PCR primer binding region 502 (primer tail) located at the 3' end of the oligo, for binding to reverse PCR primers.

The PCR primer binding regions 302, 402, and 502, are typically from about 10 to 50 nucleotides in length, such as at least 10 nucleotides in length, such as at least 15 nucleotides, at least 20 nucleotides, at least 25 nucleotides, at least 30 nucleotides, at least 35 nucleotides, at least 40 nucleotides, at least 45 nucleotides, or at least 50 nucleotides in length. In some embodiments, the PCR primer binding regions 302, 402, and 502 are from about 20 to 30, such as about 25 nucleotides in length.

In some embodiments, the 5' consensus ligation oligo 300 has a different primer binding region 302 than the primer binding region 402 of the 5' variant ligation oligo 400, to allow for detection of the presence or amount of the consensus ligation product 200 and the variant ligation product 250 in a single ligation reaction using two different sets of detection PCR primers, each set designed to detect either the consensus ligation product 200 or the variant ligation product 250.

In some embodiments, the ligation-dependent genotyping assay is a multiplexed assay comprising a plurality of sets of SNV query oligos for detecting a plurality of SNV positions of interest, such as at least 5, at least 10, at least 20, at least 40, at least 50, at least 80, at least 100, at least 200, at least 300, at least 500, at least 1000, at least 2,500, at least 5,000, at least 7,500 up to 10,000 more SNV positions of interest in a single ligation reaction. As illustrated in FIGURE 4, many different sets of SNV query oligos may be added to a genomic test sample, annealed, ligated, and assayed by qPCR. The results of each independent ligation are read out by unique, tail-specific PCR primer pairs designed to detect a particular ligation product. The advantage to this multiplexing approach is that very small amounts of precious starting material can be interrogated at many different potential mutation locations simultaneously.

For example, as shown in FIGURE 4, exemplary genes 1-6 in a target sample, each including a SNV position of interest, are assayed in a single reaction by pooling six sets of SNV query oligos, each set comprising a 5' consensus ligation oligo (300), a 5' variant ligation oligo (400), and a 3' ligation oligo (500), for each gene of interest. As further shown in FIGURE 4, each ligation oligo has a tail-specific PCR primer, for example, the forward PCR primers designated R1 to R8 (R stands for "row"), are designed to bind to the tail regions of 5' ligation primers (consensus and variant) for genes 1-4, with a common reverse PCR primer, designated C1 (C stands for "column") designed to bind to the tail region of the 3' ligation primers for genes 1-4. As further shown in FIGURE 4, a unique combination of PCR primers (e.g., R1 + C1) at a particular location in an assay plate (e.g., A1) is used to amplify a particular ligation product (e.g., Gene 1 consensus) contained in the multiplexed ligation mixture. The PCR primer pairs can be dispensed

into individual wells of a multi-well container, thereby allowing for ease of detection of the presence and/or amount of each ligation product in the multiplexed ligation reaction at a designated location.

As shown in FIGURE 9, the ligation-dependent genotyping assay 3000 includes the step 3010 of annealing a test sample with a least one set of SNV query oligonucleotides comprising at least one of a 5' ligation oligonucleotide (300) and/or (400), and a 3' ligation oligonucleotide (500) for each SNV position to be genotyped and ligating the adjacent query oligos annealed to the test sample, detecting the presence or amount of the ligation products at step 3020, and optionally, comparing the detection result to one or more reference values at step 3030 to determine the genotype of the test sample at each SNV loci of interest.

Test Samples

The methods of the invention are useful in any situation in which it is desired to detect one or more SNVs in a target nucleic acid sample (i.e., a haploid or diploid sample), such as, for example, to genotype a particular diploid subject, such as a human, with respect to one or more particular SNV positions of interest (e.g., in the context of determining whether the subject is likely to benefit from a particular therapeutic agent), to confirm the presence or absence of a variant nucleotide at a SNV position of interest that was initially detected during high-throughput sequence analysis, to compare a plurality of subjects of a particular species with respect to a particular target region of interest in order to identify new SNVs within the target region, or to monitor a subject with respect to a particular SNV position of interest over time (e.g., in the context of a therapeutic treatment regime and/or for prognosis or progression of a particular disease, such as cancer).

Examples of a test sample containing one or more target nucleic acid sequence(s) of interest for use in the methods of the invention include genomic DNA, mRNA, tRNA, rRNA, cRNA, oligonucleotides, DNA derived from RNA or DNA, ESTs, cDNA, PCR amplified products derived from RNA or DNA, microRNA, shRNA, siRNA, and mutations, variants or modifications thereof. The starting sample containing nucleic acid molecules may be isolated from a subject, such as a cell sample, tissue sample, or organ sample derived therefrom, including, for example, cultured cell lines, biopsy, blood sample, or fluid sample containing cells. The subject may be an animal, including, but not limited to, an animal such as a cow, a pig, a mouse, a rat, a chicken, a cat, a dog, etc., and is usually a mammal, such as a human. The methods of the invention are also

useful to genotype SNV locations of interest in a test sample containing a haploid genome, such as a yeast strain, as demonstrated in Example 7.

Samples containing a target nucleic acid sequence of interest to be genotyped, such as genomic DNA or RNA (e.g., mRNA, rRNA, tRNA, total RNA, microRNA), can be prepared by any of a variety of procedures. In some embodiments, the starting sample comprises genomic DNA. The genomic DNA sample may contain total genomic DNA, intact, fragmented, or enzymatically amplified portions of the same. Genomic DNA can be prepared using routine methods known in the art, (see, e.g., Sambrook, et al., *Molecular Cloning: A Laboratory Manual*, 2d ed., Cold Spring Harbor Press, Plainsview, New York (1989); and Ausubel, et al., *Current Protocols in Molecular Biology* (Supplement 47), John Wiley & Sons, New York (1999)).

In some embodiments, the starting sample comprises genomic DNA that has been amplified by whole genome amplification, using multiple displacement amplification, for example as described in Pan et al., *PNAS 40*(105):15499-15504, incorporated herein by reference.

Target Enrichment

In another embodiment, the starting sample comprises a population of nucleic acid molecules that has been enriched for one or more target regions of interest. In one embodiment, the enriched sample comprises PCR products amplified from a plurality of target-specific amplicons from a nucleic acid containing sample. In another embodiment, as illustrated in FIGURE 7, the sample is enriched for a target sequence containing an SNV position of interest 100, using solution-based capture methods from a library of DNA molecules 1000 comprising a subpopulation of nucleic acid target insert sequences of interest flanked by a first primer binding region 1022 and a second primer binding region 1032 within a larger population of nucleic acid insert sequences flanked by the first primer binding region and the second primer binding region.

The step of enriching a library for target sequences 100 with the population of DNA molecules 1000 may be carried out as illustrated in FIGURE 7. As shown in FIGURE 7, at step A, solution-based capture is carried out by first annealing a library of single-stranded capture probes 1200, each capture probe comprising a target specific region 1202 that hybridizes to a target sequence 100 contained in a library insert, with a library of nucleic acid molecules 1000 comprising nucleic acid target insert sequences of interest 100 flanked by a first primer binding region 1022 on one end and a second primer binding region 1032 on the other end. As further shown in FIGURE 7, step A, in one embodiment, the library of nucleic acid molecules 1000 is

annealed with a combination of a library of single-stranded capture probes 1200 each comprising a region 1204 that hybridizes to a universal adaptor oligo 1300 and an equimolar amount of universal adaptor oligos 1300 comprising a moiety 1310 for binding to a capture reagent 1400.

The annealing step for solution-based capture is typically carried out by mixing a molar excess of capture probes (or capture probes plus universal adaptor oligos) with the library in a high salt solution comprising from 100 mM to 2 M NaCl (osmolarity = 200 to 4000 molar). An exemplary high salt solution for annealing is 10 mM Tris pH 7.6, 0.1 mM EDTA, 1 M NaCl (osmolarity = 2000 molar). The nucleic acid molecules in the mixture are then denatured (i.e., by heating to 94 degrees) and allowed to cool to room temperature. In one embodiment, the annealing step is carried out in a high salt solution comprising from 100 mM to 2 M NaCl with the addition of 0.1% triton X100 (or Tween or NP40) nonionic detergent.

An amount of capture reagent 1400 is added to the annealed mixture sufficient to generate a plurality of complexes each containing a nucleic acid molecule, a capture probe (or a capture probe and a universal adaptor oligo), and a capture reagent. This step is carried out in a high salt solution comprising from 100 mM to 2 M NaCl (osmolarity = 200 to 4000 molar). An exemplary high salt solution for anneal is 10 mM Tris pH 7.6, 0.1 mM EDTA, 1 M NaCl (osmolarity = 2000 molar). The mixture is incubated at room temperature with mixing for about 15 minutes.

The complexes formed are then isolated or separated from solution with a sorting device 1500 (e.g., a magnet) that pulls or sorts the capture reagent 1400 out of solution.

The sorted complexes bound to the capture reagent 1400 are washed with a low salt wash buffer (less than 10 mM NaCl, and more preferably no NaCl) to remove non-target nucleic acids. An exemplary low salt wash buffer is 10 mM Tris pH 7.6, 0.1 mM EDTA (osmolarity = 10 millimolar). In some embodiments, the low salt wash optionally contains from 15% to 30% formamide, such as 25% formamide (osmolarity = 6.3 molar). For each wash step, the capture reagent 1400 bound to the complexes (e.g., magnetic beads) are resuspended in the low salt wash buffer and rocked for 5 minutes, then sorted again with the sorting device (magnet). The wash step may be repeated 2 to 4 times.

The nucleic acid molecules containing the target sequences are then eluted from the complexes bound to the capture reagent as follows. The washed complexes bound to the capture reagent 1400 are resuspended in water, or in a low salt buffer (i.e., osmolarity less than 100 millimolar), heated to 94°C for 30 seconds, the capture reagent (e.g., magnetic beads) is

pulled out using a sorting device (e.g., magnet), and the supernatant (eluate) containing the target nucleic acid molecules is collected.

The eluate may optionally be amplified in a PCR reaction with a first PCR primer that binds to the first primer binding site 1022 in the first linker and a second PCR primer that binds to the second primer binding site 1032 in the second linker, producing an enriched library which can be optionally sequenced.

The capture oligonucleotides 1200 may be designed to bind to a target region at selected positions spaced across the target region at various intervals. The capture oligo design and target selection process may also take into account genomic features of the target region such as genetic variation, G:C content, predicted oligo Tm, and the like. The length of a capture probe 1200 is typically in the range of from 10 nucleotides to about 200 nucleotides, such as from about 20 nucleotides to about 150 nucleotides, such as from about 30 nucleotides to about 100 nucleotides, such as from about 40 nucleotides to about 80 nucleotides.

The target-specific binding region 1202 of the target capture probe 1200 is typically from about 25 to about 150 nucleotides in length (e.g., 50 nucleotides, 100 nucleotides) and is chosen to specifically hybridize to a target sequence of interest. In one embodiment, the target-specific binding region 1202 comprises a sequence that is substantially complementary (i.e., at least 90% identical, at least 95% identical, at least 96% identical, at least 97% identical, at least 98% identical, at least 99% identical, or 100% identical) to a target sequence of interest.

In one embodiment, the capture probe 1200 is about 70 nucleotides in length, comprising a target-specific region of about 35 nucleotides in length.

One of skill in the art can use art-recognized methods to determine the features of a target binding region 1202 that will hybridize to the target region comprising the SNV position of interest 100 with minimal non-specific hybridization. For example, one of skill can determine experimentally the features such as length, base composition, and degree of complementarity that will enable a nucleic acid molecule (e.g., the target-specific binding region of a target capture probe) to specifically hybridize to another nucleic acid molecule (e.g., the nucleic acid target) under conditions of selected stringency, while minimizing non-specific hybridization to other substances or molecules. For example, for an exon target of interest, a target gene sequence is retrieved from a public database such as GenBank, and the sequence is searched for stretches of from 25 to 150 bp with a complementary sequence having a GC content in the range of 45% to 55%. The identified sequence may also be scanned to ensure the absence of potential secondary

structure and may also be searched against a public database (e.g., a BLAST search) to ensure a lack of complementarity to other genes.

In some embodiments, solution-based capture is used to enrich a population of nucleic acid molecules for one or more target polymorphic position(s) of interest, in order to determine the presence of a particular SNV, SNP, or deletion, addition, or other modification using the ligation-dependent genotyping assay described herein. In accordance with such embodiments, the set of target capture probes 1200 are typically designed such that there is a very dense array of capture probes that are closely spaced together such that a single target sequence, which may contain a mutation, will be bound by multiple capture probes that overlap the target sequence. For example, capture probes may be designed that cover every base of a target region, on one or both strands (i.e., head to tail) or that are spaced at intervals of every 2, 3, 4, 5, 10, 15, 20, 40, 50, 90, 100, or more bases across a sequence region.

As shown in FIGURE 8, the methods of solution-based capture 2000 include the step 2010 of providing a library of nucleic acid molecules comprising nucleic acid target insert sequences of interest flanked by a first primer binding region on one end and a second primer binding region on the other end. At step 2020, the library of nucleic acid molecules 1000 is annealed with a set of capture probes 1200, each capture probe comprising a region that hybridizes to a target sequence contained in a library insert. In one embodiment, the capture probes 1200 comprise a moiety 1310 (e.g., biotinylated) for binding to a capture reagent 1400 (e.g., streptavidin coated beads). In another embodiment, the library of nucleic acid molecules 1000 is annealed with a combination of a set of capture probes 1200, each comprising a region 1204 that hybridizes to a universal adaptor oligo 1300 and an equimolar amount of universal adaptor oligos 1300 comprising a moiety 1310 for binding to a capture reagent 1400.

The annealing step 2020 for solution-based capture is carried out by mixing a molar excess of capture probes (or capture probes plus universal adaptor oligos) with the library in a high salt solution comprising from 100 mM to 2 M NaCl (osmolarity = 200 to 4000 molar). An exemplary high salt solution for annealing is 10 mM Tris pH 7.6, 0.1 mM EDTA, 1 M NaCl (osmolarity = 2000 molar). The nucleic acid molecules in the mixture are then denatured (i.e., by heating to 94 degrees) and allowed to cool to room temperature. In one embodiment, the annealing step is carried out in a high salt solution comprising from 100 mM to 2 M NaCl with the addition of 0.1% triton X100 (or Tween or NP40) nonionic detergent.

At step 2030, an amount of capture reagent is added to the annealed mixture sufficient to generate a plurality of complexes each containing a nucleic acid molecule, a capture probe (or a capture probe and a universal adaptor oligo), and a capture reagent. This step is carried out in a high salt solution comprising from 100 mM to 2 M NaCl (osmolarity = 200 to 4000 molar). An exemplary high salt solution for anneal is 10 mM Tris pH 7.6, 0.1 mM EDTA, 1 M NaCl (osmolarity = 2000 molar). The mixture is incubated at room temperature with mixing for about 15 minutes.

At step 2040, the complexes formed in step 2030 are isolated or separated from solution with a sorting device 1500 (e.g., a magnet) that pulls or sorts the capture reagent 1400 out of solution.

At step 2050, the sorted complexes bound to the capture reagent 1400 are washed with a low salt wash buffer (less than 10 mM NaCl, and more preferably no NaCl) to remove non-target nucleic acids. An exemplary low salt wash buffer is 10 mM Tris pH 7.6, 0.1 mM EDTA (osmolarity = 10 millimolar). In some embodiments, the low salt wash optionally contains from 15% to 30% formamide, such as 25% formamide (osmolarity = 6.3 molar). For each wash step, the capture reagent 1400 bound to the complexes (i.e., magnetic beads) are resuspended in the low salt wash buffer and rocked for 5 minutes, then sorted again with the sorting device (magnet). The wash step may be repeated 2 to 4 times.

At step 2060, the nucleic acid molecules containing the target sequences are eluted from the complexes bound to the capture reagent as follows. The washed complexes bound to the capture reagent 1400 are resuspended in water, or in a low salt buffer (i.e., osmolarity less than 100 millimolar), heated to 94°C for 30 seconds, the capture reagent (i.e., magnetic beads) are pulled out using a sorting device (i.e., magnet), and the supernatant (eluate) containing the target nucleic acid molecules is collected.

At step 2070, the eluate is amplified in a PCR reaction with a first PCR primer that binds to the first primer binding site in the first linker and a second PCR primer that binds to the second primer binding site in the second linker, producing a once-enriched library which can be optionally genotyped at step 3000.

Alternatively, as shown in FIGURE 8, the once-enriched library may be further processed according to steps 2020–2070 using the same set of capture probes in each round of enrichment to generate a library that is twice-enriched, or three-times enriched, etc., for the target sequences of interest prior to performing a ligation-dependent genotyping assay 3000.

In one embodiment, the ratio of the concentration of the DNA target in the first and second round of enrichment to the concentration of capture oligo is a concentration of about 500 ng/ml DNA target to a concentration in the range of from about 1 nM to 10 nM of capture oligo. In one embodiment, the ratio of the concentration of DNA target in the third round of enrichment to concentration of capture oligo is a concentration of about 500 ng/ml of the twice-enriched library to a concentration of about 1 nM of capture oligo.

In one embodiment, the first round of enrichment (steps 2020–2070 shown in FIGURE 8) are carried out with a first set of capture probes designed to target a first set of targets, followed by a second round of enrichment that is carried out with a second set of capture probes designed to target a second set of targets.

In one embodiment, the capture reagent (1400) comprises streptavidin coated magnetic beads, each bead having a binding capacity of approximately 50 pmol of biotinylated double-stranded DNA/50 µl of beads. In one embodiment, at step 2030, about 50 µl of the streptavidin coated magnetic beads are added to about 5 µg of the annealed nucleic acids (e.g., in the first and second rounds of enrichment). In one embodiment, at step 2030, about 5 µl of the streptavidin coated magnetic beads are added to about 5 µg of the annealed nucleic acids (e.g., in the third round of enrichment).

Annealing and Ligation for the Ligation-Dependent Genotyping Assay

With reference to FIGURE 9, in one embodiment of the method, the annealing and ligation step 3010 of the ligation-dependent genotyping assay 3000 is carried out by mixing a set of SNV query oligonucleotides with the test sample comprising nucleic acids containing the target region of interest under conditions that allow hybridization between the SNV query oligonucleotides and the target nucleic acid region(s) of interest in the presence of a thermostable DNA ligase, and under conditions suitable to ligate the 5' ligation oligonucleotides having a 3' region that hybridizes to the nucleotide sequence present at the polymorphic locus of interest in the test sample and the adjacent 3' phosphorylated ligation oligonucleotides, thereby generating a plurality of ligation products indicative of the genotype of the test sample at the one or more polymorphic loci of interest.

In another embodiment of the method, the annealing and ligation step 3010 of the ligation-dependent genotyping assay 3000 is carried out by first mixing a set of SNV query oligonucleotides with the test sample comprising nucleic acids containing the target region of

interest under conditions that allow hybridization between the SNV query oligonucleotides and the target nucleic acid region(s) of interest, then contacting the annealed mixture with either a thermostable, or non-thermostable DNA ligase under conditions suitable to ligate the 5' ligation oligonucleotides having a 3' region that hybridizes to the nucleotide sequence present at the polymorphic locus of interest in the test sample and the adjacent 3' phosphorylated ligation oligonucleotides, thereby generating a plurality of ligation products indicative of the genotype of the test sample at the one or more polymorphic loci of interest.

Hybridizing conditions for hybridizing the SNV query oligos to the target nucleic acid molecules in the test sample are selected at a suitable stringency to achieve specific hybridization and are chosen based on the length of the target-specific binding region and the level of identity between the binding region and the target. The hybridization parameters that can be varied include salt concentration, buffer, pH, temperature, time of incubation, amount and type of denaturant, such as formamide, etc. (see, e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2d ed., Vols. 1–3, Cold Spring Harbor Press, New York, 1989; Hames et al., *Nucleic Acid Hybridization*, IL Press, 1985; Davis et al., *Basic Methods in Molecular Biology*, Elsevier Sciences Publishing, Inc., New York, 1986). The reaction conditions required to achieve specific interactions of the SNV query oligos and target nucleic acid molecules are routine and conventional in the art (e.g., as described in Niemeyer et al., *Nucleic Acid Res. 22*:5530-5539, 1994; Fodor et al., U.S. Patent No. 5,510,270; Pirrung et al., U.S. Patent No. 5,143,854, incorporated herein by reference).

In some embodiments, the hybridization step of a hybridization reaction followed by a ligation reaction, or a coupled hybridization/ligation reaction is carried out in a suitable reaction mixture comprising at least one monovalent cationic salt selected from the group consisting of KCl, NaCl and NH_4Cl, in order to stimulate annealing of the genotyping primers to the complementary genotyping primers, for example, as described in Example 5.

In some embodiments, the hybridization step of a hybridization reaction followed by a ligation reaction, or a coupled hybridization/ligation reaction is carried out in a suitable reaction mixture by incubating the mixture at an initial temperature greater than 90°C to denature the nucleic acids and gradually cooling to room temperature over a time period ranging from 30 minutes to 2 hours or longer, such as for at least 30 minutes, at least 60 minutes, at least 120 minutes, at least 170 minutes, or longer.

For example, hybridization of two binding partners may be carried out in a buffer such as, for example, 6X SSPE-T (0.9 M NaCl, 60 mM MaH_2PO_4, 6 mM EDTA and 0.05% Triton-X-100) for a time period from 10 minutes to at least 3 hours, at a temperature from about 4° to about 37°. In some embodiments of the invention, the reaction conditions can approximate physiological conditions. An exemplary solution for annealing is 10 mM Tris pH 7.6, 0.1 mM EDTA, 20 mM NaCl, as described in Example 1.

The amount of SNV query oligos added to the test sample per genotyping reaction is typically from about 1 pM to about 50 nM, such as from about 10 pM to about 5 nM, such as about 50 pM to about 1000 pM, such as from about 100 pM to about 500 pM. As described in Example 3, it was determined that SNV oligo concentrations in the range of 100 pM improved assay sensitivity by increasing the signal-to-noise ratio. The nucleic acids in the mixture are then denatured (i.e., by heating to 94 degrees) and allowed to cool to room temperature.

In one embodiment, a thermostable DNA ligase, such as, for example, Taq DNA ligase or 9°N DNA ligase, is utilized in the methods of the invention. The use of a thermostable ligase is advantageous because the enzyme activity is retained at the high temperatures needed for DNA melting and reannealing. In another embodiment, either a thermostable (such as Taq DNA ligase) or a non-thermostable DNA ligase (such as T4 DNA ligase) is added to the annealed mixture, as described in Example 5.

In accordance with some embodiments, a ligation reaction comprising a non-thermostable DNA ligase is typically incubated at a temperature ranging from about 15°C to about 45°C for a time period ranging from at least one minute to 30 minutes or longer (e.g., at least about 1 minute, at least 5 minutes, at least 10 minutes, at least 20 minutes, or at least 30 minutes).

In accordance with some embodiments, a ligation reaction comprising a thermostable DNA ligase is typically incubated at a temperature ranging from about 37°C to about 75°C for a time period ranging from at least one minute to 30 minutes or longer (e.g., at least about 1 minute, at least 5 minutes, at least 10 minutes, at least 20 minutes, or at least 30 minutes). In addition to the fact that thermostable ligases may be utilized at high temperatures, it has been determined that thermostable DNA ligases have greater specificity and preference for ligating nicks in dsDNA and have little ssDNA joining activity (i.e., randomly joining oligos together in the absence of template, such as a target nucleic acid of interest), whereas it has been determined that T4 DNA ligase, a non-thermostable ligase, joins oligos in the absence of template at a significant rate.

Detection of Ligation Products

At step 3020, the presence and/or amount of the ligation products in the ligation reaction are detected. The presence and/or amount of the ligation products in the ligation reaction may be determined using any suitable method of measurement. As used herein, the terms "determining," "measuring," "evaluating," "assessing," and "assaying" are used interchangeably to refer to any form of measurement, and include determining if an element, (e.g., such as the variant or consensus nucleotide, or the ligation product indicative of presence of the variant or consensus nucleotide), is present or not. These terms include both quantitative and/or qualitative determinations, which may be relative or absolute.

In one embodiment, the amount of the ligation products in the ligation reaction are measured using quantitative PCR (qPCR) comprising amplification of the ligation products with one or more pair(s) of detection primers with a DNA polymerase, each primer pair comprising a forward PCR primer that binds to the first PCR primer binding region in the 5' ligation oligonucleotide and a reverse PCR primer that binds to the second PCR primer binding region in the 3' ligation oligonucleotide. In such embodiments, it is noted that the tails 302, 402, 502 on the ligation primers 300, 400, and 500, respectively, containing primer binding sites for primers used for subsequent real-time quantitative PCR, can, in principle, be many different sequences. This allows for multiplexing of numerous assays to detect different SNVs in a single ligation reaction, as further described in Examples 3, 5, and 7.

In one embodiment, a fluorescent dye, such as SYBR green, is included in the qPCR reaction that intercalates with double-stranded DNA, causing fluorescence of the dye. An increase in DNA product during PCR therefore leads to an increase in fluorescence intensity and is measured at each cycle, thus allowing DNA concentration to be quantified. In order to reduce background levels due to the binding of the dye to non-specific PCR products, such as primer-dimers, in one embodiment, a paired set of primers is used for each PCR reaction, wherein the penultimate two or three nucleotides at the 3' end of the forward and reverse primers are selected to avoid primer-dimer formation.

In another embodiment, the qPCR reaction is carried out using a set of fluorescent reporter probes. An increase in the product targeted by the reporter probe occurs during each PCR cycle, therefore, causes a proportional increase in fluorescence.

Fluorescence is detected and measured in the real-time PCR thermocycler and its geometric increase corresponding to exponential increase of the product is used to determine the

threshold cycle (Ct) in each reaction. Relative concentrations of DNA present during the exponential phase of the reaction are determined by plotting fluorescence against cycle number on a logarithmic scale. A threshold for detection of fluorescence above background is determined. The cycle at which the fluorescence from a sample crosses the threshold is called the cycle threshold (Ct). Since the quantity of DNA doubles every cycle during the exponential phase, relative amounts of DNA can be calculated. For example, a sample whose Ct is 3 cycles earlier than another sample has $2^3 = 8$ times more template.

In some embodiments of the genotyping methods, as shown in FIGURE 9 at step 3030, the detection result is compared to one or more reference values obtained from one or more reference standards to determine the genotype of the test sample at each SNV position of interest, for example, using the methods described supra. The one or more reference values may be obtained by carrying out the ligation-dependent genotyping assay with one or more reference standards, such as a set of SNV query oligos and a pair of synthetic double-stranded templates comprising a target specific sequence region including either a consensus or variant nucleotide at the SNV position of interest, as described herein. The synthetic double-stranded templates containing an SNV position of interest are typically at least 30 to 200 nucleotides in length and may be generated by annealing complementary synthesized oligos, as described in Example 1. The SNV position of interest is typically located at or within 10 nucleotides of the middle of the synthetic template.

In another aspect, the present invention provides a method of genotyping a test sample at one or more single nucleotide variant(s) (SNVs) position(s) of interest, the method comprising: (a) for each SNV position of interest, contacting in three separate reaction mixtures: (i) a synthetic template comprising the target region of interest having a consensus nucleotide at the SNV position of interest; (ii) a synthetic template comprising the target region of interest having a variant nucleotide at the SNV position of interest; and (iii) a test sample comprising the target region of interest comprising the SNV position of interest to be genotyped; with one or more set(s) of SNV query oligonucleotides, each set comprising: (i) a pair of allele-specific 5' ligation oligonucleotides, the pair comprising a first 5' ligation oligonucleotide comprising, from the 5' to 3' end, a first PCR primer binding region, a target-specific binding region selected to hybridize 5' of the SNV nucleotide position of interest, and a 3' region chosen to hybridize to the consensus nucleotide sequence at the SNV position of interest and a second 5' ligation oligonucleotide comprising, from the 5' to 3' end, a first PCR primer binding region, a target-specific binding region selected to hybridize 5' of the SNV nucleotide position of interest, and a 3' region chosen to

hybridize to the variant nucleotide sequence at the SNV position of interest and (ii) a phosphorylated 3' ligation oligonucleotide comprising from the 5' to 3' end, a target-specific binding region selected to hybridize 3' of the SNV position of interest and a second PCR primer binding region, under conditions that allow hybridization between the SNV query oligonucleotides and the nucleic acid target regions of interest; (b) contacting the three separate reaction mixtures of step (a) with DNA ligase under conditions suitable to ligate the 5' ligation oligonucleotides having a 3' region that hybridizes to the nucleotide sequence present at the SNV nucleotide position of interest in the synthetic templates and test samples and the adjacent 3' phosphorylated ligation oligonucleotides, thereby generating three separate ligation mixtures; and (c) measuring the amount of the ligation products in each of the three ligation mixtures of step (b).

In some embodiments of the method, the hybridization and ligation steps are combined (i.e., coupled), wherein a test sample comprising one or more SNV positions of interest within one or more target nucleic acid region(s) of interest is contacted with one or more set(s) of query oligonucleotides in the presence of a thermostable DNA ligase. In other embodiments of the method, the hybridization and ligation reactions are carried out sequentially under separate reaction conditions (i.e., uncoupled), and may utilize either thermostable or non-thermostable DNA ligase.

The synthetic templates and SNV query oligos for a SNV position of interest may be generated as previously described herein.

As shown in FIGURE 10, an embodiment of the ligation-dependent multiplexed genotyping assay 4000 is carried out for a set of SNV positions of interest by providing a pool of one or more sets of SNV query oligos (i.e., an oligo pool) at step 4010, each set of SNV query oligos comprising 5' and 3' ligation oligos for each SNV position of interest to be genotyped. The pool at step 4010 may comprise at least 5 sets of SNV query oligos, at least 10, at least 20, at least 40, at least 50, at least 80, at least 100, up to 500 or more, wherein each set is designed to genotype an SNV position of interest.

At step 4022, the oligo pool according to step 4010 is annealed with a set of consensus reference templates corresponding to the SNV positions of interest and ligated in a first reaction vessel. At step 4024, the oligo pool according to step 4010 is annealed with a set of variant reference templates corresponding to the SNV positions of interest in the presence of DNA ligase and ligated in a second reaction vessel. At step 4026, the oligo pool according to step 4010 is annealed with a test sample comprising nucleic acid molecules having the SNV positions of

interest and ligated in a third reaction vessel. The annealing and ligation steps may be carried out as previously described herein.

At step 4032, the ligation mixture from step 4022 (consensus templates) is distributed over a multi-well container (e.g., a universal assay plate) comprising PCR detection primer pairs arranged in a matrix such that each well in the matrix is positionally addressable and contains a different detection primer pair, and a quantitative PCR assay is carried out in the multi-well container.

At step 4034, the ligation mixture from step 4024 (variant templates) is distributed over a multi-well container (e.g., a universal assay plate) comprising PCR detection primer pairs arranged in a matrix such that each well in the matrix is positionally addressable and contains a different detection primer pair, and a quantitative PCR assay is carried out in the multi-well container. In some embodiments, the PCR detection primer pairs in the matrix are minimally-interacting primer pairs, as described herein.

At step 4036, the ligation mixture from step 4026 (test sample) is distributed over a multi-well container (e.g., a universal assay plate) comprising PCR detection primer pairs arranged in a matrix such that each well in the matrix is positionally addressable and contains a different detection primer pair, and a quantitative PCR assay is carried out in the multi-well container. The multi-well containers used in steps 4032, 4034, and 4036, are separate, but substantially identical containers (i.e., each container contains the same primer pairs, arranged in the same grid pattern, so that the results of each assay can be compared side by side).

At step 4040, the quantitative PCR results obtained from step 4032 (consensus templates) and from step 4034 (variant templates) are used to calculate the reference values expected for a diploid genome containing homozygous consensus nucleotides, heterozygous nucleotides, or homozygous variant nucleotides, at each SNV position of interest. The quantitative PCR results may be raw cycle threshold (Ct) results (i.e., the cycle at which the fluorescence from a sample crosses the threshold), or may be processed results (such as those obtained by subtracting a background measurement, or by rejecting a reading for a feature which is below a predetermined threshold, normalizing the results, or the average Ct value of replicate samples, and the like). An exemplary method of calculating the reference values expected for a diploid genome using quantitative PCR results obtained from a pair of reference templates (consensus and variant) for each SNV position of interest is provided in Example 3.

At step 4050, the quantitative PCR results obtained from step 4036 (the test sample) are compared to the calculated reference values from step 4040 to determine the genotype of the test sample at each SNV position of interest, and assigning the genotype based on the closest pairing between the experimental value from the test sample and the calculated reference values for each potential genotype at each SNV position of interest. For example, genotyping with the consensus template may yield a Ct value of "25" in the consensus assay (assay with SNV consensus query oligos), and a Ct value of "30" in the variant assay (assay with SNV variant query oligos). The genotyping results are calculated as a result of the Ct of the variant (Ct(var)) minus the Ct of the consensus (Ct(cons)). Therefore, in the above example, the consensus template yields a Ct(var) - Ct(cons) of "30"-"25" = 5. The variant template in the above example yields a Ct(var)-Ct(cons) value of "25"-"30" = -5. Finally, a mixed template would be inferred to give a Ct(var) - Ct(cons) value of "25"-"25" = 0. Assuming the sample is a diploid, then a sample with a homozygous consensus base at the SNV position would be expected to yield a Ct(var) - Ct(cons) value of approximately "30"-"25" $\approx$ 5. Similarly, a homozygous variant base would be expected to yield a value of $\approx$ -5, and a heterozygous consensus plus variant would be expected to return a Ct(var) - Ct(cons) value of approximately zero. By comparing the actual Ct(var) - Ct(cons) value of the test sample in the genotyping assay to the reference templates, genotypes are assigned based on the closest numerical similarity to the homozygous consensus, homozygous variant or heterozygous consensus and variant values produced with the templates and the Ct(var) - Ct(cons) calculation.

Assessing the performance of the ligation-dependent genotyping assay

The performance of the ligation-dependent genotyping assay carried out using quantitative PCR may be evaluated by calculating the dynamic range of the assay as follows. The average Cts across replicate samples (e.g., quadruplicate wells) in the qPCR assay for each consensus and variant pair of an SNV assay set is calculated, wherein a Ct value below 30 is indicative of an informative qPCR assay. The Ct(variant)-Ct(consensus) = Δ consensus is calculated for each of the consensus template assays and the Ct(consensus)-Ct(variant) = Δ variant is calculated for each of the variant template assays. The sum of Δ consensus for the consensus template assays plus Δ variant for the variant template assays is then calculated. As described in Example 7, it was experimentally determined that if the sum of Δ consensus for the consensus template assays plus Δ variant for the variant template assays is ≥ 3, then genotyping calls can be made with confidence in diploid organisms.

Matrix of Detection Primer Pairs

In another aspect, the present invention provides a method of producing a multi-well container comprising a matrix of detection primer pairs for decoding a multiplexed assay, the method comprising: (a) designing a plurality of detection primer pairs, each pair comprising a forward primer and a reverse primer for amplifying a target nucleic acid molecule of interest comprising a 5' primer binding region and a 3' primer binding region, wherein each forward primer comprises a 5' region that hybridizes to the 5' primer binding region of the target nucleic acid molecule of interest and a 3' region selected to avoid primer-dimer formation with the reverse primer; and wherein each reverse primer comprises a 5' region that hybridizes to the 3' primer binding region of the target nucleic acid molecule of interest and a 3' region selected to avoid primer-dimer formation with the forward PCR primer; and (b) dispensing each of the plurality of detection primer pairs into a well in a multi-well container comprising an ordered array of wells arranged in a matrix comprising a plurality of perpendicular rows distributed along the vertical axis of the container and a plurality of columns distributed along the longitudinal axis of the container, such that each well in the matrix is positionally addressable. In one embodiment, the present invention provides multi-well containers comprising a matrix of detection primer pairs for decoding a multiplexed assay. In some embodiments, the detection primer pairs in the matrix are designed to be minimally-interacting primer pairs (i.e., primer pairs each comprising a 3' region selected to avoid primer-dimer formation), as described herein.

An exemplary multi-well container useful for carrying out the detection step of the genotyping assay is shown in FIGURE 5A. FIGURE 5A shows a perspective view of a representative multi-well container 800 of the present invention. The multiwell container 800 includes a body 802 including an upper surface 804, a lower surface 806 disposed opposite the upper surface 804, a right side 808, a left side 810 disposed opposite the right side 808, a top 814, and a bottom 812 disposed opposite the top 814. The container body 802 defines multiple wells 826. A lid 828 may optionally cover the upper surface 804. As shown in FIGURE 5A, the lid 828 includes a lid body 830 defining an outer surface 832, an inner surface 834 and a lip 836 that extends around the perimeter of lid body 830. In the embodiment shown in FIGURE 5A, the multi-well container 800 comprises an ordered array of individual wells 826. In the embodiment shown in FIGURE 5A, the ordered array of individual wells 826 is arranged in a matrix of a plurality of perpendicular rows distributed along the vertical axis (i.e., from the top 814 to the bottom 812) of the multi-well container 800 and a plurality of columns distributed along the

longitudinal axis (i.e., from the left side 810 to the right side 808). In exemplary embodiments, the multi-well container 800 includes a matrix having a dimension of 8 columns x 12 rows = 96 wells, or 16 columns x 24 rows = 384 wells, or 32 columns x 48 rows = 1536 wells.

As shown more clearly in FIGURE 5B, in the representative embodiment of the multi-well container 800, each well 826 is generally hemispherical and includes a well wall 838 defining a well lumen 840 that opens onto upper surface 804 of the container body 802 through the opening 842. The multiple wells 826 are sized for receiving and retaining aliquots (e.g., aliquots that each have a volume of from 1 µl to 1000 µl) of a liquid composition, such as a PCR reaction mixture.

In the exemplary embodiment shown in FIGURE 5A, the multi-well container 800 has a generally rectangular shape, but the multi-well container 800 can have any shape, such as square or circular. Similarly, the wells 826 can have any desired shape provided that they are capable of containing a liquid composition 844. The lid 828 is suitably dimensioned to fit over upper surface 814 of container body 802. The container body 802 and the lid 828 may be made from any suitable material, or mixtures of suitable materials. Typically, the container body 802 and the lid 828 are made from a material, such as plastic, that can withstand freezing and thawing at least once, as well as multiple cycling to temperatures up to at least 95°C (e.g., in a thermocycler). Exemplary containers include a 96 well assay plate, or a 384 well assay plate, such as commercially-available 96-well plastic plates manufactured by Island Scientific (7869 NE Day Rd West, Bainbridge Island, WA 98110), or by MWG Biotech (4170 Mendenhall Oaks Parkway, Suite 160, High Point, NC 27265), or optical plates from ABI for real time PCR analysis with the ABI 7900 (Applied Biosystems, Foster City, CA), or any other multi-well assay plate suitable for quantitative PCR assay analysis.

In one embodiment of the invention, the present invention provides a multi-well container 800 comprising a matrix of a plurality of compositions 844, each composition 844 comprising detection primer pairs dispensed into individual wells for decoding a multiplexed assay. The multi-well containers 800 are preferably produced en masse, easily stored, and reproducible, allowing multiple genotyping assays to be assayed and easily compared with each other.

In some embodiments, at least 20% of the wells 826 (e.g., at least 20% (e.g., at least 25%, at least 30%, at least 40%, at least 50%, at least 60%, or at least 70%, or at least 80%, or at least 90%, or all of the wells in the container) comprise a composition 844 comprising a PCR detection

primer pair, each pair comprising a forward PCR primer and a reverse PCR primer for amplifying a target nucleic acid molecule of interest flanked by a 5' primer binding region and a 3' primer binding region, wherein each forward PCR primer comprises a 5' region that hybridizes to the 5' primer binding region of the target nucleic acid molecule of interest and a 3' region selected to avoid primer-dimer formation with the reverse PCR primer; and wherein each reverse PCR primer comprises a 5' region that hybridizes to the 3' primer binding region of the target nucleic acid molecule of interest and a 3' region selected to avoid primer-dimer formation with the forward PCR primer, also referred to as "minimally interacting primer pairs".

In some embodiments, the 3' region of the minimally interacting forward and reverse primer pairs selected to avoid primer-dimer formation consists of from two to nine 3' terminal nucleotides (e.g., the last 2 nucleotides, the last 3 nucleotides, the last 4 nucleotides, the last 5 nucleotides, the last 6 nucleotides, the last 7 nucleotides, the last 8 nucleotides, or the last 9 nucleotides as measured from the 3' end) wherein the 3' terminal nucleotide sequence is selected to reduce background signal and provide the greatest possible dynamic range for genotyping assays.

In some embodiments, the 3' region of the forward and reverse primer pairs consists of the last two or three nucleotides at the 3' end of the respective oligonucleotides. In accordance with such embodiments, as described in Examples 2, 4, 6, and 7, the last two or three nucleotides at the 3' end of the primer pairs are designed with sequences that cannot pair with one another nor can they self anneal. For example, in one representative embodiment, each forward primer in a primer matrix is designed to end in the sequence "CT" and each reverse primer in the primer matrix is designed to end in the sequence "GA," as described in Example 2. In another representative embodiment, a set of forward primers in a primer matrix is designed to end in "ACA" and a set of reverse primers in the primer matrix is designed to end in "CAC," as described in Example 4. In some embodiments, candidate primers for use as minimally interactive primer pairs are further screened to eliminate primers containing sequences present within 9 nucleotides of the 3' end of the primer that would hybridize to the 3' terminal sequences, such as primers containing the sequence "GTG" or "TGT" within the last 9 nucleotides of the 3' end, as described in Example 4.

In another representative embodiment, a set of minimally interacting primer pairs is selected by first generating a set of candidate random 22-mer DNA sequences, screening the sequences for the presence of either "TTT" or "GGG" in the 3' terminal 6 nucleotides, and removing such candidate primers to generate a subset of candidate primers, adding the 3' terminal

sequence "CCC" to a first group of the subset of primers and adding the 3' terminal sequence "AAA" to a second group of the subset primers, to generate a set of candidate primer pairs, and performing a control assay with no template with the set of candidate primer pairs to identify primer pairs that generated a Ct value indicative of a low background level, such as a Ct value of greater than 35 (such as a Ct value greater than 36, a Ct value greater than 37, a Ct value greater than 38, a Ct value greater than 39, or a Ct value greater than 40). In some embodiments, the 3' terminal sequence "ACA" is added to the first group of the subset primers and the 3' terminal sequence "CAC" is added to the second group of primers in order to provide primer sets with closely matched Tm values.

A primer matrix is then generated that includes only the primer pairs with the desired low background level (e.g., all primer pairs generated a Ct value of greater than 35 in a no template control assay), as described in Examples 6 and 7.

In some embodiments, the PCR detection primer pairs are dispensed into a plurality of individual wells 826 (also referred to as "features") in the multi-well container such that each pair of PCR detection primers in each well 826 of the matrix is positionally addressable, i.e., is localized to a known, defined well 826 in the container 800 such that the identity (i.e., the sequence) of each amplified ligation product can be determined from its position on the matrix.

In some embodiments, at least 20% (e.g., at least 25%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or all of the wells in the container) of the wells 826 in the multi-well container 800 comprise a composition 844 comprising a pair of PCR detection primer pairs that is different from the PCR detection primer pairs contained in the other wells of the multi-well container 800.

In some embodiments, the composition 844 further comprises reagents for carrying out an enzyme reaction, such as a polymerase, such as a DNA polymerase, or such as a reverse transcriptase.

In some embodiments, the composition 844 further comprises one or more reagents for carrying out a PCR amplification reaction. PCR amplification methods are well known in the art and are described, for example, in Innis et al., eds., 1990, *PCR Protocols: A Guide to Methods and Applications*, Academic Press Inc., San Diego, CA, and Ausubel et al., *Short Protocols in Molecular Biology*, Wiley, 1995; and Innis et al., *PCR Protocols*, Academic Press, 1990. An amplification reaction typically includes the DNA that is to be amplified, a thermostable DNA

polymerase, two oligonucleotide primers, deoxynucleotide triphosphates (dNTPs), reaction buffer and magnesium.

In some embodiments, the composition 844 comprises a pair of PCR detection primers, DNA polymerase, and reagents for carrying out a quantitative PCR reaction, such as one or more of the following: a Tris buffer, a potassium salt (e.g., potassium chloride), a magnesium salt (e.g., magnesium chloride), nucleotides (e.g., adenine, cytosine, guanine and thymidine), or derivatives thereof, and a detection reagent, such as a fluorescent dye (e.g., SYBR green) or other qPCR reagents known in the art, such as TaqMan, or molecular beacons. In some embodiments, the composition 844 comprises 2X SYBR master mix, commercially available from Applied Biosystems, Foster City, CA.

In some embodiments, the method of making a matrix for decoding the results of a multiplexed assay further comprises aliquoting the liquid composition 844 into multiple wells 826 of the multi-well container 800 and freezing the liquid composition 844 or freezing and drying (i.e., lyophilizing) the composition 844, wherein each dried aliquot comprises an amount of water that is less than 0.1% by weight of the dried aliquot. Aliquots of the liquid composition 844 can be frozen by any means, such as by placing the container containing the aliquots of the liquid composition 844 into a freezer where the container is incubated at a temperature below the freezing point of the liquid mixture until the aliquots of the mixture freeze.

In some embodiments, the method further comprises storing the frozen liquid or lyophilized aliquots at a temperature below minus 15°C. In some embodiments, the method comprises packaging the multi-well container 800 comprising the aliquoted composition 844 into a packaging material, such as a plastic wrapper, or other suitable protective outer packaging material.

Kits for Ligation-Dependent Genotyping Assays

In another aspect, the invention provides a kit for genotyping a test sample at one or more polymorphic loci of interest, such as at one or more single nucleotide variant(s) (SNVs) position(s) of interest. The kit in accordance with this aspect of the invention comprises at least one set of query oligonucleotides for genotyping at least one polymorphic locus of interest, the set comprising (i) at least one 5' ligation oligonucleotide comprising, from the 5' to 3' end, a first PCR primer binding region, a target-specific binding region selected to hybridize 5' of the polymorphic locus of interest, and a 3' region chosen to hybridize to either a consensus or variant nucleotide sequence at the polymorphic locus of interest, and (ii) a phosphorylated 3' ligation oligonucleotide

comprising from the 5' to 3' end, a target-specific binding region selected to hybridize 3' of the polymorphic locus of interest and a second PCR primer binding region. The query ligation oligonucleotides (e.g., SNV query ligation oligonucleotides) may be generated as described herein.

In some embodiments, the kit may further comprise a thermostable DNA ligase, such as Taq DNA ligase or 9°N DNA ligase. In some embodiments, the kit may further comprise at least one synthetic template comprising the target region of interest having a consensus or variant nucleotide at the SNV position of interest. The synthetic templates may be generated as described herein.

In some embodiments, the kit may further comprise one or more detection primer pairs for quantitative PCR analysis of the ligation mixture. In some embodiments, the kit may comprise a multi-well container comprising a plurality of detection primer pairs arranged in a matrix (i.e., a universal plate for decoding a multiplex assay), as described herein. In some embodiments, the kit may further comprise one or more reagents for carrying out a quantitative PCR reaction, such as one or more of the following: a Tris buffer, a potassium salt (e.g., potassium chloride), a magnesium salt (e.g., magnesium chloride), nucleotides (e.g., adenine, cytosine, guanine and thymidine), or derivatives thereof, and a detection reagent, such as a fluorescent dye (e.g., SYBR green) or other qPCR reagents known in the art, such as TaqMan, or molecular beacons.

Oligonucleotide Synthesis

DNA synthesis of the various oligonucleotides of the invention (e.g., SNV query oligos, synthetic templates, PCR detection primer linkers, and capture probes) can be carried out by any art-recognized chemistry, including phosphodiester, phosphotriester, phosphate triester, or N-phosphonate and phosphoramidite chemistries (see, e.g., Froehler et al., *Nucleic Acid Res. 14*:5399-5407, 1986; McBride et al., *Tetrahedron Lett. 24*:246-248, 1983). Methods of oligonucleotide synthesis are well known in the art and generally involve coupling an activated phosphorous derivative on the 3' hydroxyl group of a nucleotide with the 5' hydroxyl group of the nucleic acid molecule (see, e.g., Gait, *Oligonucleotide Synthesis: A Practical Approach*, IRL Press, 1984).

Suitable nucleotides useful in the synthesis of the various oligonucleotides of the invention include nucleotides that contain activated phosphorus-containing groups such as phosphodiester, phosphotriester, phosphate triester, H-phosphonate and phosphoramidite groups. In some embodiments, oligonucleotides can be synthesized using modified nucleotides, or nucleotide

derivatives, such as, for example, combinations of modified phosphodiester linkages such as phosphorothiate, phosphorodithioate, and methylphosphonate, as well as nucleotides having modified bases such as inosine, 5'-nitroindole, and 3' nitropyrrole. Additionally, it is possible to vary the charge on the phosphate backbone of the nucleic acid molecule, for example, by thiolation or methylation, or to use a peptide rather than a phosphate backbone. In some embodiments, oligonucleotides may be synthesized for use in the methods described herein that include one or more nucleotide analogs at one or more positions, wherein the nucleotide analogs enhance oligonucleotide binding affinity, such as 2-O-ethyl modified nucleotides or locked nucleic acid molecules. As used herein, the term "locked nucleic acid molecule" (abbreviated as LNA molecule) refers to a nucleic acid molecule that includes a 2'-O,4'-C-methylene-β-D-ribofuranosyl moiety. Exemplary 2'-O,4'-C-methylene-β-D-ribofuranosyl moieties, and exemplary LNAs including such moieties, are described, for example, in Petersen, M. and Wengel, J., *Trends in Biotechnology 21*(2):74-81 (2003) which publication is incorporated herein by reference in its entirety. The making of such modifications is within the skill of one trained in the art.

A population of nucleic acid molecules can be synthesized on a substrate by any art-recognized means including, for example, photolithography (see, Lipshutz et al., *Nat. Genet. 21*(1 Suppl):20-24, 1999) and piezoelectric printing (see, Blanchard et al., *Biosensors and Bioelectronics 11*:687-690, 1996). In some embodiments, nucleic acid molecules are synthesized in a defined pattern on a solid substrate to form a high-density microarray. Techniques are known for producing arrays containing thousands of oligonucleotides comprising defined sequences at defined locations on a substrate (see, e.g., Pease et al., *Proc. Nat'l. Acad. Sci. 91*:5022-5026, 1994; Lockhart et al., *Nature Biotechnol. 14*:1675-80, 1996; and Lipshutz et al., *Nat. Genet. 21* (1 Suppl):20-4, 1999).

In some embodiments, populations of nucleic acid molecules are synthesized on a substrate, to form a high density microarray, by means of an ink jet printing device for oligonucleotide synthesis, such as described by Blanchard in U.S. Patent No. 6,028,189; Blanchard et al., *Biosensors and Bioelectrics 11*:687-690 (1996); Blanchard, *Synthetic DNA Arrays in Genetic Engineering*, Vol. 20, J.K. Setlow, Ed. Plenum Press, New York at pages 111-123; and U.S. Patent No. 6028189 issued to Blanchard. The nucleic acid sequences in such microarrays are typically synthesized in arrays, for example, on a glass slide, by serially depositing individual nucleotide bases in "microdroplets" of a high surface tension solvent such as propylene carbonate. The microdroplets have small volumes (e.g., 100 picoliters (pL) or less, or

50 pL or less) and are separated from each other on the microarray (e.g., by hydrophobic domains) to form surface tension wells which define the areas containing the array elements (i.e., the different populations of nucleic acid molecules). Microarrays manufactured by this ink-jet method are typically of high density, typically having a density of at least about 2,000 different nucleic acid molecules per 1 cm^2. The nucleic acid molecules may be covalently attached directly to the substrate, or to a linker attached to the substrate at either the 3' or 5' end of the polynucleotide. Exemplary chain lengths of the synthesized nucleic acid molecules suitable for use in the present methods are in the range of about 20 to about 200 nucleotides in length, such as 50 to 100, 60 to 100, 70 to 100, 80 to 100, or 90 to 100 nucleotides in length. In some embodiments, the nucleic acid molecules are in the range of 40 to 100 nucleotides in length.

Exemplary ink jet printing devices suitable for oligonucleotide synthesis in the practice of the present invention contain microfabricated ink-jet pumps, or nozzles, which are used to deliver specified volumes of synthesis reagents to an array of surface tension wells (see, Kyser et al., *J. Appl. Photographic Eng. 7*:73-79, 1981).

In some embodiments, a population of nucleic acid molecules is synthesized to form a high-density microarray. A DNA microarray, or chip, is an array of nucleic acid molecules, such as synthetic oligonucleotides, disposed in a defined pattern onto defined areas of a solid support (see, Schena, *BioEssays 18*:427, 1996). The arrays are preferably reproducible, allowing multiple copies of a given array to be produced and easily compared with each other. Microarrays are typically made from materials that are stable under nucleic acid molecule hybridization conditions. In some embodiments, the nucleic acid molecules on the array are single-stranded DNA sequences. Exemplary microarrays and methods for their manufacture and use are set forth in T.R. Hughes et al., *Nature Biotechnology 19*:342-347, April 2001, which publication is incorporated herein by reference.

In some embodiments, the methods of the invention utilize oligonucleotides that are synthesized on a multiplex parallel DNA synthesis system based on an integrated microfluidic microarray platform for parallel production of oligonucleotides, wherein the DNA synthesis system utilizes photogenerated acid chemistry, parallel microfluidics and a programmable digital light controlled synthesizer, as described in U.S. Patent Publication No. 2007/0059692; Gao et al., *Biopolymers 73*:579-596 (2004); and Zhou et al., *Nucleic Acids Research 32*(18):5409-5417 (2004), each of which is incorporated herein by reference.

In some embodiments, the methods of the invention utilize synthesized oligonucleotides that are cleaved off a substrate, such as a microarray. The synthesized nucleic acid molecules can be harvested from the substrate by any useful means. In some embodiments, the portion of the nucleic acid molecule that is directly attached to the substrate, or attached to a linker that is attached to the substrate, is attached to the substrate or linker by an ester bond which is susceptible to hydrolysis by exposure to a hydrolyzing agent, such as hydroxide ions, for example, an aqueous solution of sodium hydroxide or ammonium hydroxide. The entire substrate can be treated with a hydrolyzing agent, or alternatively, a hydrolyzing agent can be applied to a portion of the substrate. For example, a silane linker can be cleaved by exposure of the silica surface to ammonium hydroxide, yielding various silicate salts and releasing the nucleic acid molecules with the silane linker into solution. In some embodiments, ammonium hydroxide can be applied to the portion of a substrate that is covalently attached to the nucleic acid molecules, thereby releasing the nucleic acid molecules into the solution (see, Scott and McLean, *Innovations and Perspectives in Solid Phase Synthesis*, 3rd International Symposium, 1994, Mayflower Worldwide, pp. 115-124).

The following examples merely illustrate the best mode now contemplated for practicing the invention, but should not be construed to limit the invention.

EXAMPLE 1

This Example describes a method for validating single nucleotide variants (SNV) using oligonucleotide ligation and detection of the ligation product by PCR to confirm the presence of a panel of potential SNVs identified during massively parallel sequencing analysis.

Rationale: One of the unforeseen issues that has emerged with SNV and/or mutation detection in the context of massively parallel sequencing platforms is that allele calls are often ambiguous. A combination of factors including sequence read depth, sequence quality, misaligned reads, alignment algorithms, etc., all likely contribute to the error rate associated with high throughput sequence analysis. None of the current methods for single nucleotide variant (SNV) validation are simple, economical, and orthogonal solutions that are suitable to validate thousands of potential SNVs. Therefore, it is important to have a follow-on validation assay that unambiguously detects polymorphisms in a high-throughput manner.

This Example describes a high throughput assay for SNV detection for genotyping genomic DNA samples in which ligation primers are annealed directly to a genomic DNA

template in the presence of DNA ligase, followed by a real-time PCR assay for the ligation product. The oligonucleotide ligation occurs when query 5' and 3' ligation oligonucleotides bind with perfect complementarity to adjacent sites on target DNA, leaving a gap that can be sealed by DNA ligase. The joining of upstream (5') and downstream (3') query ligation oligonucleotides creates a ligation product that serves as a PCR template. A single nucleotide mismatch at the site of the gap significantly impairs ligation efficiency, and therefore decreases the amount of ligation product (i.e., PCR template) that is created. The amount of ligation product generated, which is read out by quantitative PCR, is indicative of the genotype of the target DNA. As further described and demonstrated in Example 3, because only minute quantities of query ligation oligonucleotides are added to the ligation reaction and each annealing event is independent of other annealing events, hundreds of SNV validation assays can be multiplexed in a single reaction vessel for any given test sample.

Methods:

In this Example, three target synthetic templates containing an SNV of interest were designed and investigated in a total of 18 different assays. The first two synthetic template sets were based on actual human SNPs, and the third synthetic template set was based on an actual mouse SNP. Each double-stranded synthetic SNP template was 61 bp in length, which was generated by synthesizing complementary oligos, in which the SNP base (polymorphic site) was located precisely in the center of the synthetic template (i.e., 30 bp on either side of the SNP). All genotypes described herein are oriented to the forward strand, (e.g., A/G) with the first nucleotide (e.g., "A") listed as the SNV position of interest.

Template set #1 (hSNP1:A/G) contains a synthetic template corresponding to a wild-type (consensus) human allele (SEQ ID NO:1/SEQ ID NO:2) (A/T), and a synthetic template corresponding to a variant human allele (SEQ ID NO:3/SEQ ID NO:4) (G/C), for use as control templates in an assay to distinguish between the presence or absence of the human SNP1 (A/G).

Template set #2 (hSNP2:G/T) contains a synthetic template corresponding to a wild-type human allele (SEQ ID NO:5/SEQ ID NO:6) (G/C), and a synthetic template corresponding to a variant human allele (SEQ ID NO:7/SEQ ID NO:8) (T/A), for use as control templates in an assay to distinguish between the presence or absence of the human SNP (G/T).

Template set #3 (mSNP: A/G) contains a synthetic template corresponding to a wild-type mouse allele (SEQ ID NO:9/SEQ ID NO:10) (A/T), and a synthetic template corresponding to a

variant mouse allele (SEQ ID NO:11/SEQ ID NO:12) (G/C), for use as control templates in an assay to distinguish between the presence or absence of the mouse SNP (A/G).

TABLE 1: OLIGONUCLEOTIDES FOR SYNTHETIC TEMPLATES:

Ref No.	Tem-plate Set	Sequence	SEQ ID NO:
rs949895FA	1 (FA)	5'CTTCTGGCAATTGAAGAAAAAAAATTGAGCAGCTGTAACTGCATGCACATTATGCAAATTT3'	1
rs949895RT	1 (RT)	5'AAATTTGCATAATGTGCATGCAGTTACAGCTGCTCAATTTTTTTTCTTCAATTGCCAGAAG3'	2
rs949895FG	1 (FG)	5'CTTCTGGCAATTGAAGAAAAAAAATTGAGCGGCTGTAACTGCATGCACATTATGCAAATTT3'	3
rs949895RC	1 (RC)	5'AAATTTGCATAATGTGCATGCAGTTACAGCCGCTCAATTTTTTTTCTTCAATTGCCAGAAG3'	4
rs11042937FG	2 (FG)	5'TGCAGCACAAGGGCTGGCACACAGCAGGCCGCCATATTCATGTGCTGTTCTGCCAGACGTT3'	5
rs11042937RC	2 (RC)	5'AACGTCTGGCAGAACAGCACATGAATATGGCGGCCTGCTGTGTGCCAGCCCTTGTGCTGCA3'	6
rs11042937FT	2 (FT)	5'TGCAGCACAAGGGCTGGCACACAGCAGGCCTCCATATTCATGTGCTGTTCTGCCAGACGTT3'	7
rs11042937RA	2 (RA)	5'AACGTCTGGCAGAACAGCACATGAATATGGAGGCCTGCTGTGTGCCAGCCCTTGTGCTGCA3'	8
FTA	3 (FA)	5'GGAGGCCTCGGTGAAGGGCATGCTGGGACGACTCACTAGCACATTGGGTGGCTCAGCTTCC3'	9
RTT	3 (RT)	5'GGAAGCTGAGCCACCCAATGTGCTAGTGAGTCGTCCCAGCATGCCCTTCACCGAGGCCTCC3'	10
FTG	3 (FG)	5'GGAGGCCTCGGTGAAGGGCATGCTGGGACGGCTCACTAGCACATTGGGTGGCTCAGCTTCC3'	11
RTC	3 (RC)	5'GGAAGCTGAGCCACCCAATGTGCTAGTGAGCCGTCCCAGCATGCCCTTCACCGAGGCCTCC3'	12

Pooling the templates

Because each template shown in TABLE 1 has a length of 61 bp, and a molecular weight (MW) of ~40,000 amu, therefore, a 250 nM solution is 10 ng/µl. Complementary

oligonucleotides at a concentration of 10 μM were mixed in buffer containing TEzero (10 mM Tris pH 7.6, 0.1 mM EDTA) plus 20 mM NaCl, diluted to 250 nM, and then diluted 10-fold in TEzero plus 20 mM NaCl that contained 10 ng/μl of human genomic DNA (hgDNA obtained from Clontech). Because hgDNA (diploid) is $6x10^9$ bases, and the templates are $6x10^1$ bases, the template was present in 100,000,000-fold excess. The templates were then diluted in buffered hgDNA 10^6-fold to produce a solution that had 100-fold excess of template over hgDNA. Controls were set up that contained no template (TEzero plus 20 mM NaCl) and the 10 ng/μl hgDNA diluted in TEzero plus 20 mM NaCl.

Ligation Oligonucleotides:

Each assay described in this Example was carried out with two different 5' allele-specific ligation oligos 300, 400, and one common, phosphorylated 3' ligation oligo 500 (e.g., as illustrated in FIGURE 2). For this Example, three sets of 5' ligation oligos were designed for each synthetic template, wherein each set had a different primer binding tail sequence 302, in order to determine whether different tail sequences influence the ability to detect a ligation product by PCR amplification. The 5' ligation oligonucleotides for the human synthetic templates (template sets 1 and 2) were each designed to have 30 nt of complementarity to the target template, and the 5' ligation oligos for the mouse synthetic templates (template set 3) were designed to have 25 nt of complementarity to the target template. The sequences for the ligation oligos are provided below in TABLE 2.

For each 5' ligation allele-specific oligo (SEQ ID NO:13-30) the tail sequence 302 containing the PCR primer binding site is underlined, and the 3' allele-specific region 306 is shown as underlined in bold. For each 3' common phosphorylated [P] ligation oligo (SEQ ID NO:31-33), the tail sequence 502 containing the PCR binding site is underlined.

TABLE 2: 5' AND 3' LIGATION OLIGONUCLEOTIDES

Ref. number	Template Target	Sequence	SEQ ID NO:
5' ligation oligos			
rs949895_FP1_5'LPA	Template Set 1: consensus	5' AGTATAGCCCCAGCGTGTCTACGAGCTTCT GGCAATTGAAGAAAAAAAATTGAGCA 3'	13
rs949895_FP1_5'LPG	Template Set 1: variant	5' AGTATAGCCCCAGCGTGTCTACGAGCTTCT GGCAATTGAAGAAAAAAAATTGAGCG 3'	14

Ref. number	Template Target	Sequence	SEQ ID NO:
5' ligation oligos			
rs949895_FP2_5'LPA	Template Set 1: consensus	5' AATCGCTACTGTCGCAAGGGGTCCTCTTCT GGCAATTGAAGAAAAAAAATTGAGCA 3'	15
rs949895_FP2_5'LPG	Template Set 1: variant	5' AATCGCTACTGTCGCAAGGGGTCCTCTTCT GGCAATTGAAGAAAAAAAATTGAGCG 3'	16
rs949895_FP4_5'LPA	Template Set 1: consensus	5' GGCTGTAGTCATACCATAGTGCATCCTTCT GGCAATTGAAGAAAAAAAATTGAGCA 3'	17
rs949895_FP4_5'LPG	Template Set 1: variant	5' GGCTGTAGTCATACCATAGTGCATCCTTCT GGCAATTGAAGAAAAAAAATTGAGCG 3'	18
rs11042937_FP1_5'LPA	Template Set 2: consensus	5' AGTATAGCCCCAGCGTGTCTACGAGTGCA GCACAAGGGCTGGCACACAGCAGGCCG 3'	19
rs11042937_FP1_5'LPG	Template Set 2: variant	5' AGTATAGCCCCAGCGTGTCTACGAGTGCA GCACAAGGGCTGGCACACAGCAGGCCT 3'	20
rs11042937_FP2_5'LPA	Template Set 2: consensus	5' AATCGCTACTGTCGCAAGGGGTCCTTGCAG CACAAGGGCTGGCACACAGCAGGCCG 3'	21
rs11042937_FP2_5'LPG	Template Set 2: variant	5' AATCGCTACTGTCGCAAGGGGTCCTTGCAG CACAAGGGCTGGCACACAGCAGGCCT 3'	22
rs11042937_FP4_5'LPA	Template Set 2: consensus	5' GGCTGTAGTCATACCATAGTGCATCTGCAG CACAAGGGCTGGCACACAGCAGGCCG 3'	23
rs11042937_FP4_5'LPG	Template Set 2: variant	5' GGCTGTAGTCATACCATAGTGCATCTGCAG CACAAGGGCTGGCACACAGCAGGCCT 3'	24
mouse_1PEA	Template Set 3: consensus	5' AGTATAGCCCCAGCGTGTCTACGAGCTCGG TGAAGGGCATGCTGGGACGA 3'	25
mouse_1PEG	Template Set 3: variant	5' AGTATAGCCCCAGCGTGTCTACGAGCTCGG TGAAGGGCATGCTGGGACGG 3'	26
mouse_2PEA	Template Set 3: consensus	5' AATCGCTACTGTCGCAAGGGGTCCTCTCGG TGAAGGGCATGCTGGGACGA 3'	27
mouse_2PEG	Template Set 3: variant	5' AATCGCTACTGTCGCAAGGGGTCCTCTCGG TGAAGGGCATGCTGGGACGG 3'	28
mouse_4PEA	Template Set 3: consensus	5' GGCTGTAGTCATACCATAGTGCATCCTCGG TGAAGGGCATGCTGGGACGA 3'	29
mouse_4PEG	Template Set 3: variant	5' GGCTGTAGTCATACCATAGTGCATCCTCGG TGAAGGGCATGCTGGGACGG 3'	30

Ref. number	Template Target	Sequence	SEQ ID NO:
5' ligation oligos			
3' ligation oligos			
rs949895_FP3_3'LP	Template set 1	5' [P]GCTGTAACTGCATGCACATTATGCAAAT TTTTCCAGCTATCCTGTAAGGCAACGT 3'	31
rs11042937_FP3_3'LP	Template set 2	5' [P]CCATATTCATGTGCTGTTCTGCCAGACG TTTTCCAGCTATCCTGTAAGGCAACGT 3'	32
mouse_SNP_FP3_3'LP	Template set 3	5' [P]CTCACTAGCACATTGGGTGGCTCAGCTT CCTTCCAGCTATCCTGTAAGGCAACGT 3'	33

Annealing and Ligation Reaction

The annealing/ligation reactions were carried out under very dilute conditions (5 fmol). A 5 nM stock solution was prepared for each ligation primer in water.

Two different thermo-stable ligases were tested in this Example: Taq DNA ligase, and 9°N DNA Ligase (both obtained from New England Biolabs).

Ligation reactions:

1 µl 10X ligase buffer (New England Biolabs)

5 µl H_2O

0.1 µl of 5 M NaCl (to increase salt for annealing)

2 µl of annealed template (25 fM)

1 µl of 5' ligation oligo (5 nM) (allele specific)

1 µl of 3' ligation primer (5 nM) (phosphorylated common primer)

0.2 µl (40 U/µl) ligase (Taq DNA ligase or 9°N DNA Ligase, New England Biolabs)

A ligation cocktail was prepared that contained the following: each ligase enzyme type (Taq DNA ligase or 9°N DNA Ligase), water, salt, buffer and the common 3' phosphorylated ligation oligo. 7 µl of the ligation cocktail was aliquoted into wells of a 96 well plate. 2 µl of annealed template and 1 µl of 5' allele-specific ligation oligo was then added.

The following assays were carried out:

Set #1 (hSNP1:A/G)

Templates tested: SEQ ID NO:1/SEQ ID NO:2 (A) and SEQ ID NO:3/SEQ ID NO:4 (G)

5' allele-specific ligation oligos tested: SEQ ID NO:13-18

3' common ligation oligo tested: SEQ ID NO:31

Set #2 (hSNP2:G/T)

Templates tested: SEQ ID NO:5/SEQ ID NO:6 (G) and SEQ ID NO:7/SEQ ID NO:8 (T)

5' allele-specific ligation oligos tested: SEQ ID NOS:19-24

3' common ligation oligo tested: SEQ ID NO:32

Set #3 (mSNP:A/G)

Templates tested: SEQ ID NO:9/SEQ ID NO:10 (A) and SEQ ID NO:11/SEQ ID NO:12 (G).

5' allele-specific ligation oligos tested: SEQ ID NOS:25-30

3' common ligation oligo tested: SEQ ID NO:33

The ligation reactions were aliquoted into a grid pattern in a 96-well assay plate as shown below in TABLE 3 and incubated in a thermal cycler across the following temperatures:

95°C for 5 minutes;

75°C for 15 minutes;

70°C for 15 minutes;

65°C for 15 minutes;

60°C for 15 minutes;

55°C for 15 minutes;

50°C for 15 minutes;

45°C for 15 minutes;

4°C rest;

The ligation reactions were then diluted to 100 µl with 90 µl of TEzero (10 mM Tris pH 7.6, 0.1 mM EDTA), and quantitative PCR assays were carried out as described below.

TABLE 3: QPCR DETECTION SCHEME FOR LIGATED TEMPLATES

Sample Series #	**qPCR Primers**	**Template Set 1 (hSNP1-A)**	**Template Set 1 (hSNP1-G)**	**Template Set 2 (hSNP2-G)**	**Template Set 2 (hSNP2-T)**	**Template Set 3 (mSNP-A)**	**Template Set 3 (mSNP-G)**
1	FP1+RP3	SEQ ID NO:13	SEQ ID NO:13	SEQ ID NO:19	SEQ ID NO:19	SEQ ID NO:25	SEQ ID NO:25
2	FP1+RP3	SEQ ID NO:14	SEQ ID NO:14	SEQ ID NO:20	SEQ ID NO:20	SEQ ID NO:26	SEQ ID NO:26
3	FP2+RP3	SEQ ID	SEQ ID	SEQ ID	SEQ ID	SEQ ID	SEQ ID

Sample Series #	qPCR Primers	Template Set 1 (hSNP1-A)	Template Set 1 (hSNP1-G)	Template Set 2 (hSNP2-G)	Template Set 2 (hSNP2-T)	Template Set 3 (mSNP-A)	Template Set 3 (mSNP-G)
		NO:15	NO:15	NO:21	NO:21	NO:27	NO:27
4	FP2+RP3	SEQ ID NO:16	SEQ ID NO:16	SEQ ID NO:22	SEQ ID NO:22	SEQ ID NO:28	SEQ ID NO:28
5	FP4+RP3	SEQ ID NO:17	SEQ ID NO:17	SEQ ID NO:23	SEQ ID NO:23	SEQ ID NO:29	SEQ ID NO:29
6	FP4+RP3	SEQ ID NO:18	SEQ ID NO:18	SEQ ID NO:24	SEQ ID NO:24	SEQ ID NO:30	SEQ ID NO:30

Quantitative PCR Assays

PCR primers

Three forward PCR primers were designed to hybridize to the tail regions of the 5' allele-specific ligation oligos (300, 400) containing three different PCR Primer Binding sites (302, 402) as follows:

FP1 5' AGTATAGCCCCAGCGTGTCTACGAG3' (SEQ ID NO:34)

FP2 5' AATCGCTACTGTCGCAAGGGGTCCT3' (SEQ ID NO:35)

FP4 5' GGCTGTAGTCATACCATAGTGCATC3' (SEQ ID NO:36)

One reverse PCR Primer was designed to hybridize to the tail region of the 3' common ligation oligo (500):

RP3 5' ACGTTGCCTTACAGGATAGCTGGAA3' (SEQ ID NO:37)

Power SYBR master mix (Applied Biosystems): a premix of all the components (SYBR Green Dye, AmpliTaq Gold® DNA Polymerase, dNTPs, and buffer components) except primers, template, and water, necessary to perform real-time PCR. The SYBR Green dye, which binds to double-stranded DNA, provides a fluorescent signal that reflects the amount of dsDNA product generated during PCR. The master mix includes AmpliTaq Gold® DNA Polymerase, provided in an inactive state to allow pre-mixing of PCR reagents at room temperate and allows for an automated, hot start. Upon thermal activation, the enzyme is activated.

qPCR Assay:

2 µl of each diluted ligation reaction was used as a template in four 10 µl qPCR reactions as follows:

A PCR reaction cocktail was prepared so that each sample would contain:

5 µl of 2X master mix (Power SYBR master mix, manufactured by Applied Biosystems)

1.4 µl H_2O

0.8 µl of Forward PCR primer (10 µM)

0.8 µl of Reverse PCR primer (10 µM)

8 µl total volume was aliquoted into wells of the 96 well plate, and 2 µl of diluted ligation template was added.

qPCR was run for 40 cycles with a denaturation step at the end to assess product integrity.

The qPCR abundance ratios for correct over mismatched calls for all 36 assays that were run is shown below in TABLE 4.

TABLE 4: QPCR RATIOS OF TARGET/OFF-TARGET FOR LIGATION DEPENDENT GENOTYPING ASSAYS

qPCR primer	**ligation template**	**9°N data**	**Taq data**
FP1	hSNP1-A	19	77
FP1	hSNP1-G	93	17
FP2	hSNP1-A	8*	126
FP2	hSNP1-G	71	20
FP4	hSNP1-A	10	111
FP4	hSNP1-G	59	20
FP1	hSNP2-G	211	311
FP1	hSNP2-T	195	34
FP2	hSNP2-G	53	309
FP2	hSNP2-T	312	95
FP4	hSNP2-G	489	288
FP4	hSNP2-T	217	0*
FP1	mSNP-A	29	154
FP1	mSNP-G	50	42
FP2	mSNP-A	41	72
FP2	mSNP-G	62	29
FP4	mSNP-A	58	101
FP4	mSNP-G	31	20

Results:

The data shown above in TABLE 4 is a measure of the specificity of the SNV detection assay, with all but two assay sets (shown with *) registering ratios of correct versus mismatch > 10. The magnitude of this differential detection is an adequate foundation for a genotyping assay because the ten-fold and greater difference in absolute abundances between correctly matched assays and mismatched assays translates into a correct-allele-to-incorrect allele Ct difference in qPCR measurements of > 3, which is a threshold value well above random deviations that are observed within an experiment.

These data demonstrate that all three of the target synthetic SNV templates were accurately detected with the total of 18 different assays using at least one of the two tested thermostable DNA ligase enzymes. DNA ligase and, in particular, the thermostable Taq DNA ligase, are ideal enzymes for interrogating nucleotide polymorphisms because they can only seal nicks at sites of perfect base pairing. The thermostable nature of the DNA ligase is advantageous because the enzyme activity is retained at the high temperatures needed for DNA melting and reannealing. It is noted that the ligation oligos worked at very dilute concentrations (5 fmol), and all the tested arbitrary PCR binding tails all appeared to work; therefore, the multiplexing aspect of the assays is likely to be successfully implemented.

This Example demonstrates the successful use of a ligation-dependent assay to detect SNVs in synthetic templates mixed with genomic DNA. However, the allele calls were not perfect in this experiment, likely due to the fact that the mismatch generated significant background, which is typical of many genotyping assays. In order to improve the accuracy of this detection assay, each set of ligation oligos was calibrated against a control set of synthetic reference and variant templates, as described in Example 3.

EXAMPLE 2

This Example describes the manufacture of a 96-well assay plate comprising a 12 column by 8 row primer matrix of detection primer pairs (also referred to as a "universal PCR decoding matrix"), which can be pre-made and stored in a freezer, for decoding a multiplex assay, such as a multiplex ligation-dependent genotyping assay for genotyping a test sample at a plurality of SNV positions of interest.

PCR Primer Matrix Design

As described in Example 1, the 5' and 3' ligation oligos (300, 400, and 500) for each genotyping assay are tailed with unique PCR primer binding regions (302, 402, 502) that correspond to a pair of PCR detection primers that are present in a particular well (also referred to as an "address") in the universal assay matrix. Therefore, each address (for example, a well in a 96-well plate) "decodes" the result from an individual genotyping assay.

An important element of the universal PCR decoding matrix is that the last two or three (penultimate) 3' bases of the PCR primers are chosen to reduce and preferably eliminate primer-dimer formation, and the remaining bases are specificity tags chosen to provide a unique address at an intersection position (well) in the matrix, disposed into one or more assay plates.

A matrix comprising 20 "universal" paired decoding PCR primers (provided in TABLE 5) was produced for use in a universal detection assay carried out on a 96-well plate 800 (e.g., as shown in FIGURES 4-6), as follows.

Each of the 12 column "C" PCR primers were aliquoted into a separate well 826 along the horizontal axis of the 96 well assay plate (columns 1-12).

Each of the 8 row "R" PCR primers were aliquoted into a separate well 826 along the vertical axis of the 96 well assay plate (rows A-H).

As shown below in TABLE 6, each well 826 located at the intersection of a row and column of the 96 well assay plate contained a unique PCR primer pair, thereby providing a unique "address" at a designated physical location on the matrix (i.e., a positionally addressable array). The universal PCR plate containing the 96 unique pairs of PCR primers was then used to "decode" the results of a multiplexed ligation-dependent genotyping assay. The allele-specific ligation oligonucleotides in the genotyping assay were designed with tail sequences that are complementary to the PCR primers at a specific well location in the assay plate.

The PCR primer design for the universal PCR decoding plate:

It was previously determined that almost any two 25mer oligonucleotides having DNA sequences with a balanced A, C, G, and T content can serve as quantitative PCR primer pairs, provided that they terminate in a di- or tri- nucleotide sequence that inhibits primer-dimer formation (data not shown). In this Example, each PCR primer was 25 nucleotides in length, with the 23 bases at the 5' end of the primer 602, 702 serving as specificity "addresses," due to the fact that each well of the matrix contained a unique pair of primers which would bind to and amplify the ligation product resulting from an individual genotyping assay.

As shown in FIGURE 3C, each forward PCR primer 600 has a 5' region 602 that binds to a primer binding region in the 5' tail of a 5' ligation oligo, and a region 606 at the 3' end having a sequence selected to inhibit primer-dimer interactions with the reverse PCR primer. For example, as shown in FIGURE 2B, forward primer 600 has a region 602 that binds to primer binding region 302 in the 5' tail of the 5' ligation oligo 300, and forward primer 600' has a region 602 that binds to primer region 402 in the 5' tail of the 5' ligation oligo 400. Similarly, each reverse PCR primer 700 has a 5' region 702 that binds to a primer binding region in the 3' tail of a 3' ligation oligo, and a region 706 at the 3' end having a sequence selected to inhibit primer-dimer interactions with the forward PCR primer.

The "C" series was designed as the reverse primer set 700 to bind to the 3' common tail region 502 on the ligation products 200, 250.

The "R" series was designed as the forward primer set 600, each forward primer having a region 606 designed to specifically bind to the 5' tail region 302, 402 on the ligation products 200, 250.

To alleviate primer dimer formation, each "R" PCR primer sequence ended in "CT" and each "C" PCR primer sequence ended in "GA." These terminal dinucleotides cannot pair with one another nor can they self anneal, hence they prevent the formation of primer dimers. It will be understood by those of skill in the art that other di- or tri-nucleotide sequences could be chosen to avoid primer-dimer formation. Exemplary tri-nucleotide sequences chosen to avoid the formation of primer-dimers are provided in Example 4 herein.

The PCR primers were synthesized by MWG/Operon, Huntsville, Alabama, resuspended in water to a concentration of 100 μM and a 1 ml of a 10 μM working stock was made for each primer.

The sequences of the 20 universal PCR primers used to generate the universal PCR decoding matrix are provided below in TABLE 5.

TABLE 5: UNIVERSAL DECODING PCR PRIMERS

Reference number	**Sequence**	**SEQ ID NO:**
C1	5' ACTCTGCGCTCTGGAACTTACCG**GA** 3'	38
C2	5'GATCTTGGACGAAGTCGTCCTAT**GA** 3'	39

Reference number	Sequence	SEQ ID NO:
C3	5'GTTTGCATGAAGACCTCTATACA**GA** 3'	40
C4	5'CTGAAAGGTCCATGGCCTGTACT**GA** 3'	41
C5	5'TTGTATCGATGCAGCCAGGATCC**GA** 3'	42
C6	5'CACAGAATTAGCGATCTATGCCG**GA** 3'	43
C7	5'CACGCCTCATCGTAGTGTAGGAG**GA** 3'	44
C8	5'ATAACATTGAACGCTGCCGTTGC**GA** 3'	45
C9	5'CAGACTACGGCAATATAACGCTG**GA** 3'	46
C10	5'ACGTAACTATTACGGTGAGCGCC**GA** 3'	47
C11	5'ATGGATAGCCGCTGTTTAACTAC**GA** 3'	48
C12	5'TTCGGCTTCCACAGAGCAAGGTA**GA** 3'	49
R1	5'TGTATCAGCATCTGGCTCAGCGT**CT** 3'	50
R2	5'CTTTGGGGTAAGCGACCATCAGC**CT** 3'	51
R3	5'TACATAGAATCTACCGTGGTGAC**CT** 3'	52
R4	5'ACGATGGCGTTGCAGGCGCTTAC**CT** 3'	53
R5	5'TTGACTGAGACTCCTCATGACCT**CT** 3'	54
R6	5'GCCGTTTCATATCGAACAAGGCG**CT** 3'	55
R7	5'GGGCTACTCGCAATTTCAAATTG**CT** 3'	56
R8	5'CGCCAGCAATCAGCTTTGATACA**CT** 3'	57

Design of Universal Assay Matrix

The layout of the universal assay matrix for qPCR to detect ligation products in a multiplexed ligation-dependent genotyping assay for multiple SNV positions of interest, was a matrix of wells (i.e., features), the matrix comprising a plurality of columns and rows. For example, with reference to FIGURE 6A, wells A1 and B1 represent the qPCR assay result

detecting the ligation products resulting from ligation of a 5' consensus ligation oligo (at A1) or a 5' variant ligation oligo (at B1) with the 3' ligation oligo in an assay for a determining the nucleotide present at a particular SNV position of interest.

For example, as shown in more detail in FIGURE 4, the tail region of the 5' consensus allele-specific ligation oligo 300 for SNV (Gene#1) had a first primer binding sequence (e.g., for binding to PCR primer R1 = SEQ ID NO:50), and the tail region of the 5' variant allele-specific ligation oligo 400 for SNV (Gene#1) had a second primer binding sequence (e.g., for binding to PCR primer R2 = SEQ ID NO:51). The 3' phosphorylated common primer 500, common to both assays for SNV (Gene#1), had a common primer binding sequence (e.g., for binding to PCR primer C1 = SEQ ID NO:38). As further shown in FIGURE 4, assay #1 is decoded in the universal assay matrix dispensed into a 96-well plate at well A1 (containing PCR primers R1+C1) for measuring the amount of consensus ligation product 200, and at well B1 (containing PCR primers R2 and C1), for measuring the amount of variant ligation product 250 present in the multiplexed ligation mixture. Thus, a 96 well assay plate can accommodate a universal matrix for carrying out consensus versus variant assays for 48 SNV positions of interest.

As shown in FIGURE 6, a set of three identical 96-well assay plates 800, each comprising a universal matrix can be used to generate a set of data representing: a pool of synthetic consensus templates (FIGURE 6A); a pool of synthetic variant templates (FIGURE 6B); and a test sample (FIGURE 6C), each assayed with the same pool of SNV query consensus and variant allele-specific ligation oligos for a particular SNV having PCR tails corresponding to a particular location on the assay plate.

Preparation of the Assay Plate(s) Containing the Universal Matrix

The assay plates were prepared for quantitative PCR (qPCR) assays as follows:

35 mls of 2X Power SYBR master mix (Applied Biosystems, Foster City, CA) was combined with 10 mls of H_2O. 450 µl of the mixture was aliquoted into each well of a 96 well assay plate. 55 µl of the 12 "C" (reverse) primers (10 µM) were aliquoted into the wells of the columns (C) of the assay plate, and 55 µl of the 8 "R" (forward) primers (10 µM) were aliquoted into the wells of the rows (R) of the 96 well assay plate, as shown below in TABLE 6. The assay plate can be run in a 96 well plate format. Alternatively, for an assay done in quadruplicate, the reagents were mixed, then 8 µl per well was aliquoted in quadruplicate into a 384 qPCR plate, in order to carry out 4 identical reactions for each qPCR primer pair, as described in Example 3.

For example, as described in Example 3, 2 µl aliquots were dispensed into all wells of the prepared 384 well qPCR plate (4 x 96). The samples were mixed, and the qPCR assay was run on an ABI 7900 instrument set on SYBR detection channel.

TABLE 6: 96-WELL ASSAY PLATE CONTAINING A UNIVERSAL PCR PRIMER MATRIX

	C1 (SEQ ID NO:38)	**C2 (SEQ ID NO:39)**	**C3 (SEQ ID NO:40)**	**C4 (SEQ ID NO:41)**	**C5 (SEQ ID NO:42)**	**C6 (SEQ ID NO:43)**	**C7 (SEQ ID NO:44)**	**C8 (SEQ ID NO:45)**	**C9 (SEQ ID NO:46)**	**C10 (SEQ ID NO:47)**	**C11 (SEQ ID NO:48)**	**C12 (SEQ ID NO:49)**
A: R1 (SEQ ID NO:50)	R1 x C1	R1 x C2	R1 x C3	R1 x C4	R1 x C5	R1 x C6	R1 x C7	R1 x C8	R1 x C9	R1 x C10	R1 x C11	R1 x C12
B: R2 (SEQ ID NO:51)	R2 x C1	R2 x C2	R2 x C3	R2 x C4	R2 x C5	R2 x C6	R2 x C7	R2 x C8	R2 x C9	R2 x C10	R2 x C11	R2 x C12
C: R3 (SEQ ID NO:52)	R3 x C1	R3 x C2	R3 x C3	R3 x C4	R3 x C5	R3 x C6	R3 x C7	R3 x C8	R3 x C9	R3 x C10	R3 x C11	R3 x C12
D: R4 (SEQ ID NO:53)	R4 x C1	R4 x C2	R4 x C3	R4 x C4	R4 x C5	R4 x C6	R4 x C7	R4 x C8	R4 x C9	R4 x C10	R4 x C11	R4 x C12
E: R5 (SEQ ID NO:54)	R5 x C1	R5 x C2	R5 x C3	R5 x C4	R5 x C5	R5 x C6	R5 x C7	R5 x C8	R5 x C9	R5 x C10	R5 x C11	R5 x C12
F: R6 (SEQ ID NO:55)	R6 x C1	R6 x C2	R6 x C3	R6 x C4	R6 x C5	R6 x C6	R6 x C7	R6 x C8	R6 x C9	R6 x C10	R6 x C11	R6 x C12
G: R7 (SEQ ID NO:56)	R7 x C1	R7 x C2	R7 x C3	R7 x C4	R7 x C5	R7 x C6	R7 x C7	R7 x C8	R7 x C9	R7 x C10	R7 x C11	R7 x C12
H: R8 (SEQ ID NO:57)	R8 x C1	R8 x C2	R8 x C3	R8 x C4	R8 x C5	R8 x C6	R8 x C7	R8 x C8	R8 x C9	R8 x C10	R8 x C11	R8 x C12

Results:

In order to validate the universal PCR decoding matrix, an initial experiment was carried out in which TEzero (no template control) was added to a universal PCR decoding matrix, prepared as described above, and a qPCR assay was carried out with the 8 ("R") forward PCR

primers (SEQ ID NOS:50-57) and the 12 ("C") reverse PCR primers (SEQ ID NOS:38-49). The qPCR data was analyzed at the level of raw Cts. In this initial experiment, only three wells in column C10 gave background Ct values less than 30 (lower Ct values represent higher amounts of product), and this background did not significantly impact assay performance. Therefore, the concept of a universal matrix of PCR primers was validated by this initial experiment. It will be understood by those of skill in the art that the design of the universal PCR matrix described herein, and the design principles of the PCR primers can be expanded to accommodate approximately 1000 or more samples, which can be assayed on an appropriately sized multi-well assay plate. For example, a primer matrix with > 1000 addresses can be constructed from as few as 32 row and 36 column primers (32 x 36 = 1152 unique addresses, otherwise referred to as "unique features").

EXAMPLE 3

This Example describes a multiplexed, high throughput assay for SNV genotyping using oligonucleotide ligation and detection of the ligation product by PCR to validate the presence or absence of a panel of 96 potential SNVs that were initially detected during high-throughput sequence analysis.

Rationale:

In order to further develop the ligation-dependent SNV detection assay described in Example 1 for high-throughput analysis, an experiment was set up to genotype 96 potential SNVs that were initially identified during massively-parallel sequencing of a genomic DNA library representing 139 cancer-related genes from a Calu6 cell line. This assay, referred to as "oligonucleotide ligation validation of potential SNVs" or "OLIVES" combines single-tube multiplexing with assay read-out in a universal PCR decoding plate (described in Example 2) to provide both validated genotypes and assay reagents for follow-on genotyping studies.

In the first step of the assay, 5' allele-specific oligos and 3' phosphorylated common ligation oligonucleotides (up to 1000 or more) are annealed to the test DNA and ligated. In the second step, the ligation mixture is distributed across universal PCR "decoding" plates, as described in Example 2, which can be pre-made and stored in a freezer prior to use.

Methods:

1. Prepare DNA Samples for Genotyping:

Genotyping signal-to-noise ratios typically improve when DNA samples are enriched in target sequences. Therefore, in this Example a comparison was made between genotyping total genomic DNA and genotyping a genomic DNA library enriched for target sequences.

A. Total genomic DNA 85 ng/µl was isolated from the Calu6 cell line from cells grown in culture using a standard genomic DNA purification kit (Qiagen, Valencia, CA).

B. A genomic DNA library was generated from a panel of 139 cancer-related genes from the Calu6 cell line and was enriched using solution-based capture as follows.

Preparation of Capture Probes

All the exons of the set of 139 genes were identified. An algorithm was then applied for picking alternating sense and antisense strand chimeric oligos with a 5' target-specific region (35 nt) with a sequence that hybridizes to either the sense or antisense strand of each of these exons, and a 3' region that hybridizes to the biotinylated adaptor capture oligo.

These capture oligonucleotides were chosen as follows. For exons less than 69 nucleotides in length, two oligonucleotides, both targeting the same strand and oriented in the same direction, and not overlapping one another in sequence by more than 10 nucleotides were chosen. In some cases where exons were very short (i.e., <60 nucleotides), these capture oligonucleotides included flanking exon sequences.

For exons between 70 and 115 nucleotides in length, two oligonucleotides targeting opposite Watson and Crick strands and oriented in the opposite orientations were selected. The first oligonucleotide covered exon base positions 1-35 and the second oligonucleotide was positioned from base positions 80-115, which often included flanking intron sequences, so that the oligos were each about 35 nt in length, and spaced about 45 nt apart.

For exonic sequences greater than 115 nucleotides in length, the first capture oligonucleotide was placed at exon positions 1-35 and successive oligos were placed in alternating orientations with a spacing of 45 nucleotides between oligonucleotides.

The oligos designed as described above were synthesized by Operon and provided in a plate at 100 μM and pooled into a single 50 ml sample using a Biomek robot. The pooled 3229 capture oligos were then diluted to 10 μM and 1 μM.

TaqMan assays were developed for the 139 target genes. TaqMan assays were also developed for off-target genes for use as negative controls. These genes were not targeted by capture oligonucleotides, and it was shown that their representation diminished during the course of target library enrichment.

Library Generation:

Genomic DNA libraries were generated by fragmenting Calu6 genomic DNA and ligating on linkers containing a first and second primer binding site, followed by PCR amplification for 20 cycles with PCR forward primer and PCR reverse primer, then the PCR product was purified over a Qiaquick column.

Solution-Based Capture and Enrichment of Libraries for Target Sequences

Capture reagents: 10 μM of the capture oligos for the 139 candidate genes described were mixed with 10 μM of the biotinylated adaptor oligo.

Capture Mixture: 125 μl of 2X binding buffer (2 M NaCl, 20 mM Tris pH 7.6, 0.2 mM EDTA), 60 μl (4.3 μg) of gDNA library, 5 μl capture oligo pool (50 pM of 10 μM of oligo pool + adaptor oligo), and 60 μl water, for a total volume of 250 μl.

The reaction mixture was annealed as follows:

94°C for 1 minute
90°C for 1 minute
85°C for 1 minute
80°C for 1 minute
75°C for 1 minute
70°C for 1 minute
65°C for 1 minute
60°C for 1 minute
55°C for 1 minute
50°C for 1 minute
45°C for 1 minute
40°C for 1 minute
25°C - hold

Capture Reagents: Washed beads were prepared by combining six aliquots of 50 µl beads (in principle, each 50 µl of beads is capable of binding 50 pmol of dsDNA complex), 500 µl 2X binding buffer and 440 µl water. The beads were pulled over with a magnet and washed twice with 1 ml 1X binding buffer.

1st Round of Capture/Enrichment: The aliquots of washed oligos were combined with the annealed oligos into a total volume of 1 ml of 1X binding buffer and mixed gently for 15 minutes.

Wash solutions:

A series of wash buffers with increasing formamide were tested, each with 100 mM Tris pH 7.6, 1 mM EDTA, and a range of formamide from 15%, 20%, 25%, 30%, and 50%.

It was previously determined that the presence of 20 mM NaCl in the 10 mM Tris pH 7.6, 1 mM EDTA buffer enhanced non-specific binding (data not shown), therefore the NaCl was eliminated in the wash buffer in this experiment.

The capture oligos/library/bead complexes were washed four times with the above-described wash buffers including formamide, 1 ml each wash for 5 minutes.

Elution: The DNA bound to the beads was eluted with two aliquots of 50 µl of water by incubation at 94°C for 1 minute each, pulling over the beads and removing the eluate, for a total eluate volume of 100 µl.

Amplification of eluate (once enriched library):

PCR Reaction Mixture (5% DMSO)

29 µl	H_2O
20 µl	5X buffer (supplied by manufacturer with the EXPAND$^{plus®}$ kit, Roche)
10 µl	25 mM $MgCl_2$
10 µl	template (1/10th eluate from once enriched fragment library)
5 µl	dNTPs (10 nM each dNTP)
5 µl	DMSO
10 µl	10 µM Forward PCR primer
10 µl	10 µM Reverse PCR primer
1 µl	Expand$^{PLUS®}$ polymerase (Roche)
100 µl total volume	

PCR cycling Conditions

1 cycle:

94°C for 2 minutes

10 cycles:

94°C for 30 sec

60°C for 30 sec

72°C for 1 minute

10 or 15 cycles:

94°C for 30 sec

60°C for 30 sec

72°C for 1 minute plus 10 sec/cycle

1 cycle:

72°C for 7 minutes

4°C hold

The PCR reaction products were purified over a Qiaquick column and quantified.

1 µl of PCR product was analyzed on a 2% agarose gel.

2. Design of SNV Query Oligos for Ligation-Dependent Genotyping Assay

A set of SNV query oligos were designed to determine the presence or absence of a panel of potential SNVs that had been identified during the sequencing of 139 genes from the Calu6 cell line using massively-parallel sequencing techniques (data not shown). From this initial sequencing analysis, 96 non-synonymous SNV calls were identified, whose confidence was ranked from high to low based on the degree of overlapping bioinformatic evidence. Assays 1 to 96 listed in TABLE 11 correspond to the 96 distinct putative SNVs that were initially detected as potential polymorphic loci during massively parallel sequencing. The lowest numbered assays in TABLE 11 (starting at assay #1), correspond to the highest confidence ranking, based on the degree of overlapping bioinformatic evidence (e.g., the presence of the SNV in dbSNP). As shown in TABLE 11, many known SNPs from the dbSNP database and two known mutations identified in the Wellcome Trust COSMIC database were included in the set of 96 non-synonymous SNV calls.

For each of the 96 SNV positions of interest, the following reagents were generated:

- A 5' allele-specific consensus ligation oligo (51mer) (TABLE 7)
- A 5' allele-specific variant ligation oligo (51mer) (TABLE 8)

- A 3' common ligation oligo (50mer) (TABLE 9)
- A pair of oligos to generate an annealed reference template with the consensus SNV sequence (51mers) (not shown)
- A pair of oligos to generate an annealed reference template with the variant SNV sequence (51mers) (not shown)

Ligation Oligonucleotides

The paired set of allele-specific (consensus and variant) 5' and 3' ligation oligonucleotide pairs for each SNV target of interest were designed as follows:

Each 5' ligation oligo 300, 400 had a total length of 51 nucleotides, with a target-specific complementary region 304, 404 of 25 nucleotides, an allele-specific region 306, 406 of 1 nucleotide, and a primer-binding tail region 302, 402 of 25 nucleotides in length, including 2 nt at the 3' end corresponding to the forward PCR primer region 606 selected to avoid primer dimer (e.g., "CT").

The target-specific binding region 304, 404 of the 5' ligation oligos was designed to have a length of 25 nt that were 100% complementary to the target region of interest immediately 5' of each of the panel of 96 SNV loci of interest. The allele-specific binding region 306, 406 of the 5' ligation oligos were designed to have a length of 1 nt that was complementary to the consensus or variant allele for a particular SNV of interest. The sequence of the tail region 302, 402 was selected to bind to a forward PCR primer 600 in the universal assay plate 800 made as described in Example 2. The 5' consensus ligation oligos for SNV assays 1-96 are provided in TABLE 7. The 5' variant ligation oligos for SNV assays 1-96 are provided in TABLE 8.

Each 3' ligation oligo 500 had a total length of 50 nucleotides, with a target-specific complementary region 504 of 25 nucleotides, and a primer-binding tail region 502 of 25 nucleotides in length. The target-specific complementary region of the 3' ligation oligos was designed to have a length of 25 nt that was 100% complementary to the target region of interest starting at the nucleotide immediately 3' to the SNV position of interest. The sequence of the tail region 502 was selected to bind to a reverse PCR primer 700 in the universal assay plate 800 made as described in Example 2. The 3' ligation oligos were phosphorylated prior to use in the assay.

The 3' ligation sequences for SNV assays 1-96 are provided below in TABLE 9.

Ligation Primers:

6 plates of 96 well plates were ordered for synthesis as follows:

Plate 1: 5' consensus and variant ligation oligos for 1-48

Plate 2: 3' common ligation oligos for 1-48

Plate 3: consensus templates for 1-96

Plate 4: 5' consensus and variant ligation oligos for 49-96

Plate 5: 3' common ligation oligos for 49-96

Plate 6: variant templates for 1-96

TABLE 7: CONSENSUS 5' LIGATION OLIGOS

SNV Assay #	SEQUENCE (5' to 3')	SEQ ID NO:
1	TGTATCAGCATCTGGCTCAGCGTCTGCTTTCATTCATATCTGCAGGTTCAA	58
2	TACATAGAATCTACCGTGGTGACCTGCAGTCCAGAGCACCGTGGTCCTGCT	59
3	TTGACTGAGACTCCTCATGACCTCTCACCACTGGGGTAAGGTTTTCTAGGG	60
4	GGGCTACTCGCAATTTCAAATTGCTAGCCAGTTTTCCATGGGTTCTACTAC	61
5	TGTATCAGCATCTGGCTCAGCGTCTCTTCCCGGTCAGCTACTCCTCTTCCG	62
6	TACATAGAATCTACCGTGGTGACCTTGATCCATTAGATTCAAATGTAGCAA	63
7	TTGACTGAGACTCCTCATGACCTCTACCTGCTGGTGCCACTCTGGAAAGGC	64
8	GGGCTACTCGCAATTTCAAATTGCTCTTGCTGCTTCCAGTAAATAAGGTGA	65
9	TGTATCAGCATCTGGCTCAGCGTCTATCCTTGTCCAAGGAGGCTGTTTCTG	66
10	TACATAGAATCTACCGTGGTGACCTTCCACACGCAAATTTCCTTCCACTCG	67
11	TTGACTGAGACTCCTCATGACCTCTAGGAGCTGCTGGTGCAGGGGCCACGG	68
12	GGGCTACTCGCAATTTCAAATTGCTATTCATCGGACATGTTACTGTTTTTC	69
13	TGTATCAGCATCTGGCTCAGCGTCTGTGTCATCAACTTGGTCCACAGTCGT	70
14	TACATAGAATCTACCGTGGTGACCTGGCCCTCTAGGGACTCGAACAGAGAT	71
15	TTGACTGAGACTCCTCATGACCTCTCAGAGGGAGGACGAGCTGACCTTCAT	72

SNV Assay #	SEQUENCE (5' to 3')	SEQ ID NO:
16	GGGCTACTCGCAATTTCAAATTGCTCCAACTCGAAATTCCCCGTGACCAGA	73
17	TGTATCAGCATCTGGCTCAGCGTCTACCAGCAGATACTCAGCCGGAGGATA	74
18	TACATAGAATCTACCGTGGTGACCTGAGGACCCCAAGTCCCATAGGGACCC	75
19	TTGACTGAGACTCCTCATGACCTCTGCAGCGCACCACGGGACCCAAGCCCG	76
20	GGGCTACTCGCAATTTCAAATTGCTCAGAGTCTGAGGTAGCTGCCCTGGCA	77
21	TGTATCAGCATCTGGCTCAGCGTCTTGCTGTTTTCTTCCTTCAGGCATACA	78
22	TACATAGAATCTACCGTGGTGACCTCCTGAACAGCTCGCGGCTCAGCAGGG	79
23	TTGACTGAGACTCCTCATGACCTCTATCTTCAAAGTTGCAGTAAAAACCCA	80
24	GGGCTACTCGCAATTTCAAATTGCTCTTGATTCATGATATTTTACTCCAAG	81
25	TGTATCAGCATCTGGCTCAGCGTCTTCCCTCATTGCACTGTACTCCTCTTG	82
26	TACATAGAATCTACCGTGGTGACCTGTCCGAGGACAACGATGAGGCGGCGC	83
27	TTGACTGAGACTCCTCATGACCTCTTTCCGCCTGGTGTTGGAAGAGACAGG	84
28	GGGCTACTCGCAATTTCAAATTGCTTGGTCTTTCAGTGCCTCCACTATGAC	85
29	TGTATCAGCATCTGGCTCAGCGTCTTGAAGAGAAATATAAGAAGGCTATGG	86
30	TACATAGAATCTACCGTGGTGACCTCCTCCAGGTGCAGGAGTTCATGCTCA	87
31	TTGACTGAGACTCCTCATGACCTCTAAAGGCAATGTGGGATCCTGAATTGC	88
32	GGGCTACTCGCAATTTCAAATTGCTGCCCGAACAGCCGCTGGATATGGGAC	89
33	TGTATCAGCATCTGGCTCAGCGTCTCTAAAAAGGACCCTGAAGGTTGTGAC	90
34	TACATAGAATCTACCGTGGTGACCTTCATATGGATGATAATGATGGAGAAC	91
35	TTGACTGAGACTCCTCATGACCTCTCTTGGCTGTGCTCCTGCTGCTGGCCG	92
36	GGGCTACTCGCAATTTCAAATTGCTATCAACTATAGGTTGCTTTGGTGGTG	93
37	TGTATCAGCATCTGGCTCAGCGTCTAAATTTCTGAATAACTGAAGTTGGTC	94
38	TACATAGAATCTACCGTGGTGACCTGGTAGCAGACAAACCTGTGGTTGATC	95
39	TTGACTGAGACTCCTCATGACCTCTGAGCTTTGGGTTGTTCCTTAGGACCC	96

SNV Assay #	SEQUENCE (5' to 3')	SEQ ID NO:
40	GGGCTACTCGCAATTTCAAATTGCTGGAATTACGGCAGCCCTTCTTTCCCA	97
41	TGTATCAGCATCTGGCTCAGCGTCTCGGGGCCTCTGCTTGGATGTGATGAC	98
42	TACATAGAATCTACCGTGGTGACCTGGCGGCCGTGGTGGCGGCAGTGGTGG	99
43	TTGACTGAGACTCCTCATGACCTCTGCAGAAGTCATATTTAGGATGTGTAC	100
44	GGGCTACTCGCAATTTCAAATTGCTGGACTTTTTTTCCAAGGCTATTCAGT	101
45	TGTATCAGCATCTGGCTCAGCGTCTGAGATGTGTAAGCGCAGCCTTGAGTC	102
46	TACATAGAATCTACCGTGGTGACCTATTTATGCTATACATGATGAAACATC	103
47	TTGACTGAGACTCCTCATGACCTCTGAAATTGATAGAAGCAGAAGATCGGC	104
48	GGGCTACTCGCAATTTCAAATTGCTCTCACCTCCCATGTTGCTCAAAGAAC	105
49	TGTATCAGCATCTGGCTCAGCGTCTAGCATTCTCTGCAGTACATCAACCGT	106
50	TACATAGAATCTACCGTGGTGACCTACTTTACTCACGTTTTTCCCATCTAG	107
51	TTGACTGAGACTCCTCATGACCTCTGGCATGGTGGTGGATGTAGTGGTGGT	108
52	GGGCTACTCGCAATTTCAAATTGCTATGGTGGTGGATGTAGTGGTGGTGGA	109
53	TGTATCAGCATCTGGCTCAGCGTCTAACATGAGTTTTTTATGGCGGGAGGT	110
54	TACATAGAATCTACCGTGGTGACCTGGACACCGGCAAGGCCACCCTGACCT	111
55	TTGACTGAGACTCCTCATGACCTCTCTGACCCACTCATCCCAAGACACACC	112
56	GGGCTACTCGCAATTTCAAATTGCTGAAAGTAACAGCTTGACTATATCCAC	113
57	TGTATCAGCATCTGGCTCAGCGTCTGTGTCCTGGAATGGGGCCCATGAGAT	114
58	TACATAGAATCTACCGTGGTGACCTTGGAATTTCCTCCTCGAGTCTGAACC	115
59	TTGACTGAGACTCCTCATGACCTCTTTCCCTCCAGCCCCAGGTTACCCCTG	116
60	GGGCTACTCGCAATTTCAAATTGCTCGGGGGGTCTTGGATGTGCCGGCTTG	117
61	TGTATCAGCATCTGGCTCAGCGTCTTGGCCTGTTGGCCGTATCTGCTAACA	118
62	TACATAGAATCTACCGTGGTGACCTCTGTTTTGTTCCGAATGTCTGAGGAC	119
63	TTGACTGAGACTCCTCATGACCTCTGTGTCAACAATTCTAAGGAGGAAGAT	120

SNV Assay #	SEQUENCE (5' to 3')	SEQ ID NO:
64	GGGCTACTCGCAATTTCAAATTGCTGAGATCCAGATGTTTTGGAATATTAC	121
65	TGTATCAGCATCTGGCTCAGCGTCTTGGACTGTGTATGAAACCTGGTTTTA	122
66	TACATAGAATCTACCGTGGTGACCTCAATCTTTTTAACCATTTTGTCATCG	123
67	TTGACTGAGACTCCTCATGACCTCTCCAGGGGAGAAAAGTACATTGGAAAC	124
68	GGGCTACTCGCAATTTCAAATTGCTCTTTATTCAGGTGGATGCCCCTGACC	125
69	TGTATCAGCATCTGGCTCAGCGTCTCAACATCTCTTTTCCCTGGAAGTTTC	126
70	TACATAGAATCTACCGTGGTGACCTGGACATGGATCTTGTTTTTCTCTTTG	127
71	TTGACTGAGACTCCTCATGACCTCTCTCAATCTGTAGTGCTCCTGGTCGGC	128
72	GGGCTACTCGCAATTTCAAATTGCTGTTTTTTCAGGAGGCCATCTTTCTCC	129
73	TGTATCAGCATCTGGCTCAGCGTCTAGAATGAGCCTGTTCTGTTGACATTG	130
74	TACATAGAATCTACCGTGGTGACCTCAGGGGAGGGTGTGGGCAGGCGGTTC	131
75	TTGACTGAGACTCCTCATGACCTCTGACTATTCAGACATCAATGAGGTGGC	132
76	GGGCTACTCGCAATTTCAAATTGCTCTGTTCCTCTACAGGGCCAAAACACT	133
77	TGTATCAGCATCTGGCTCAGCGTCTGATGATACTCACTGTCCATCAGCCTC	134
78	TACATAGAATCTACCGTGGTGACCTAGTATCCTCACCTGTAGCCAGGTATC	135
79	TTGACTGAGACTCCTCATGACCTCTGTTAATTCAGCATCCAGCAGGTCCCT	136
80	GGGCTACTCGCAATTTCAAATTGCTACACCAACATTCCCAGCTGCTGGAAC	137
81	TGTATCAGCATCTGGCTCAGCGTCTCCTGGGCCAGGTGTGCATCAAAGCGC	138
82	TACATAGAATCTACCGTGGTGACCTGTGTCAAGCTACTCTCAGGACTGCTC	139
83	TTGACTGAGACTCCTCATGACCTCTCAAGGTGCCAGGTGCAAGACCCACCA	140
84	GGGCTACTCGCAATTTCAAATTGCTTGTCGCGATGAATGTGAAATCCTGGA	141
85	TGTATCAGCATCTGGCTCAGCGTCTTCCACAAACTCGTCACTCATCCTCCG	142
86	TACATAGAATCTACCGTGGTGACCTAGACATGGAAGCCAGTGATTATGAGC	143
87	TTGACTGAGACTCCTCATGACCTCTCCACAGCCAGGCAGTCTGTATCTTGC	144

SNV Assay #	SEQUENCE (5' to 3')	SEQ ID NO:
88	GGGCTACTCGCAATTTCAAATTGCTATATGTGGAGGCCCAACAAAAGAGAC	145
89	TGTATCAGCATCTGGCTCAGCGTCTTGGAAGTTGCGTATTGTAAGCTATTC	146
90	TACATAGAATCTACCGTGGTGACCTGGTCAGAACAGGAGTGCACGGATAGC	147
91	TTGACTGAGACTCCTCATGACCTCTTGCTTTCAATCCCAAATTATGTGTTT	148
92	GGGCTACTCGCAATTTCAAATTGCTCATCAGTGTGTCTGAACATGTGGTCC	149
93	TGTATCAGCATCTGGCTCAGCGTCTCTGCCAGCCTGCCCTGGAGGAAGACA	150
94	TACATAGAATCTACCGTGGTGACCTACTGGAACTATCTGTAATACTGGAAC	151
95	TTGACTGAGACTCCTCATGACCTCTAACTCTTTCACTTTTACATATTAAAG	152
96	GGGCTACTCGCAATTTCAAATTGCTGCAGCCAGAGTGGTTTTTTCAGGGGA	153

TABLE 8: VARIANT 5' LIGATION OLIGOS

SNV Assay #	SEQUENCE (5' to 3')	SEQ ID NO:
1	CTTTGGGGTAAGCGACCATCAGCCTGCTTTCATTCATATCTGCAGGTTCAG	154
2	ACGATGGCGTTGCAGGCGCTTACCTGCAGTCCAGAGCACCGTGGTCCTGCC	155
3	GCCGTTTCATATCGAACAAGGCGCTCACCACTGGGGTAAGGTTTTCTAGGA	156
4	CGCCAGCAATCAGCTTTGATACACTAGCCAGTTTTCCATGGGTTCTACTAT	157
5	CTTTGGGGTAAGCGACCATCAGCCTCTTCCCGGTCAGCTACTCCTCTTCCA	158
6	ACGATGGCGTTGCAGGCGCTTACCTTGATCCATTAGATTCAAATGTAGCAC	159
7	GCCGTTTCATATCGAACAAGGCGCTACCTGCTGGTGCCACTCTGGAAAGGG	160
8	CGCCAGCAATCAGCTTTGATACACTCTTGCTGCTTCCAGTAAATAAGGTGG	161
9	CTTTGGGGTAAGCGACCATCAGCCTATCCTTGTCCAAGGAGGCTGTTTCTA	162
10	ACGATGGCGTTGCAGGCGCTTACCTTCCACACGCAAATTTCCTTCCACTCA	163
11	GCCGTTTCATATCGAACAAGGCGCTAGGAGCTGCTGGTGCAGGGGCCACGC	164

SNV Assay #	SEQUENCE (5' to 3')	SEQ ID NO:
12	CGCCAGCAATCAGCTTTGATACACTATTCATCGGACATGTTACTGTTTTTG	165
13	CTTTGGGGTAAGCGACCATCAGCCTGTGTCATCAACTTGGTCCACAGTCGG	166
14	ACGATGGCGTTGCAGGCGCTTACCTGGCCCTCTAGGGACTCGAACAGAGAC	167
15	GCCGTTTCATATCGAACAAGGCGCTCAGAGGGAGGACGAGCTGACCTTCAC	168
16	CGCCAGCAATCAGCTTTGATACACTCCAACTCGAAATTCCCCGTGACCAGT	169
17	CTTTGGGGTAAGCGACCATCAGCCTACCAGCAGATACTCAGCCGGAGGATG	170
18	ACGATGGCGTTGCAGGCGCTTACCTGAGGACCCCAAGTCCCATAGGGACCT	171
19	GCCGTTTCATATCGAACAAGGCGCTGCAGCGCACCACGGGACCCAAGCCCC	172
20	CGCCAGCAATCAGCTTTGATACACTCAGAGTCTGAGGTAGCTGCCCTGGCG	173
21	CTTTGGGGTAAGCGACCATCAGCCTTGCTGTTTTCTTCCTTCAGGCATACC	174
22	ACGATGGCGTTGCAGGCGCTTACCTCCTGAACAGCTCGCGGCTCAGCAGGA	175
23	GCCGTTTCATATCGAACAAGGCGCTATCTTCAAAGTTGCAGTAAAAACCCG	176
24	CGCCAGCAATCAGCTTTGATACACTCTTGATTCATGATATTTTACTCCAAA	177
25	CTTTGGGGTAAGCGACCATCAGCCTTCCCTCATTGCACTGTACTCCTCTTT	178
26	ACGATGGCGTTGCAGGCGCTTACCTGTCCGAGGACAACGATGAGGCGGCGT	179
27	GCCGTTTCATATCGAACAAGGCGCTTTCCGCCTGGTGTTGGAAGAGACAGA	180
28	CGCCAGCAATCAGCTTTGATACACTTGGTCTTTCAGTGCCTCCACTATGAT	181
29	CTTTGGGGTAAGCGACCATCAGCCTTGAAGAGAAATATAAGAAGGCTATGT	182
30	ACGATGGCGTTGCAGGCGCTTACCTCCTCCAGGTGCAGGAGTTCATGCTCG	183
31	GCCGTTTCATATCGAACAAGGCGCTAAAGGCAATGTGGGATCCTGAATTGA	184
32	CGCCAGCAATCAGCTTTGATACACTGCCCGAACAGCCGCTGGATATGGGAA	185
33	CTTTGGGGTAAGCGACCATCAGCCTCTAAAAAGGACCCTGAAGGTTGTGAA	186
34	ACGATGGCGTTGCAGGCGCTTACCTTCATATGGATGATAATGATGGAGAAA	187
35	GCCGTTTCATATCGAACAAGGCGCTCTTGGCTGTGCTCCTGCTGCTGGCCA	188

SNV Assay #	SEQUENCE (5' to 3')	SEQ ID NO:
36	CGCCAGCAATCAGCTTTGATACACTATCAACTATAGGTTGCTTTGGTGGTA	189
37	CTTTGGGGTAAGCGACCATCAGCCTAAATTTCTGAATAACTGAAGTTGGTT	190
38	ACGATGGCGTTGCAGGCGCTTACCTGGTAGCAGACAAACCTGTGGTTGATA	191
39	GCCGTTTCATATCGAACAAGGCGCTGAGCTTTGGGTTGTTCCTTAGGACCT	192
40	CGCCAGCAATCAGCTTTGATACACTGGAATTACGGCAGCCCTTCTTTCCCC	193
41	CTTTGGGGTAAGCGACCATCAGCCTCGGGGCCTCTGCTTGGATGTGATGAT	194
42	ACGATGGCGTTGCAGGCGCTTACCTGGCGGCCGTGGTGGCGGCAGTGGTGT	195
43	GCCGTTTCATATCGAACAAGGCGCTGCAGAAGTCATATTTAGGATGTGTAA	196
44	CGCCAGCAATCAGCTTTGATACACTGGACTTTTTTTCCAAGGCTATTCAGG	197
45	CTTTGGGGTAAGCGACCATCAGCCTGAGATGTGTAAGCGCAGCCTTGAGTT	198
46	ACGATGGCGTTGCAGGCGCTTACCTATTTATGCTATACATGATGAAACATA	199
47	GCCGTTTCATATCGAACAAGGCGCTGAAATTGATAGAAGCAGAAGATCGGA	200
48	CGCCAGCAATCAGCTTTGATACACTCTCACCTCCCATGTTGCTCAAAGAAA	201
49	CTTTGGGGTAAGCGACCATCAGCCTAGCATTCTCTGCAGTACATCAACCGC	202
50	ACGATGGCGTTGCAGGCGCTTACCTACTTTACTCACGTTTTTCCCATCTAT	203
51	GCCGTTTCATATCGAACAAGGCGCTGGCATGGTGGTGGATGTAGTGGTGGG	204
52	CGCCAGCAATCAGCTTTGATACACTATGGTGGTGGATGTAGTGGTGGTGGG	205
53	CTTTGGGGTAAGCGACCATCAGCCTAACATGAGTTTTTTATGGCGGGAGGG	206
54	ACGATGGCGTTGCAGGCGCTTACCTGGACACCGGCAAGGCCACCCTGACCG	207
55	GCCGTTTCATATCGAACAAGGCGCTCTGACCCACTCATCCCAAGACACACT	208
56	CGCCAGCAATCAGCTTTGATACACTGAAAGTAACAGCTTGACTATATCCAT	209
57	CTTTGGGGTAAGCGACCATCAGCCTGTGTCCTGGAATGGGGCCCATGAGAC	210
58	ACGATGGCGTTGCAGGCGCTTACCTTGGAATTTCCTCCTCGAGTCTGAACA	211
59	GCCGTTTCATATCGAACAAGGCGCTTTCCCTCCAGCCCCAGGTTACCCCTC	212

SNV Assay #	SEQUENCE (5' to 3')	SEQ ID NO:
60	CGCCAGCAATCAGCTTTGATACACTCGGGGGGTCTTGGATGTGCCGGCTTT	213
61	CTTTGGGGTAAGCGACCATCAGCCTTGGCCTGTTGGCCGTATCTGCTAACT	214
62	ACGATGGCGTTGCAGGCGCTTACCTCTGTTTTGTTCCGAATGTCTGAGGAA	215
63	GCCGTTTCATATCGAACAAGGCGCTGTGTCAACAATTCTAAGGAGGAAGAA	216
64	CGCCAGCAATCAGCTTTGATACACTGAGATCCAGATGTTTTGGAATATTAA	217
65	CTTTGGGGTAAGCGACCATCAGCCTTGGACTGTGTATGAAACCTGGTTTTG	218
66	ACGATGGCGTTGCAGGCGCTTACCTCAATCTTTTTAACCATTTTGTCATCT	219
67	GCCGTTTCATATCGAACAAGGCGCTCCAGGGGAGAAAAGTACATTGGAAAA	220
68	CGCCAGCAATCAGCTTTGATACACTCTTTATTCAGGTGGATGCCCCTGACA	221
69	CTTTGGGGTAAGCGACCATCAGCCTCAACATCTCTTTTCCCTGGAAGTTTA	222
70	ACGATGGCGTTGCAGGCGCTTACCTGGACATGGATCTTGTTTTTCTCTTTT	223
71	GCCGTTTCATATCGAACAAGGCGCTCTCAATCTGTAGTGCTCCTGGTCGGA	224
72	CGCCAGCAATCAGCTTTGATACACTGTTTTTTCAGGAGGCCATCTTTCTCA	225
73	CTTTGGGGTAAGCGACCATCAGCCTAGAATGAGCCTGTTCTGTTGACATTT	226
74	ACGATGGCGTTGCAGGCGCTTACCTCAGGGGAGGGTGTGGGCAGGCGGTTA	227
75	GCCGTTTCATATCGAACAAGGCGCTGACTATTCAGACATCAATGAGGTGGA	228
76	CGCCAGCAATCAGCTTTGATACACTCTGTTCCTCTACAGGGCCAAAACACC	229
77	CTTTGGGGTAAGCGACCATCAGCCTGATGATACTCACTGTCCATCAGCCTT	230
78	ACGATGGCGTTGCAGGCGCTTACCTAGTATCCTCACCTGTAGCCAGGTATT	231
79	GCCGTTTCATATCGAACAAGGCGCTGTTAATTCAGCATCCAGCAGGTCCCA	232
80	CGCCAGCAATCAGCTTTGATACACTACACCAACATTCCCAGCTGCTGGAAA	233
81	CTTTGGGGTAAGCGACCATCAGCCTCCTGGGCCAGGTGTGCATCAAAGCGA	234
82	ACGATGGCGTTGCAGGCGCTTACCTGTGTCAAGCTACTCTCAGGACTGCTA	235
83	GCCGTTTCATATCGAACAAGGCGCTCAAGGTGCCAGGTGCAAGACCCACCT	236

SNV Assay #	SEQUENCE (5' to 3')	SEQ ID NO:
84	CGCCAGCAATCAGCTTTGATACACTTGTCGCGATGAATGTGAAATCCTGGG	237
85	CTTTGGGGTAAGCGACCATCAGCCTTCCACAAACTCGTCACTCATCCTCCA	238
86	ACGATGGCGTTGCAGGCGCTTACCTAGACATGGAAGCCAGTGATTATGAGA	239
87	GCCGTTTCATATCGAACAAGGCGCTCCACAGCCAGGCAGTCTGTATCTTGA	240
88	CGCCAGCAATCAGCTTTGATACACTATATGTGGAGGCCCAACAAAAGAGAA	241
89	CTTTGGGGTAAGCGACCATCAGCCTTGGAAGTTGCGTATTGTAAGCTATTA	242
90	ACGATGGCGTTGCAGGCGCTTACCTGGTCAGAACAGGAGTGCACGGATAGA	243
91	GCCGTTTCATATCGAACAAGGCGCTTGCTTTCAATCCCAAATTATGTGTTC	244
92	CGCCAGCAATCAGCTTTGATACACTCATCAGTGTGTCTGAACATGTGGTCT	245
93	CTTTGGGGTAAGCGACCATCAGCCTCTGCCAGCCTGCCCTGGAGGAAGACT	246
94	ACGATGGCGTTGCAGGCGCTTACCTACTGGAACTATCTGTAATACTGGAAA	247
95	GCCGTTTCATATCGAACAAGGCGCTAACTCTTTCACTTTTACATATTAAAT	248
96	CGCCAGCAATCAGCTTTGATACACTGCAGCCAGAGTGGTTTTTTCAGGGGG	249

TABLE 9: 3' PHOSPHORYLATED LIGATION OLIGOS

SNV Assay #	SEQUENCE (5' to 3')	SEQ ID NO:
1	TTTTCACATGGTTTTCCAGGCTTGCTCCGGTAAGTTCCAGAGCGCAGAGT	250
2	GCGCCAGCTCCAGCAAAGCCAGCACTCCGGTAAGTTCCAGAGCGCAGAGT	251
3	TTGGCTTCGACAACTTTGCTGCTTGTCCGGTAAGTTCCAGAGCGCAGAGT	252
4	TAAACTAGAAAACATACAAAATAGGTCCGGTAAGTTCCAGAGCGCAGAGT	253
5	GTGCCCGCCGGCCCTCGCTGGACTCTCATAGGACGACTTCGTCCAAGATC	254
6	ATCAGAAGCCCTTTGAGAGTGGAAGTCATAGGACGACTTCGTCCAAGATC	255
7	CCAAGACTCTCTCCCCAGGGAAGAATCATAGGACGACTTCGTCCAAGATC	256

SNV Assay #	SEQUENCE (5' to 3')	SEQ ID NO:
8	GGTACTGTACTTTAAAGAGGTCACTTCATAGGACGACTTCGTCCAAGATC	257
9	TCTGCAAAGGAGTAAGTCGATTTGGTCTGTATAGAGGTCTTCATGCAAAC	258
10	GATAAGATGCTGAGGAGGGGCCAGATCTGTATAGAGGTCTTCATGCAAAC	259
11	GGGGAGCAGCCTCTGGCATTCTGGGTCTGTATAGAGGTCTTCATGCAAAC	260
12	CTCCCTGATGTACCACCAACTTTACTCTGTATAGAGGTCTTCATGCAAAC	261
13	GTCAGGAGGGGCATCAGGCGCTAAGTCAGTACAGGCCATGGACCTTTCAG	262
14	CTCTGCAGCTGTGGGTTTCTTTGCATCAGTACAGGCCATGGACCTTTCAG	263
15	CAAGAGCGCCATCATCCAGAATGTGTCAGTACAGGCCATGGACCTTTCAG	264
16	CTTTTGGACACCAGGTTGGTGAATCTCAGTACAGGCCATGGACCTTTCAG	265
17	TTTCAGAGGTGAGAGTAGGGCAATTTCGGATCCTGGCTGCATCGATACAA	266
18	CTCGAATAGGCACAGTTACCCCCAGTCGGATCCTGGCTGCATCGATACAA	267
19	CCGAGCTCGCGCCAGCCCGCGCCACTCGGATCCTGGCTGCATCGATACAA	268
20	TATTTAACAACATCAGCCGAGACGTTCGGATCCTGGCTGCATCGATACAA	269
21	GAGGATGACCCCAAAGATAGTGGATTCCGGCATAGATCGCTAATTCTGTG	270
22	CGTCCCAGAGCTGGTCCACCTGCAGTCCGGCATAGATCGCTAATTCTGTG	271
23	CAGGCAGTTTCCCTATGGAGAGAGCTCCGGCATAGATCGCTAATTCTGTG	272
24	ATACAAATGAATCATGGAGAAATCTTCCGGCATAGATCGCTAATTCTGTG	273
25	ACCTGCTGTGTCGAGAATATCCAAGTCCTCCTACACTACGATGAGGCGTG	274
26	CCGGGCCCTGGGCGGTGGCAACGGCTCCTCCTACACTACGATGAGGCGTG	275
27	CATGGGTTTGGTGACCTGGCCCTTGTCCTCCTACACTACGATGAGGCGTG	276
28	GTTGTAGGTGGCACCTCTGGTGAGGTCCTCCTACACTACGATGAGGCGTG	277
29	TTTCCAATGCTCAGCTAGACAATGATCGCAACGGCAGCGTTCAATGTTAT	278
30	GCTTCCTCCGAGACCCCTTACGAGATCGCAACGGCAGCGTTCAATGTTAT	279
31	AAAAACCTTCACAACGACCAGGCCTTCGCAACGGCAGCGTTCAATGTTAT	280

SNV Assay #	SEQUENCE (5' to 3')	SEQ ID NO:
32	GAACAGCCGCAAGTTTGAGTTTGAATCGCAACGGCAGCGTTCAATGTTAT	281
33	AAAAGTGATGACAAAAACACTGTAATCCAGCGTTATATTGCCGTAGTCTG	282
34	TAGATACACCAATAAATTATAGTCTTCCAGCGTTATATTGCCGTAGTCTG	283
35	GGCTGTATCGAGGGCAGGCGCTCCATCCAGCGTTATATTGCCGTAGTCTG	284
36	TTGCCAACACAGCCTCTGCTTCTTCTCCAGCGTTATATTGCCGTAGTCTG	285
37	CTGAATTCTATGAAAAGTAGGTCTTTCGGCGCTCACCGTAATAGTTACGT	286
38	CTAAATTAGTGAAAAGAAAAATGTATCGGCGCTCACCGTAATAGTTACGT	287
39	GGTAGGGGGTGTGCTTATAAGGTAATCGGCGCTCACCGTAATAGTTACGT	288
40	CCCACATGGGGCCCATCAAACTCCGTCGGCGCTCACCGTAATAGTTACGT	289
41	TTGCAAAGACGGTGCTATGGACTGATCGTAGTTAAACAGCGGCTATCCAT	290
42	CGTTGGTGATGTTGGCCCCGCTGGCTCGTAGTTAAACAGCGGCTATCCAT	291
43	TATCTGTATAAATAAGAAAAAAAGGTCGTAGTTAAACAGCGGCTATCCAT	292
44	GTGCGAGGTAATCTAATCTCTTTTTTCGTAGTTAAACAGCGGCTATCCAT	293
45	TGTGTATTCGCTCTATCCCACACTTTCTACCTTGCTCTGTGGAAGCCGAA	294
46	TTATAAAGGAAAAAAAATACCGAAATCTACCTTGCTCTGTGGAAGCCGAA	295
47	TATAAAAAAGATAATGGAAAGGGATTCTACCTTGCTCTGTGGAAGCCGAA	296
48	CATATAGTAAGTATTTAATTTATGCTCTACCTTGCTCTGTGGAAGCCGAA	297
49	GACCTGTCAAAATAGAATGTGAGTTTCCGGTAAGTTCCAGAGCGCAGAGT	298
50	CAATTCCATGCACTTCTCATTTCTGTCCGGTAAGTTCCAGAGCGCAGAGT	299
51	GGACATGCTTCGTCGTCTGCTTGGTTCCGGTAAGTTCCAGAGCGCAGAGT	300
52	CATGCTTCGTCGTCTGCTTGGTCACTCCGGTAAGTTCCAGAGCGCAGAGT	301
53	AGACTGACCCTTTTTGGACTTCAGGTCATAGGACGACTTCGTCCAAGATC	302
54	CGAGCCCACTGGGTGCATCCTGAGATCATAGGACGACTTCGTCCAAGATC	303
55	TGCTACGGAGAAGTTGTTTAAGGGGTCATAGGACGACTTCGTCCAAGATC	304

SNV Assay #	SEQUENCE (5' to 3')	SEQ ID NO:
56	ATGCCCATTCTTGGCTGCATCGTGATCATAGGACGACTTCGTCCAAGATC	305
57	GGTTGTCTGAGAGAGAGCTTCTTGTTCTGTATAGAGGTCTTCATGCAAAC	306
58	AAAACTGCCAGGAACAATACACAACTCTGTATAGAGGTCTTCATGCAAAC	307
59	TCGTGTGGCTCCTTCTTTGCTATAGTCTGTATAGAGGTCTTCATGCAAAC	308
60	TCAGATTGGTGAGCTCCCATCTGTTTCTGTATAGAGGTCTTCATGCAAAC	309
61	GGCTGTGCCACTGCTGGGGAAGGCCTCAGTACAGGCCATGGACCTTTCAG	310
62	AAGCCACAAGATTACAAGAAACGGCTCAGTACAGGCCATGGACCTTTCAG	311
63	CCTCCAAGGATGTACTGCAGTACAGTCAGTACAGGCCATGGACCTTTCAG	312
64	AAAAATGATCATGCCAAGAAGCCTATCAGTACAGGCCATGGACCTTTCAG	313
65	TTACTTTTCTTCTCTTGATGTGCAATCGGATCCTGGCTGCATCGATACAA	314
66	TCACTTTCTAAGAACTTCTTTATGGTCGGATCCTGGCTGCATCGATACAA	315
67	TAAAAAGATAGAATCTGAAAGTAAATCGGATCCTGGCTGCATCGATACAA	316
68	AAAAAGGAACTGAGATAAAACCAGGTCGGATCCTGGCTGCATCGATACAA	317
69	AAACGGCCACTGCAGACTTCACCGATCCGGCATAGATCGCTAATTCTGTG	318
70	ACAACAGTAAAATCACCTATGAGACTCCGGCATAGATCGCTAATTCTGTG	319
71	AAAAGACAACATTCTCTATTTTAGGTCCGGCATAGATCGCTAATTCTGTG	320
72	AACCTGCCATATAAATCTAAGATCTTCCGGCATAGATCGCTAATTCTGTG	321
73	TGCGCACCACATCAATCACTTCCCATCCTCCTACACTACGATGAGGCGTG	322
74	AAGCCGTTGGCTGGAGACACCTATTTCCTCCTACACTACGATGAGGCGTG	323
75	AGAAGATGAAAGCCGAAGATACCAGTCCTCCTACACTACGATGAGGCGTG	324
76	CATTGACTCAGATCTCTCAATCCATTCCTCCTACACTACGATGAGGCGTG	325
77	CAGTTCAGCAAGGGGTCATAGACAATCGCAACGGCAGCGTTCAATGTTAT	326
78	AATCTGGATGGCTTTCACCCCCTCCTCGCAACGGCAGCGTTCAATGTTAT	327
79	GGCCTTGTCAATGCACTAGAAGAGATCGCAACGGCAGCGTTCAATGTTAT	328

SNV Assay #	SEQUENCE (5' to 3')	SEQ ID NO:
80	AAAAACTGAAATGGACAAGAGGTCATCGCAACGGCAGCGTTCAATGTTAT	329
81	TCGTCCAGGGACGCCAAGACACAGTTCCAGCGTTATATTGCCGTAGTCTG	330
82	AACTAAAACAAAACGATGACAAATTTCCAGCGTTATATTGCCGTAGTCTG	331
83	TGAGCAGATAGCCTCCCACCACACGTCCAGCGTTATATTGCCGTAGTCTG	332
84	GAATGTCCTGTGTCAAACAGAGTACTCCAGCGTTATATTGCCGTAGTCTG	333
85	GAGCTCGCGGCCATAGCGCTGTGCTTCGGCGCTCACCGTAATAGTTACGT	334
86	TTGAAGATGAAACAAGACCTGCTAATCGGCGCTCACCGTAATAGTTACGT	335
87	AAAAACATCCACTCTGCCTCGAATCTCGGCGCTCACCGTAATAGTTACGT	336
88	TAGAAGCCTTATTCACTAAAATTCATCGGCGCTCACCGTAATAGTTACGT	337
89	AAAAAAAGAAAAAGATTCAGGTAAGTCGTAGTTAAACAGCGGCTATCCAT	338
90	AAAACAACTAGAAAATGATACAAGATCGTAGTTAAACAGCGGCTATCCAT	339
91	CCGAAATTTACCGCATGGAGGAAGTTCGTAGTTAAACAGCGGCTATCCAT	340
92	GCAGGTAGCGGGACTGTCGGGTGGGTCGTAGTTAAACAGCGGCTATCCAT	341
93	GTACAGCATCACACCCACGCTGAGATCTACCTTGCTCTGTGGAAGCCGAA	342
94	CTAAATAAAACAAAGCAGCCAAAAATCTACCTTGCTCTGTGGAAGCCGAA	343
95	CCTCATGAGGATCACTGGCCAGTAATCTACCTTGCTCTGTGGAAGCCGAA	344
96	GTCTTATATAAGTAATTTAAAAAAATCTACCTTGCTCTGTGGAAGCCGAA	345

Step 1. Pooling of synthesized oligos.

All of the oligos from each of the 6 plates were pooled into separate, labeled pools of 100 µM oligos, resulting in a pool of 96 variant templates, a pool of 96 consensus templates, a pool of consensus plus variant 5' ligation oligos for templates 1-48, a pool of consensus plus variant 5' ligation oligos for templates 49-96, and a pool of 3' common ligation oligos for templates 1-48, and a pool of 3' common ligation oligos for templates 49-96.

100 µl of the Ligation oligos were diluted to a stock solution of 0.5 µM = 5 nM in each ligation oligo.

The pooled templates were diluted to a working concentration of 100 pM.

Step 2. Kinase treatment of the 3' common ligation oligo pools.

The 3' common ligation oligo pools (from plate 2 and plate 5) were kinased as follows:

A 100 µl reaction of 1 µM pooled common oligos = 10 µl of 10 µM 3' oligos = 100 pmoles of termini (optimal molarity of ends in a 100 µl kinase reaction).

The kinase reaction was carried out as follows:

10 µl 10X T4 kinase buffer (New England Biolabs, Ipswich, MA)

10 µl 10 mM ATP

10 µl of 10 µM 3' common ligation oligo pool

70 µl H_2O

100 µl total volume, mix, add 2 µl T4 kinase (New England Biolabs, Ipswich, MA), mix and incubate at 37°C for 30 minutes, then incubate at 65°C for 20 minutes.

The kinase reaction was then diluted by adding 300 µl of H_2O to a 400 µl mixture of 250 nM 3' common ligation primer that was 5 nM in each primer.

Step 3. The Ligation-Dependent Genotyping Assays were carried out as follows:

For each assay, the ligation mixture contains

1. 96 consensus templates with 500 pM ligation oligos (high) (assays 1-48)
2. 96 variant templates with 500 pM ligation oligos (high) (assays 1-48)
3. No template control with 500 pM ligation oligos (high) (assays 1-48)
4. 96 consensus templates with 100 pM ligation oligos (low) (assays 1-48)
5. 96 variant templates with 100 pM ligation oligos (low) (assays 1-48)
6. No template control with 100 pM ligation oligos (low) (assays 1-48)
7. 96 consensus templates with 500 pM ligation oligos (high) (assays 49-96)
8. 96 variant templates with 500 pM ligation oligos (high) (assays 49-96)
9. Calu6 gDNA library with 500 pM ligation oligos (high) (assays 1-48)
10. Calu6 gDNA library with 500 pM ligation oligos (high) (assays 49-96)
11. Calu6 enriched (E1) library* with 500 pM ligation oligos (high) (assays 1-48)
12. Calu6 enriched (E1) library* with 500 pM ligation oligos (high) (assays 49-96)

* The Calu6 enriched (E1) library is a pool of PCR Products from a Calu6 gDNA library that was enriched with a single round of solution-based capture for the Maxwell 139 gene set followed by PCR amplification, as described above.

For each genotyping ligation reaction, the following reagents were combined:

50 µl H_2O

20 µl target DNA (100 pM synthetic templates, or DNA samples: Calu6 gDNA (85 ng/µl); or Calu6 E1 (75 ng/µl))

10 µl (high) or 2 µl (low) of 500 nM 5' ligation oligo pool (consensus and variant)

10 µl (high) or 2 µl (low) of 250 nM kinased, 3' common oligo pool

10 µl of 10X Taq DNA ligase buffer (New England Biolabs, MA)

1 µl 5M NaCl

100 µl total volume, mix and add 2 µl Taq DNA ligase (New England Biolabs)

The ligation mixture was then incubated in a thermal cycler across the following temperatures:

95°C for 5 minutes;

75°C for 15 minutes;

70°C for 15 minutes;

65°C for 30 minutes;

60°C for 45 minutes;

55°C for 30 minutes;

50°C for 15 minutes;

45°C for 15 minutes;

4°C rest.

The ligation reactions were diluted to 1 ml with 900 µl of TEzero.

To measure the performance of the ligation-dependent genotyping assay, the following 6 ligation reactions were carried out on synthetic templates (all using high concentration (500 pM) ligation oligos), followed by qPCR analysis of each ligation reaction on a universal 384 well qPCR plate.

Templates: consensus templates, variant templates, no template control, Calu6 genomic DNA, and Calu6 enriched (E1) library.

Ligation Oligo pools: pool of 5' consensus and variant ligation oligos for assays 1-48 plus 3' common ligation primers for assays 1-48; pool of 5' consensus and variant ligation oligos for assays 49-96, plus 3' common ligation oligos for assays 49-96.

Therefore, for each set of SNVs of interest (e.g., assays 1-48, represent 48 different potential SNVs), a total of 5 ligation reactions were carried out:

1. ligation oligo pool (5' consensus, 5' variant, and 3' common) plus synthetic consensus templates; and
2. ligation oligo pool (5' consensus, 5' variant, and 3' common) plus synthetic variant templates;
3. ligation oligo pool (5' consensus, 5' variant, and 3' common) plus no template control.
4. ligation oligo pool (5' consensus, 5' variant, and 3' common) plus Calu6 genomic DNA.
5. ligation oligo pool (5' consensus, 5' variant, and 3' common) plus Calu6 enriched (E1) library.

Each ligation reaction was then plated onto a separate prepared universal qPCR plate and assayed, providing a set of qPCR results for ligation reaction #1 consensus template (qPCR plate 1), #2 variant template (qPCR plate 2), #3 no template (qPCR plate 3), #4 Calu6 gDNA (plate 9), and #5 Calu6 E1 library (plate 11).

Step 4: Quantitative PCR (qPCR):

Manufacture of Universal Assay Plate

The assay plates were prepared for quantitative PCR (qPCR) assays using the PCR primers as described in Example 2:

Briefly described, 35 mls of 2X SYBR master mix (ABI) was combined with 10 mls of H_2O. 450 µl of the mixture was aliquoted into each well of a 96 well assay plate. 55 µl of the "C" (reverse) primers (10 µM) were added to the wells along the columns of the assay plate, and 55 µl of the "R" (forward) primers (10 µM) were added to the wells along the rows of the assay plate, as shown above in TABLE 6. The reagents were mixed, then 8 µl per well was aliquoted in quadruplicate into a 384 qPCR plate, in order to carry out 4 identical reactions for each qPCR primer pair.

Quantitative PCR Assay:

120 µl aliquots of each diluted genotyping ligation reaction were distributed into 8 wells of the 384 well qPCR plate. Then, 2 µl aliquots were dispensed into all wells of the prepared 384 well qPCR plate (4 x 96). The samples were mixed, and the qPCR assay was run on an ABI 7900 instrument set on SYBR detection channel.

qPCR Results of Ligation-Dependent Genotyping Assay

As a measurement of assay performance, the average raw Ct data from each of the qPCR assays was first determined across four wells of each quadruplicate for assays 1-96 (high primer input). The results of the ligation with consensus templates (plates 1 and 4) or variant templates (plates 2 and 5) were measured against a no template control (plates 3 and 6), to obtain a set of raw Ct data (data not shown).

Dynamic Range of the Ligation-dependent Genotyping Assays

In order to determine the dynamic range of each assay for a SNV position, from the raw Ct data, the Ct spread between consensus and variant ligation assays using consensus templates (e.g., plate 1) was determined. Then, the Ct spreads for variant versus consensus ligation assays when variant templates were measured (e.g., plate 2) was calculated. The sum of the Ct spreads for plate 1 and plate 2 were calculated, which represents the complete dynamic range of the assay.

For example, for assay #1, the Ct spread between consensus and variant ligation assays using a consensus template (Ct(var)-Ct(cons) = 3. The Ct spread for variant versus consensus ligation assays when variant template was measured (Ct(cons)-Ct(var) = 2. The sum of the Ct spreads (3+2 = 5) represents the complete dynamic range of assay #1.

It was determined that the sum of the Ct spreads for plate 1 (consensus template) and plate 2 (variant template) was ≥ 5 Cts for all but two of assays (assay #15 and #17), which is a very tractable Ct spread. Significantly, when the same analysis was performed on ligation reactions using a lower concentration of ligation oligos (100 pM), every single assay registered a dynamic range greater than 5 Cts (data not shown). The average dynamic range for the ligation-dependent genotyping assays carried out with high ligation oligo concentration (500 pM oligos) was 9.1 Cts, whereas the average dynamic range for the ligation-dependent genotyping assays carried out with low ligation oligo concentration (100 pM oligos) was 10.4 Cts. These results

demonstrate that the use of ligation oligos in the range of 100 pM improves assay performance by proving a greater dynamic range.

Scoring Scheme for Genotyping

The ligation-dependent genotyping assay results generated using the synthetic template for the consensus and variant versions of the target sequence were then used to generate a calibrating "truth," or "reference" value for the Ct values that are expected from a test sample (diploid) that contains a homozygous consensus (con/con), heterozygous (con/var), or homozygous variant (var/var) for a particular polymorphic site of interest (e.g., SNV or SNP), as follows.

If the actual test sample contains a diploid homozygous consensus sequence (con/con) at the polymorphic locus of interest, then on average Ct(var) > Ct(cons) and the term [Ct(var)-Ct(cons)] is expected to return a positive integer value.

If the actual test sample contains a diploid heterozygote sequence (con/var) at the polymorphic locus of interest, then on average, Ct(var) ≈ Ct(cons) and the term [Ct(var)-Ct(cons)] is expected to return a value near zero.

If the actual test sample contains a diploid homozygous variant sequence (var/var) at the polymorphic locus of interest, then on average, Ct(var) < Ct(cons) and the term [Ct(var)-Ct(cons)] is expected to return a negative integer value.

The calibrating consensus and variant synthetic templates are scored as follows:

- Value homozygous consensus base = [Ct(var) - Ct(cons)] for consensus template measurements.
- Value heterozygous = Ct(var) for variant template - Ct(cons) for consensus template.
- Value homozygous variant base = [Ct(var) - Ct(cons)] for variant template.

The above scoring matrix was applied to the ligation-dependent genotyping assays using synthetic templates, and the results are shown below in TABLE 10, Column 2. The key observation is that all ligation-dependent genotyping assays 1-96, with the exception of assays 15 and 26, returned discrete integer values for each of the three genetic states. Importantly, it is noted that assays 15 and 26 did return discrete integer values when repeated with the more dilute ligation oligos (100 pmol) (data not shown).

TABLE 10: QPCR ASSAY RESULTS OF LIGATION-DEPENDENT GENOTYPING

Assay #	Scoring Matrix (based on synthetic template assays):	Test Sample: E1 Calu6 DNA Ct Value
1	homo-cons: 3 hetero: 0 homo-var: -2	0 (hetero)
2	homo-cons: 9 hetero: 1 homo-var: -3	-3 (homo-var)
3	homo-cons: 5 hetero: 1 homo-var: -2	-1 (homo-var)
4	homo-cons: 6 hetero: 1 homo-var: -3	-4 (homo-var)
5	homo-cons: 6 hetero: 1 homo-var: -2	2 (hetero)
6	homo-cons: 9 hetero: 1 homo-var: -3	-5 (homo-var)
7	homo-cons: 9 hetero: 1 homo-var: -5	-2 (homo-var)
8	homo-cons: 5 hetero: 1 homo-var: -2	-2 (homo-var)
9	homo-cons: 7 hetero: 3 homo-var: 0	1 (homo-var)
10	homo-cons: 9 hetero: 2 homo-var: -2	-3 (homo-var)
11	homo-cons: 11 hetero: 1 homo-var: -3	-3 (homo-var)

Assay #	Scoring Matrix (based on synthetic template assays):	Test Sample: E1 Calu6 DNA Ct Value
12	homo-cons: 7 hetero: 0 homo-var: -4	0 (hetero)
13	homo-cons: 8 hetero: 0 homo-var: -4	-3 (homo-var)
14	homo-cons: 4 hetero: 1 homo-var: -1	-1 (homo-var)
15	homo-cons: 0 hetero: 0 homo-var: -3	-3 (homo-var)
16	homo-cons: 4 hetero: 1 homo-var: -4	-5 (homo-var)
17	homo-cons: 2 hetero: 0 homo-var: -2	-1 (homo-var)
18	homo-cons: 3 hetero: 1 homo-var: -3	-3 (homo-var)
19	homo-cons: 6 hetero: 1 homo-var: -3	2 (hetero)
20	homo-cons: 8 hetero: 1 homo-var: -3	0 (hetero)
21	homo-cons: 7 hetero: 1 homo-var: -5	-5 (homo-var)
22	homo-cons: 5 hetero: 1 homo-var: -2	-2 (homo-var)
23	homo-cons: 7 hetero: 0 homo-var: -4	0 (hetero)

Assay #	Scoring Matrix (based on synthetic template assays):	Test Sample: E1 Calu6 DNA Ct Value
24	homo-cons: 4 hetero: 1 homo-var: -3	-2 (homo-var)
25	homo-cons: 8 hetero: 1 homo-var: -3	0 (hetero)
26	homo-cons: 9 hetero: 4 homo-var: 4	1 (hetero)
27	homo-cons: 8 hetero: 2 homo-var: -2	-3 (homo-var)
28	homo-cons: 5 hetero: 1 homo-var: -2	-2 (homo-var)
29	homo-cons: 6 hetero: 2 homo-var: -2	2 (hetero)
30	homo-cons: 4 hetero: 1 homo-var: -2	1 (hetero)
31	homo-cons: 8 hetero: 0 homo-var: -3	11 (homo-cons)
32	homo-cons: 7 hetero: 1 homo-var: -1	8 (homo-cons)
33	homo-cons: 6 hetero: 1 homo-var: -2	8 (homo-cons)
34	homo-cons: 9 hetero: 2 homo-var: -2	9 (homo-cons)
35	homo-cons: 1 hetero: -1 homo-var: -4	-3 (hetero-var)

Assay #	Scoring Matrix (based on synthetic template assays):	Test Sample: E1 Calu6 DNA Ct Value
36	homo-cons: 7 hetero: 1 homo-var: -2	-2 (hetero-var)
37	homo-cons: 4 hetero: 1 homo-var: -3	0 (hetero)
38	homo-cons: 5 hetero: 1 homo-var: -3	6 (homo-cons)
39	homo-cons: 4 hetero: 0 homo-var: -4	-4 (homo-var)
40	homo-cons: 7 hetero: 1 homo-var: -2	8 (homo-cons)
41	homo-cons: 4 hetero: 1 homo-var: -1	3 (homo-cons)
42	homo-cons: 7 hetero: 1 homo-var: -3	8 (homo-cons)
43	homo-cons: 9 hetero: 1 homo-var: -4	10 (homo-cons)
44	homo-cons: 10 hetero: 0 homo-var: -5	11 (homo-cons)
45	homo-cons: 5 hetero: 0 homo-var: -6	5 (homo-cons)
46	homo-cons: 9 hetero: 1 homo-var: -4	9 (homo-cons)
47	homo-cons: 11 hetero: 1 homo-var: -2	12 (homo-cons)

Assay #	Scoring Matrix (based on synthetic template assays):	Test Sample: E1 Calu6 DNA Ct Value
48	homo-cons: 10 hetero: 2 homo-var: -2	14 (homo-cons)
49	homo-cons: 5 hetero: 0 homo-var: -4	1 (hetero)
50	homo-cons: 6 hetero: 2 homo-var: -5	6 (homo-cons)
51	homo-cons: 6 hetero: 0 homo-var: -4	6 (homo-cons)
52	homo-cons: 5 hetero: 0 homo-var: -4	5 (homo-cons)
53	homo-cons: 11 hetero: 2 homo-var: -1	13 (homo-cons)
54	homo-cons: 6 hetero: 0 homo-var: -3	-3 (homo-var)
55	homo-cons: 3 hetero: 0 homo-var: -5	3 (homo-cons)
56	homo-cons: 6 hetero: 0 homo-var: -4	0 (hetero)
57	homo-cons: 6 hetero: 1 homo-var: -3	1 (hetero)
58	homo-cons: 7 hetero: 2 homo-var: -3	8 (homo-cons)
59	homo-cons: 5 hetero: 0 homo-var: -6	1 (hetero)

Assay #	Scoring Matrix (based on synthetic template assays):	Test Sample: E1 Calu6 DNA Ct Value
60	homo-cons: 6 hetero: 1 homo-var: -6	7 (homo-cons)
61	homo-cons: 5 hetero: 1 homo-var: -5	5 (homo-cons)
62	homo-cons: 4 hetero: 1 homo-var: -3	6 (homo-cons)
63	homo-cons: 5 hetero: 0 homo-var: -5	5 (homo-cons)
64	homo-cons: 6 hetero: 1 homo-var: -6	7 (homo-cons)
65	homo-cons: 3 hetero: 0 homo-var: -5	3 (homo-cons)
66	homo-cons: 5 hetero: 2 homo-var: -6	5 (homo-cons)
67	homo-cons: 6 hetero: 0 homo-var: -6	6 (homo-cons)
68	homo-cons: 6 hetero: 1 homo-var: -6	7 (homo-cons)
69	homo-cons: 6 hetero: 1 homo-var: -6	7 (homo-cons)
70	homo-cons: 6 hetero: 2 homo-var: -6	7 (homo-cons)
71	homo-cons: 6 hetero: 0 homo-var: -5	8 (homo-cons)

Assay #	Scoring Matrix (based on synthetic template assays):	Test Sample: E1 Calu6 DNA Ct Value
72	homo-cons: 6 hetero: 0 homo-var: -6	8 (homo-cons)
73	homo-cons: 5 hetero: 0 homo-var: -4	6 (homo-cons)
74	homo-cons: 7 hetero: 1 homo-var: -4	7 (homo-cons)
75	homo-cons: 5 hetero: 0 homo-var: -5	8 (homo-cons)
76	homo-cons: 6 hetero: 0 homo-var: -4	7 (homo-cons)
77	homo-cons: 4 hetero: 1 homo-var: -6	4 (homo-cons)
78	homo-cons: 4 hetero: 1 homo-var: -5	4 (homo-cons)
79	homo-cons: 6 hetero: 0 homo-var: -5	8 (homo-cons)
80	homo-cons: 6 hetero: 1 homo-var: -5	8 (homo-cons)
81	homo-cons: 4 hetero: 0 homo-var: -6	5 (homo-cons)
82	homo-cons: 6 hetero: 1 homo-var: -9	6 (homo-cons)
83	homo-cons: 5 hetero: 3 homo-var: -3	5 (homo-cons)

Assay #	Scoring Matrix (based on synthetic template assays):	Test Sample: E1 Calu6 DNA Ct Value
84	homo-cons: 3 hetero: 0 homo-var: -5	2 (homo-cons)
85	homo-cons: 5 hetero: 1 homo-var: -3	3 (homo-cons)
86	homo-cons: 5 hetero: 1 homo-var: -4	6 (homo-cons)
87	homo-cons: 3 hetero: 0 homo-var: -7	4 (homo-cons)
88	homo-cons: 5 hetero: 0 homo-var: -5	6 (homo-cons)
89	homo-cons: 7 hetero: 1 homo-var: -4	7 (homo-cons)
90	homo-cons: 8 hetero: 1 homo-var: -4	7 (homo-cons)
91	homo-cons: 6 hetero: 0 homo-var: -5	7 (homo-cons)
92	homo-cons: 6 hetero: 1 homo-var: -5	6 (homo-cons)
93	homo-cons: 3 hetero: 0 homo-var: -5	2 (homo-cons)
94	homo-cons: 9 hetero: 2 homo-var: -5	11 (homo-cons)
95	homo-cons: 7 hetero: 0 homo-var: -5	7 (homo-cons)

Assay #	Scoring Matrix (based on synthetic template assays):	Test Sample: E1 Calu6 DNA Ct Value
96	homo-cons: 6 hetero: 1 homo-var: -5	6 (homo-cons)

Genotyping of Calu6 test samples

There were two samples tested that were derived from Calu6 gDNA. The first was genomic DNA (gDNA), and the second was a population of PCR products that were generated from a library made from Calu6 gDNA (E1 Calu6 DNA) which was enriched for the Maxwell 139 set of genes by solution-based capture, as described above.

In an initial experiment, genomic DNA gave little signal above background when tested in the ligation-dependent genotyping assay. For example, the average decrease in Ct (corresponding to an increase in signal) for Calu6 gDNA versus background for 96 assays (plate 9 versus plate 3) was 1.3 Cts (data not shown). However, it is noted that the 96 assays with Calu6 gDNA were carried out with a high concentration (500 pM) of ligation oligos. As described above, it was determined in experiments with the synthetic templates that reducing the primer concentration to 100 pM increased the dynamic range, thereby improving the sensitivity of the assay (i.e., increased signal-to-noise ratio). Such improved sensitivity with a lower concentration of ligation oligos may allow for genotyping of gDNA using the ligation-dependent assay.

For the enriched (E1) Calu6 DNA test samples, the average decrease in Ct (corresponding to an increase in signal) was 5 Cts, as shown in TABLE 10, Column 3, which was adequate sensitivity for genotyping. As shown in TABLE 10, assignments of homozygous consensus alleles (con/con), heterozygous alleles (var/con), or homozygous variant alleles (var/var) for the E1 Calu6 DNA samples at each of the 96 polymorphic loci of interest were made by comparing the experimental values obtained from the E1 Calu6 DNA to the "truth set" shown as the "scoring matrix" in Column 2 of TABLE 10, based on the genotyping assays carried out using the synthetic templates. Genotypes were then assigned to the test samples based on the closest pairing between the experimental value and the scoring matrix.

TABLE 11 provides a comparison of the results of the ligation-dependent genotyping assay shown in TABLE 10 with the genotype initially determined from massive parallel sequencing. Assays 1 to 96 correspond to 96 distinct putative SNVs that were initially detected as potential polymorphic loci during massively parallel sequencing. The list of 96 assays is sorted by

highest (assay #1) to lowest (assay #96) confidence levels, with known database SNPs dominating the top portion of the list. In the situations where the polymorphic locus of interest corresponded to a SNP that is present in the dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/) or COSMIC (http://www.sanger.ac.uk/genetics/CGP/cosmic/), the corresponding SNP reference number is provided in Column 2. A "0" value in Column 2 means that the potential polymorphic loci was not present in the dbSNP or COSMIC.

TABLE 11: LIGATION DEPENDENT GENOTYPING ASSAY RESULTS

Assay #	Target sequence: dbSNP or COSMIC reference number	Initial Sequencing call (massively parallel sequence platform)	Scoring Matrix (based on synthetic template assays):	Test Sample: E1 Calu6 DNA	Comparison of ligation method to sequence method (confirm, differs)
1	rs6537825	homo-var	homo-cons: 3 hetero: 0 homo-var: -2	0 (hetero)	D
2	rs1052576	homo-var	homo-cons: 9 hetero: 1 homo-var: -3	-3 (homo-var)	C
3	rs2308941	homo-var	homo-cons: 5 hetero: 1 homo-var: -2	-1 (homo-var)	C
4	rs2230804	homo-var	homo-cons: 6 hetero: 1 homo-var: -3	-4 (homo-var)	C
5	rs1799939	hetero	homo-cons: 6 hetero: 1 homo-var: -2	2 (hetero)	C
6	rs144848	homo-var	homo-cons: 9 hetero: 1 homo-var: -3	-5 (homo-var)	C
7	rs1058808	homo-var	homo-cons: 9 hetero: 1 homo-var: -5	-2 (homo-var)	C
8	rs4986764	homo-var	homo-cons: 5 hetero: 1 homo-var: -2	-2 (homo-var)	C
9	rs6504459	homo-var	homo-cons: 7 hetero: 3 homo-var: 0	1 (homo-var)	C

Assay #	Target sequence: dbSNP or COSMIC reference number	Initial Sequencing call (massively parallel sequence platform)	Scoring Matrix (based on synthetic template assays):	Test Sample: E1 Calu6 DNA	Comparison of ligation method to sequence method (confirm, differs)
10	Cosmic	homo-var	homo-cons: 9 hetero: 2 homo-var: -2	-3 (homo-var)	C
11	rs1042522	homo-var	homo-cons: 11 hetero: 1 homo-var: -3	-3 (homo-var)	C
12	rs2229571	homo-var	homo-cons: 7 hetero: 0 homo-var: -4	0 (hetero)	D
13	rs2577301	homo-var	homo-cons: 8 hetero: 0 homo-var: -4	-3 (homo-var)	C
14	rs1670283	homo-var	homo-cons: 4 hetero: 1 homo-var: -1	-1 (homo-var)	C
15	rs753381	homo-var	homo-cons: 0 hetero: 0 homo-var: -3	-3 (homo-var)	C
16	rs235768	homo-var	homo-cons: 4 hetero: 1 homo-var: -4	-5 (homo-var)	C
17	rs20551	homo-var	homo-cons: 2 hetero: 0 homo-var: -2	-1 (homo-var)	C
18	rs376618	homo-var	homo-cons: 3 hetero: 1 homo-var: -3	-3 (homo-var)	C
19	rs2854746	hetero	homo-cons: 6 hetero: 1 homo-var: -3	2 (hetero)	C
20	rs1073123	homo-var	homo-cons: 8 hetero: 1 homo-var: -3	0 (hetero)	D
21	rs682632	homo-var	homo-cons: 7 hetero: 1 homo-var: -5	-5 (homo-var)	C

Assay #	Target sequence: dbSNP or COSMIC reference number	Initial Sequencing call (massively parallel sequence platform)	Scoring Matrix (based on synthetic template assays):	Test Sample: E1 Calu6 DNA	Comparison of ligation method to sequence method (confirm, differs)
22	rs1052571	homo-var	homo-cons: 5 hetero: 1 homo-var: -2	-2 (homo-var)	C
23	rs35093491	hetero	homo-cons: 7 hetero: 0 homo-var: -4	0 (hetero)	C
24	rs1801516	homo-var	homo-cons: 4 hetero: 1 homo-var: -3	-2 (homo-var)	C
25	Cosmic	hetero	homo-cons: 8 hetero: 1 homo-var: -3	0 (hetero)	C
26	rs1805097	homo-var	homo-cons: 9 hetero: 4 homo-var: 4	1 (hetero)	D
27	rs2240308	homo-var	homo-cons: 8 hetero: 2 homo-var: -2	-3 (homo-var)	C
28	rs1250209	homo-var	homo-cons: 5 hetero: 1 homo-var: -2	-2 (homo-var)	C
29	0	hetero	homo-cons: 6 hetero: 2 homo-var: -2	2 (hetero)	C
30	rs2228246	hetero	homo-cons: 4 hetero: 1 homo-var: -2	1 (hetero)	C
31	0	hetero	homo-cons: 8 hetero: 0 homo-var: -3	11 (homo-cons)	D
32	0	hetero	homo-cons: 7 hetero: 1 homo-var: -1	8 (homo-cons)	D
33	0	hetero	homo-cons: 6 hetero: 1 homo-var: -2	8 (homo-cons)	D

Assay #	Target sequence: dbSNP or COSMIC reference number	Initial Sequencing call (massively parallel sequence platform)	Scoring Matrix (based on synthetic template assays):	Test Sample: E1 Calu6 DNA	Comparison of ligation method to sequence method (confirm, differs)
34	0	hetero	homo-cons: 9 hetero: 2 homo-var: -2	9 (homo-cons)	D
35	rs351855	homo-var	homo-cons: 1 hetero: -1 homo-var: -4	-3 (hetero-var)	C
36	0	homo-var	homo-cons: 7 hetero: 1 homo-var: -2	-2 (hetero-var)	C
37	rs529038	hetero	homo-cons: 4 hetero: 1 homo-var: -3	0 (hetero)	C
38	0	homo-var	homo-cons: 5 hetero: 1 homo-var: -3	6 (homo-cons)	D
39	rs502209	homo-var	homo-cons: 4 hetero: 0 homo-var: -4	-4 (homo-var)	C
40	0	homo-var	homo-cons: 7 hetero: 1 homo-var: -2	8 (homo-cons)	D
41	0	hetero	homo-cons: 4 hetero: 1 homo-var: -1	3 (homo-cons)	D
42	0	hetero	homo-cons: 7 hetero: 1 homo-var: -3	8 (homo-cons)	D
43	0	hetero	homo-cons: 9 hetero: 1 homo-var: -4	10 (homo-cons)	D
44	0	hetero	homo-cons: 10 hetero: 0 homo-var: -5	11 (homo-cons)	D
45	0	homo-var	homo-cons: 5 hetero: 0 homo-var: -6	5 (homo-cons)	D

Assay #	Target sequence: dbSNP or COSMIC reference number	Initial Sequencing call (massively parallel sequence platform)	Scoring Matrix (based on synthetic template assays):	Test Sample: E1 Calu6 DNA	Comparison of ligation method to sequence method (confirm, differs)
46	rs1046984	hetero	homo-cons: 9 hetero: 1 homo-var: -4	9 (homo-cons)	D
47	0	hetero	homo-cons: 11 hetero: 1 homo-var: -2	12 (homo-cons)	D
48	0	hetero	homo-cons: 10 hetero: 2 homo-var: -2	14 (homo-cons)	D
49	rs7190823	hetero	homo-cons: 5 hetero: 0 homo-var: -4	1 (hetero)	D
50	0	hetero	homo-cons: 6 hetero: 2 homo-var: -5	6 (homo-cons)	D
51	0	hetero	homo-cons: 6 hetero: 0 homo-var: -4	6 (homo-cons)	D
52	0	hetero	homo-cons: 5 hetero: 0 homo-var: -4	5 (homo-cons)	D
53	0	hetero	homo-cons: 11 hetero: 2 homo-var: -1	13 (homo-cons)	D
54	rs243383	homo-var	homo-cons: 6 hetero: 0 homo-var: -3	-3 (homo-var)	D
55	0	hetero	homo-cons: 3 hetero: 0 homo-var: -5	3 (homo-cons)	D
56	rs2070094	hetero	homo-cons: 6 hetero: 0 homo-var: -4	0 (hetero)	D
57	rs17449032	hetero	homo-cons: 6 hetero: 1 homo-var: -3	1 (hetero)	D

Assay #	Target sequence: dbSNP or COSMIC reference number	Initial Sequencing call (massively parallel sequence platform)	Scoring Matrix (based on synthetic template assays):	Test Sample: E1 Calu6 DNA	Comparison of ligation method to sequence method (confirm, differs)
58	0	hetero	homo-cons: 7 hetero: 2 homo-var: -3	8 (homo-cons)	D
59	rs1881421	homo-var	homo-cons: 5 hetero: 0 homo-var: -6	1 (hetero)	D
60	0	hetero	homo-cons: 6 hetero: 1 homo-var: -6	7 (homo-cons)	D
61	0	hetero	homo-cons: 5 hetero: 1 homo-var: -5	5 (homo-cons)	D
62	0	homo-var	homo-cons: 4 hetero: 1 homo-var: -3	6 (homo-cons)	D
63	0	hetero	homo-cons: 5 hetero: 0 homo-var: -5	5 (homo-cons)	D
64	0	hetero	homo-cons: 6 hetero: 1 homo-var: -6	7 (homo-cons)	D
65	0	hetero	homo-cons: 3 hetero: 0 homo-var: -5	3 (homo-cons)	D
66	0	homo-var	homo-cons: 5 hetero: 2 homo-var: -6	5 (homo-cons)	D
67	0	homo-var	homo-cons: 6 hetero: 0 homo-var: -6	6 (homo-cons)	D
68	0	hetero	homo-cons: 6 hetero: 1 homo-var: -6	7 (homo-cons)	D
69	0	hetero	homo-cons: 6 hetero: 1 homo-var: -6	7 (homo-cons)	D

Assay #	Target sequence: dbSNP or COSMIC reference number	Initial Sequencing call (massively parallel sequence platform)	Scoring Matrix (based on synthetic template assays):	Test Sample: E1 Calu6 DNA	Comparison of ligation method to sequence method (confirm, differs)
70	0	hetero	homo-cons: 6 hetero: 2 homo-var: -6	7 (homo-cons)	D
71	0	hetero	homo-cons: 6 hetero: 0 homo-var: -5	8 (homo-cons)	D
72	0	hetero	homo-cons: 6 hetero: 0 homo-var: -6	8 (homo-cons)	D
73	0	hetero	homo-cons: 5 hetero: 0 homo-var: -4	6 (homo-cons)	D
74	0	hetero	homo-cons: 7 hetero: 1 homo-var: -4	7 (homo-cons)	D
75	0	hetero	homo-cons: 5 hetero: 0 homo-var: -5	8 (homo-cons)	D
76	0	homo-var	homo-cons: 6 hetero: 0 homo-var: -4	7 (homo-cons)	D
77	0	hetero	homo-cons: 4 hetero: 1 homo-var: -6	4 (homo-cons)	D
78	0	homo-var	homo-cons: 4 hetero: 1 homo-var: -5	4 (homo-cons)	D
79	0	hetero	homo-cons: 6 hetero: 0 homo-var: -5	8 (homo-cons)	D
80	0	hetero	homo-cons: 6 hetero: 1 homo-var: -5	8 (homo-cons)	D
81	0	hetero	homo-cons: 4 hetero: 0 homo-var: -6	5 (homo-cons)	D

Assay #	Target sequence: dbSNP or COSMIC reference number	Initial Sequencing call (massively parallel sequence platform)	Scoring Matrix (based on synthetic template assays):	Test Sample: E1 Calu6 DNA	Comparison of ligation method to sequence method (confirm, differs)
82	0	hetero	homo-cons: 6 hetero: 1 homo-var: -9	6 (homo-cons)	D
83	0	hetero	homo-cons: 5 hetero: 3 homo-var: -3	5 (homo-cons)	D
84	0	hetero	homo-cons: 3 hetero: 0 homo-var: -5	2 (homo-cons)	D
85	0	hetero	homo-cons: 5 hetero: 1 homo-var: -3	3 (homo-cons)	D
86	0	hetero	homo-cons: 5 hetero: 1 homo-var: -4	6 (homo-cons)	D
87	0	hetero	homo-cons: 3 hetero: 0 homo-var: -7	4 (homo-cons)	D
88	0	hetero	homo-cons: 5 hetero: 0 homo-var: -5	6 (homo-cons)	D
89	0	hetero	homo-cons: 7 hetero: 1 homo-var: -4	7 (homo-cons)	D
90	0	hetero	homo-cons: 8 hetero: 1 homo-var: -4	7 (homo-cons)	D
91	0	hetero	homo-cons: 6 hetero: 0 homo-var: -5	7 (homo-cons)	D
92	0	hetero	homo-cons: 6 hetero: 1 homo-var: -5	6 (homo-cons)	D
93	0	hetero	homo-cons: 3 hetero: 0 homo-var: -5	2 (homo-cons)	D

Assay #	Target sequence: dbSNP or COSMIC reference number	Initial Sequencing call (massively parallel sequence platform)	Scoring Matrix (based on synthetic template assays):	Test Sample: E1 Calu6 DNA	Comparison of ligation method to sequence method (confirm, differs)
94	0	hetero	homo-cons: 9 hetero: 2 homo-var: -5	11 (homo-cons)	D
95	0	hetero	homo-cons: 7 hetero: 0 homo-var: -5	7 (homo-cons)	D
96	0	hetero	homo-cons: 6 hetero: 1 homo-var: -5	6 (homo-cons)	D

As shown above in TABLE 11, all but one dbSNP call and both of the COSMIC SNP calls were validated by the ligation-dependent genotyping assay. Also, in most cases (31/36 = 86%), the heterozygous versus homozygous assignment from the ligation-dependent genotyping assay agreed with the results from sequence analysis. Two novel missense alleles identified by sequencing were validated in the ligation-dependent genotyping assay. All the other sequencing calls that indicated a potential SNV that were tested in the ligation-dependent genotyping assay proved to be false.

Conclusion: This Example demonstrates that the ligation-dependent genotyping assay can be successfully multiplexed in a single reaction tube and read out on a universal PCR matrix. The use of reference consensus and reference variant templates in a multiplex ligation assay allows for a simple scoring scheme for genotyping a test sample that is amenable to high throughput automation and analysis. The results described in Example 1 and in this Example demonstrate the successful genotyping of 144 of 144 SNV loci of interest, a 100% conversion rate (i.e., the percentage of designed assays that produce meaningful results).

As further described in this Example, it was determined that the use of lower concentrations of ligation primers (e.g., about 100 pM) reduce the background signal in the qPCR assay that was observed at high concentrations of ligation primers (e.g., about 500 pM). A 5-fold decrease in input ligation primer concentration (at fixed template) decreased signal by only 1.5 Cts, but decreased background signal by 3 Cts in real time qPCR measurements. This improved signal at decreased ligation oligo concentration indicates that it will be possible to

multiplex hundreds of genotyping assays in a single reaction without compromising assay readout accuracy.

Taken together, the results described herein form the basis for an inexpensive and very high throughput two-step sequence validation/genotyping system. In step one, ligation oligos (potentially 1000 or more at once) are mixed with a sample, annealed and ligated in a single reaction mixture. In step two, the ligation mixture is distributed across a universal PCR "decoding" matrix, which can be dispensed into one or more multi-well assay plates and stored in a freezer prior to use, as described in Examples 2 and 4. The magnitude of the qPCR signal is indicative of the underlying genotype at a given SNV position of interest. As demonstrated herein, the ligation-dependent assay can distinguish between heterozygous and homozygous states in a diploid genome.

EXAMPLE 4

This Example describes the manufacture of a 576 feature matrix of detection primers (also referred to as a "universal PCR decoding matrix"), which can be pre-made and stored in a freezer, for decoding a multiplex assay, such as a multiplex ligation-dependent genotyping assay for genotyping a test sample at a plurality of SNV positions of interest.

Rationale:

As described in Example 2, an important element of the universal PCR decoding matrix is that the last (i.e., penultimate) two or three 3' bases of the PCR primers are chosen to reduce and preferably eliminate primer-dimer formation, and the remaining bases are specificity tags chosen to provide a unique address at an intersection position (also referred to as a "feature"), in the matrix, such as a particular well on a multi-well assay plate. The universal PCR decoding matrix may be disposed into one or more multi-well assay plates.

This Example describes the manufacture of a 576 feature matrix of detection primers (universal PCR decoding matrix), that has minimal primer-dimer background due to the fact that the last three 3' bases of the PCR primers were chosen to avoid primer-dimer formation. The 576 feature matrix was dispensed into a total of six 384-well assay plates, wherein each plate contained 96 primer pairs (i.e., features) in adjacent quadruplicate wells, and stored in a freezer for use in decoding a multiplex PCR assay.

PCR Primer Matrix Design

The goal of this Example was to design a larger matrix of minimally interacting primer pairs to manufacture a 576 feature matrix of detection primer pairs. A combined bioinformatic and empirical approach was used to create the 576 feature primer matrix that has minimal primer-dimer background and therefore the greatest possible measurement dynamic range for genotyping assays.

Since A residues and C residues cannot base pair with themselves or with C or A, respectively, these sequences were used as trinucleotides on the 3' ends of primers as the basis of a minimally interactive, non-primer-dimer forming primer matrix. Specifically, one set of 36 potential primers was designed to end in "ACA," and a second set of 36 primers was designed to end in "CAC". Both primer sets were composed entirely of 25 nucleotide sequences. The 22 nucleotide "address" portions of each primer that are located at the 5' end of each primer were screened from a computationally selected randomized list of 22 nt sequences that were specified to contain at least four of each A, C, G, or T DNA residues. Each candidate 22 nt sequence was screened for "GTG" and "TGT" sequences within 9 nt of the 3' end of the 22 nt sequence, and those terminal 9 nt sequences containing these trinucleotides were eliminated. The rationale for this screening step is that the terminal "ACA" can pair with "TGT" and the terminal "CAC" can pair with "GTG". Hence by eliminating potential 22 nt address sequences that possess 9 nt terminal "GTG" or "TGT" sequences, the probability of spurious primer-dimer formation is further reduced.

As shown in FIGURE 3C, each forward PCR primer 600 has a 5' region 602 that binds to a primer binding region 302, 403 in the 5' tail of a 5' ligation oligo 300, 400, and a region 606 at the 3' end having a sequence selected to inhibit primer-dimer interactions with the reverse PCR primer 700. In this Example, the forward PCR primers 600 are located in rows in the primer matrix, and the "ACA" series was arbitrarily chosen to occupy these row positions. Similarly, as shown in FIGURE 3C, each reverse PCR primer 700 has a 5' region 702 that binds to a primer binding region 502 in the 3' tail of a 3' ligation oligo, and a region 706 at the 3' end having a sequence selected to inhibit primer-dimer interactions with the forward PCR primer 600. In this Example, the reverse PCR primers 700 are located in columns in the primer matrix, and the "CAC" series was arbitrarily chosen to represent the column primers.

The 36 primer sequences in the "ACA" row series were designed as the forward primer set 600 to bind to the 5' tail region 302, 402 on the ligation products 200, 250.

The 36 primer sequences in the "CAC" column series were designed as the reverse primer set 700 to bind to the 3' common tail region 502 on the ligation products 200, 250.

The set of 36 "column" primers ("CAC" series) and 36 "row" primers ("ACA" series) was empirically tested in a complete "all-by-all" matrix for the formation of primer dimers, as follows.

The primers were synthesized by MWG/Operon (Huntsville, Alabama), diluted to a working stock concentration of 10 μM, and 4 μl of "row" primers and 4 μl of "column" primers were added in rows and columns, respectively, to a 96 well plate that contained 42 μl of PCR mix in each well. The PCR mix was composed of 25 μl of 2X Power SYBR master mix (Applied Biosystems, Foster City, CA) and 17 μl of water. The entire matrix collection of 36 row primers and 36 column primers occupied fifteen 96 well plates. For each 96 well plate, ten microliters of PCR mix from each unique well was aliquoted in quadruplicate to 384 well optical PCR plates (Applied Biosystems) and these were run for 40 cycles under standard SYBR green PCR cycling conditions on an ABI7900 qPCR instrument (Applied Biosystems).

Results:

Each set of quadruplicate wells was analyzed for the average Ct value with the goal of identifying a primer matrix where all Cts are 35 or higher. While certain addresses in the 36 by 36 primer matrix had Cts lower than this, by eliminating 12 of the "CAC" column primers and 12 of the "ACA" row primers, a matrix where all primer pairs yield background Cts > 35 was identified, as shown in TABLE 12 and TABLE 13.

TABLE 12: SET OF 24 ROW "ACA" AND 24 COLUMN "CAC" PRIMERS

Reference No.	Sequence (5' to 3')	SEQ ID NO:
CAC1	5' TTATCCCGAGAATTCAGACAGTCAC 3'	346
CAC3	5' CACGGGAGTTGATCCTGGTTTTCAC 3'	347
CAC4	5' TATAGCCGCTTAAGTCTACACTCAC 3'	348
CAC5	5' GTTTCGTAGCGTCCTGGAGTATCAC 3'	349
CAC6	5' AAAATGTTCTATATGACCGTTCCAC 3'	350
CAC8	5' ATCAAGAGTTTAGCACTTCGCGCAC 3'	351
CAC9	5' GGCGATGATAGATTCCCCTCGTCAC 3'	352
CAC13	5' TATGCGCTGGCAACATCGACACCAC 3'	353
CAC14	5' CAGAGATCATCCGAAGGCTTCTCAC 3'	354
CAC16	5' ATTATGAGACTCCCCGACGTCCCAC 3'	355

CAC18	5' TGTGATCGACGGCCTTTCAAATCAC 3'	356
CAC19	5' AACCCAACTCTGGCAAGCGTTACAC 3'	357
CAC20	5' GCGAAGGATTTGCTGACTTAAGCAC 3'	358
CAC21	5' ATATTCATGTGCAAAAGCCTCCCAC 3'	359
CAC24	5' CGTAACCCCAGACATAGGCCTTCAC 3'	360
CAC25	5' CTGAAAGCGGTCGACTAACGGGCAC 3'	361
CAC28	5' AATTGGCGTATACGGCCCCAAGCAC 3'	362
CAC29	5' CGTCTCAACTTAAGCCAGCCGACAC 3'	363
CAC30	5' AATAGCCCGGCTTTATACGCTGCAC 3'	364
CAC31	5' CGCTTGCGACCTCTTAAAACGTCAC 3'	365
CAC32	5' GTCATACATAACTCTTGAGATCCAC 3'	366
CAC33	5' TCGATCGCTTCAGACTATTTCGCAC 3'	367
CAC34	5' ACGATGGTTTGTTTCAGGAAACCAC 3'	368
CAC35	5' ACACACTTCCAGGCGATGGAAACAC 3'	369
ACA1	5' AACCGCTACAAGGCGGGGCACCACA 3'	370
ACA2	5' ACCAAACCTAGTAGCGCTATCCACA 3'	371
ACA3	5' TGCAGGACCAGAGAATTCGAATACA 3'	372
ACA6	5' ACTCAACATCGGCATCGGGCCTACA 3'	373
ACA7	5' ATTTCTACAAACGCTCGCCACAACA 3'	374
ACA8	5' AAAAATCCAAGTTTTAGGCGTTACA 3'	375
ACA9	5' TCTATCTATGGCCATGGTCTAAACA 3'	376
ACA10	5' GATTGCGCGGTAATAGCGCCCTACA 3'	377
ACA12	5' TCTTACGTGATGATATGGCAACACA 3'	378
ACA13	5' TCCCCTAGCACCCTAGGGTATGACA 3'	379
ACA17	5' TAAGTATTCCATGCACCCCTAAACA 3'	380
ACA19	5' CCTTACCTCGTAACTAACTAAGACA 3'	381
ACA21	5' TCTGGACAAGATTAGCTTACCAACA 3'	382
ACA22	5' TAACCGATACGTACGAGAGGCAACA 3'	383
ACA24	5' CCTGGACGAGGATTGACTCTACACA 3'	384
ACA25	5' TTCGGTTAGGTCCTACCGTACAACA 3'	385
ACA26	5' ATTCGAACGCTATCGAAAGGTTACA 3'	386

ACA27	5' AAGGATTGAGTCACATGGCGCAACA 3'	387
ACA28	5' TCAGCTAAGCCCTTATGATCCGACA 3'	388
ACA29	5' CATAAGCGAGTCATACTGACGAACA 3'	389
ACA30	5' CGAATGGATCAGTAACTCGAGAACA 3'	390
ACA31	5' GAAAGCAGGCAGGCCACTGACTACA 3'	391
ACA32	5' TAGAAACTCGACCAGAGGAGCTACA 3'	392
ACA33	5' GCTATCGGGGAATCCGCATCACACA 3'	393

TABLE 13: ASSAY RESULTS OF PCR PRIMER MATRIX FOR THE PRESENCE OF PRIMER DIMERS.

Ct values	CAC1	CAC3	CAC4	CAC5	CAC6	CAC8	CAC9	CAC13	CAC14	CAC16	CAC18	CAC19	CAC20	CAC21	CAC24	CAC25	CAC28	CAC29	CAC30	CAC31	CAC32	CAC33	CAC34	CAC35
ACA1	40	40	40	40	40	40	40	40	39	39	40	40	40	40	40	40	40	40	40	40	40	40	40	40
ACA2	40	40	40	40	40	40	40	40	40	40	40	40	40	38	40	40	40	40	39	40	40	40	39	40
ACA3	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40
ACA6	40	40	40	39	39	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40
ACA7	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40
ACA8	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40
ACA9	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	39	40	40
ACA10	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	39	40
ACA12	40	40	40	40	40	40	39	40	40	40	40	40	40	39	40	40	40	40	40	40	40	40	40	40
ACA13	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40
ACA17	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	39	40
ACA19	40	40	40	40	40	40	40	40	40	40	40	40	40	40	39	40	40	40	40	40	40	40	40	40
ACA21	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40
ACA22	40	40	40	40	40	38	40	40	40	40	40	40	40	40	40	40	40	40	39	40	40	40	40	40
ACA24	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40
ACA25	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40
ACA26	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	39	40	40	40	40	40	40	40	40
ACA27	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	36	40	40
ACA28	39	40	40	40	40	40	40	40	40	40	40	40	40	40	40	39	40	40	40	40	40	40	40	40
ACA29	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40
ACA30	40	40	40	40	39	40	40	40	40	39	40	40	40	40	40	40	40	40	40	40	40	40	40	40
ACA31	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40
ACA32	40	40	40	40	40	39	40	40	40	40	40	40	40	40	40	39	40	40	40	40	39	40	40	40
ACA33	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	38	40	40	40

As shown above in TABLE 13, a matrix of 24 "CAC" column primers and 24 "ACA" row primers was identified where all primer pairs yielded a background level of Cts > 35.

As demonstrated in this Example, by using the described combined informatic and empirical approach, a set of 24 "CAC" column primers (SEQ ID NOS:346-369) and 24 "ACA" row primers (SEQ ID NOS:370-393) have been identified that fulfill the criterion of being a minimally interactive, low primer dimer forming matrix. The complete set of primers that comprise this matrix are shown in TABLE 12.

The universal PCR decoding matrix containing 24 column primers and 24 row primers (576 features) was dispensed into a total of six 384 well assay plates, wherein each plate contained 96 primer pairs (features) in adjacent quadruplicate wells. The assay plates containing the universal PCR decoding matrix were stored in a freezer for use in decoding a multiplex PCR assay as described herein.

EXAMPLE 5

This Example describes a method of ligation-dependent genotyping using separate annealing and ligation steps, and various other assay modifications that result in improved assay performance.

Rationale:

This Example describes a series of experiments that were carried out to determine the effect of various assay modifications on the performance of the ligation-dependant genotyping assay, including the use of separate annealing and ligation reaction conditions, the effect of different monovalent cations (e.g., Na+, K+, NH4+) on ligation efficiencies, the effect of ligation temperature, the effect of different ligases (TAQ or T4 DNA ligase), and the effect of ligase enzyme concentration and the length of ligation.

Methods:

A set of eight genotyping assays were designed to measure 8 SNV positions of interest under the various assay conditions as follows:

1. Preparation of Reagents for Ligation-Dependent Genotyping Assays

Synthetic Templates: The synthetic templates corresponding to the wild-type (consensus) allele, and the variant allele for each of the 8 SNV positions is provided in TABLE 14 (reverse complement sequences are shown). The length of each synthetic template is 51 nucleotides, with

the polymorphic site (shown as underlined) located in the center of the template (i.e., 25 nucleotides on either side of the SNV position of interest).

Ligation Oligonucleotides: Each assay described in this Example was carried out with two different 5' allele-specific ligation oligos 300, 400 and one common, phosphorylated 3' ligation oligo 500 (e.g., as illustrated in FIGURE 2).

The 5' ligation oligos 300, 400 for assaying the 8 SNV positions of interest, shown in TABLE 15, were designed to have a total length of 51 nucleotides, with a 25 nt first primer binding tail region 302, 402 (underlined) at the 5' most end, a 25 nt region of complementarity to the target template 304, 404, and a one nucleotide 3' allele-specific region 306, 406 shown as underlined in bold.

The 3' common phosphorylated [P] ligation oligos 500 for assaying the 8 SNV positions of interest, also shown in TABLE 15, were designed to have a total length of 50 nucleotides, with a 5' target-specific binding region 504 of 25 nucleotides selected to hybridize immediately 3' of the SNV position of interest, and a region 502 at the 3' end that contains a second PCR primer binding region that is 25 nucleotides (underlined).

TABLE 14: SYNTHETIC TEMPLATES
(REVERSE COMPLEMENT SEQUENCE IS SHOWN)

SNV #	Ref. #	Template	Sequence (5' to 3')	SEQ ID NO:
1	CT5	consensus	CTGGGGGTAACTGTGCCTATTCGAG**G**GGTCCCTATGGGACTTGGGGTCCTC	394
1	VT5	variant	CTGGGGGTAACTGTGCCTATTCGAG**A**GGTCCCTATGGGACTTGGGGTCCTC	395
2	CT6	consensus	GTGCTGGCTTTGCTGGAGCTGGCGC**A**GCAGGACCACGGTGCTCTGGACTGC	396
2	VT6	variant	GTGCTGGCTTTGCTGGAGCTGGCGC**G**GCAGGACCACGGTGCTCTGGACTGC	397
3	CT7	consensus	TCCAGCACTCTGTCATGAGGCTGTA**C**ATTCTGGGTGGGCAGTCTTCAGAGC	398
3	VT7	variant	TCCAGCACTCTGTCATGAGGCTGTA**G**ATTCTGGGTGGGCAGTCTTCAGAGC	399
4	CT8	consensus	CAACATCGACTTTGGCGAGCCCGGG**G**CCCGCCTGTCGCCGCCCGCGCCTCC	400
4	VT8	variant	CAACATCGACTTTGGCGAGCCCGGG**C**CCCGCCTGTCGCCGCCCGCGCCTCC	401
5	CT13	consensus	GCGAGTATTACTGCTACTCGAAATG**C**AAAAGCCACTCCAAGGCTCCGGAAA	402
5	VT13	variant	GCGAGTATTACTGCTACTCGAAATG**A**AAAAGCCACTCCAAGGCTCCGGAAA	403
6	CT14	consensus	CCAAATCGACTTACTCCTTTGCAGA**C**AGAAACAGCCTCCTTGGACAAGGAT	404
6	VT14	variant	CCAAATCGACTTACTCCTTTGCAGA**T**AGAAACAGCCTCCTTGGACAAGGAT	405
7	CT15	consensus	TCTCAGGATGCACCCAGTGGGCTCG**A**GGTCAGGGTGGCCTTGCCGGTGTCC	406
7	VT15	variant	TCTCAGGATGCACCCAGTGGGCTCG**C**GGTCAGGGTGGCCTTGCCGGTGTCC	407
8	CT16	consensus	CCTCACCAGAGGTGCCACCTACAAC**G**TCATAGTGGAGGCACTGAAAGACCA	408

SNV #	Ref. #	Template	Sequence (5' to 3')	SEQ ID NO:
8	VT16	variant	CCTCACCAGAGGTGCCACCTACAAC**A**TCATAGTGGAGGCACTGAAAGACCA	409

TABLE 15: 5' AND 3' LIGATION OLIGONUCLEOTIDES

Ref #	Template Target	Sequence (5' to 3')	SEQ ID NO:
5' Ligation Olios			
FC5	CT5	AACCGCTACAAGGCGGGGCACCACAGAGGACCCCAAGTCCCATAGGGACC**C**	410
FV5	VT5	ACCAAACCTAGTAGCGCTATCCACAGAGGACCCCAAGTCCCATAGGGACC**T**	411
FC6	CT6	TGCAGGACCAGAGAATTCGAATACAGCAGTCCAGAGCACCGTGGTCCTGC**T**	412
FV6	VT6	ACTCAACATCGGCATCGGGCCTACAGCAGTCCAGAGCACCGTGGTCCTGC**C**	413
FC7	CT7	ATTTCTACAAACGCTCGCCACAACAGCTCTGAAGACTGCCCACCCAGAAT**G**	414
FV7	VT7	AAAAATCCAAGTTTTAGGCGTTACAGCTCTGAAGACTGCCCACCCAGAAT**C**	415
FC8	CT8	TCTATCTATGGCCATGGTCTAAACAGGAGGCGCGGGCGGCGACAGGCGGG**C**	416
FV8	VT8	GATTGCGCGGTAATAGCGCCCTACAGGAGGCGCGGGCGGCGACAGGCGGG**G**	417
FC13	CT13	AACCGCTACAAGGCGGGGCACCACATTTCCGGAGCCTTGGAGTGGCTTTT**G**	418
FV13	VT13	ACCAAACCTAGTAGCGCTATCCACATTTCCGGAGCCTTGGAGTGGCTTTT**T**	419
FC14	CT14	TGCAGGACCAGAGAATTCGAATACAATCCTTGTCCAAGGAGGCTGTTTCT**G**	420
FV14	VT14	ACTCAACATCGGCATCGGGCCTACAATCCTTGTCCAAGGAGGCTGTTTCT**A**	421
FC15	CT15	ATTTCTACAAACGCTCGCCACAACAGGACACCGGCAAGGCCACCCTGACC**T**	422
FV15	VT15	AAAAATCCAAGTTTTAGGCGTTACAGGACACCGGCAAGGCCACCCTGACC**G**	423
FC16	CT16	TCTATCTATGGCCATGGTCTAAACATGGTCTTTCAGTGCCTCCACTATGA**C**	424
FV16	VT16	GATTGCGCGGTAATAGCGCCCTACATGGTCTTTCAGTGCCTCCACTATGA**T**	425
3' Ligation Oligos (5' to 3')			
CP5	CT5 VT5	[P]CTCGAATAGGCACAGTTACCCCCAGGTGAAAACCAGGATCAACTCCCGTG	426
CP6	CT6 VT6	[P]GCGCCAGCTCCAGCAAAGCCAGCACGTGAAAACCAGGATCAACTCCCGTG	427
CP7	CT7 VT7	[P]TACAGCCTCATGACAGAGTGCTGGAGTGAAAACCAGGATCAACTCCCGTG	428
CP8	CT8 VT8	[P]CCCGGGCTCGCCAAAGTCGATGTTGGTGAAAACCAGGATCAACTCCCGTG	429
CP13	CT13 VT13	[P]CATTTCGAGTAGCAGTAATACTCGCGTGATACTCCAGGACGCTACGAAAC	430
CP14	CT14 VT14	[P]TCTGCAAAGGAGTAAGTCGATTTGGGTGATACTCCAGGACGCTACGAAAC	431
CP15	CT15 VT15	[P]CGAGCCCACTGGGTGCATCCTGAGAGTGATACTCCAGGACGCTACGAAAC	432

Ref #	Template Target	Sequence (5' to 3')	SEQ ID NO:
CP16	CT16 VT16	[P]GTTGTAGGTGGCACCTCTGGTGAGGGTGATACTCCAGGACGCTACGAAAC	433

2. Pooling the Template Oligos

For each target SNV position of interest to be assayed, a set of control oligonucleotides were synthesized to generate double-stranded synthetic consensus and variant templates, with the reverse complement template sequences shown in TABLE 14.

Template oligonucleotides (sense and anti-sense template oligonucleotides) were mixed in two separate pools of 8 templates, resulting in a first pool containing 8 synthetic templates containing the consensus alleles for the 8 SNV positions of interest, and a second pool containing 8 synthetic templates containing the variant alleles for the 8 SNV positions of interest, and each pool was diluted to 10 pM.

3. Pooling the ligation oligos

The consensus and variant 5' ligation oligos were combined and diluted to 500 nM (31.25 nM in each individual sequence).

The 3' common ligation primers were kinased in a 100 µl reaction containing a 1 µM mixture of primers (62.5 nM in each sequence), 1X kinase buffer (New England Biolabs, Ipswich, MA), 1 mM ATP, and 20 U of T4 polynucleotide kinase. The reaction mixture was incubated at 37°C for 30 minutes and 65°C for 20 minutes. The kinased 3' common ligation primers were then diluted to a final working concentration of 250 nM.

4. Quantitative PCT Assay (qPCR)

qPCR primers were synthesized as shown below in TABLE 16.

TABLE 16: QPCR PRIMERS

Ref No	Sequence (5' to 3')	SEQ ID NO:
CAC3	CACGGGAGTTGATCCTGGTTTTCAC	347
CAC5	GTTTCGTAGCGTCCTGGAGTATCAC	349
ACA1	AACCGCTACAAGGCGGGGCACCACA	370
ACA2	ACCAAACCTAGTAGCGCTATCCACA	371
ACA3	TGCAGGACCAGAGAATTCGAATACA	372

Ref No	Sequence (5' to 3')	SEQ ID NO:
ACA6	ACTCAACATCGGCATCGGGCCTACA	373
ACA7	ATTTCTACAAACGCTCGCCACAACA	374
ACA8	AAAAATCCAAGTTTTAGGCGTTACA	375
ACA9	TCTATCTATGGCCATGGTCTAAACA	376
ACA10	GATTGCGCGGTAATAGCGCCCTACA	377

The qPCR primers were used in qPCR assays at a final concentration of 800 nM in each primer.

The qPCR assay plates used in each experiment described in this Example were configured to test 8 consensus assays and 8 variant assays (16 total), across six different experimental conditions, in an assay plate format shown below in TABLE 17.

TABLE 17: qPCR ASSAY PLATE FORMAT

	CAC3	CAC5	CAC3	CAC5	CAC3	CAC5	CAC3	CAC5	CAC3	CAC5	CAC3	CAC5
ACA1	1	1	2	2	3	3	4	4	5	5	6	6
ACA2	1	1	2	2	3	3	4	4	5	5	6	6
ACA3	1	1	2	2	3	3	4	4	5	5	6	6
ACA6	1	1	2	2	3	3	4	4	5	5	6	6
ACA7	1	1	2	2	3	3	4	4	5	5	6	6
ACA8	1	1	2	2	3	3	4	4	5	5	6	6
ACA9	1	1	2	2	3	3	4	4	5	5	6	6
ACA10	1	1	2	2	3	3	4	4	5	5	6	6

Note: each number (1-6) indicates a different assay condition that was tested.

Although 96 wells are shown in the assay plate format depicted above in TABLE 17, it will be understood that each of the 96 positions represents a quadruplicate set of assay wells in a 384 well PCR plate.

Each qPCR assay was carried out in quadruplicate, with 10 µl of SYBR green PCR reaction mix (5 µl of 2X power SYBR master mix, Applied Biosystems, Foster City CA), 1.4 µl H_2O, 0.8 µl of 10 µM row and column primers and 2 µl of template (e.g., 2 µl of a genotyping assay reaction). The genotyping assay reactions are described below.

5. Annealing and Ligation Reactions

A. Determination of the effect of different monovalent cations Na^+, K^+, and NH_4^+, on ligation efficiencies.

Methods:

A coupled annealing/ligation reaction was performed in which different monovalent cationic salts were added to stimulate annealing of the genotyping primers to the complementary genotyping targets.

Stock solutions of 2.5 M KCl, 2.5 M NH_4Cl, and 2.5 M NaCl were prepared.

Genotyping Reactions:

Consensus synthetic templates or no template controls were assayed using 5' ligation oligos (consensus and variant) primer pools.

For each genotyping ligation reaction, the following reagents were combined:

75 µl H_2O

10 µl of 10 pM consensus synthetic template or water (no template control)

2 µl of 500 nM combined 5' consensus and variant primer pools (each individual query oligo was present in the final genotyping mix at a final concentration of 625 pM)

2 µl of 250 nM 3' kinased common primer pool (each individual query oligo was present in the final genotyping mix at a final concentration of 625 pM)

10 µl of 10X Taq DNA ligase buffer (New England Biolabs, Ipswich, MA (NEB))

2 µl of 2.5 M NaCl or 2.5 M KCl or 2.5 M NH_4Cl

100 µl total volume. 2 µl ligase enzyme (40 U/µl Taq DNA ligase, NEB) was added and the ligation mixture was then incubated in a thermal cycler across the following temperatures:

95°C for 5 minutes;

75°C for 15 minutes;

70°C for 15 minutes;

65°C for 30 minutes;

60°C for 45 minutes;

55°C for 30 minutes;

50°C for 15 minutes;

45°C for 15 minutes;

4°C rest.

The ligation reactions were diluted to 1 ml with 900 µl of TEzero (10 mM Tris pH 7.6, 0.1 mM EDTA) and 2 µl of each ligation reaction was assayed in quadruplicate qPCR reactions as described above in Section 4.

Results:

The average raw Ct data from each of the qPCR assays was first determined across four wells of each quadruplicate assay. The results of the ligation with consensus templates were measured against a no template control to obtain a set of raw Ct data (data not shown). The scoring scheme of genotyping was then applied to the Ct data as described in Example 3.

Table 18 below shows the Ct(variant) - Ct(consensus) assay results for each of the eight assays under the three salt conditions tested (NaCl, KCl and NH_4Cl), and the average Ct(consensus), Ct(variant), and Ct(background) for each monovalent cation.

TABLE 18: CONSENSUS TEMPLATE GENOTYPING CALLS FOR SNV POSITIONS 1-8 AND AVERAGE CTS FROM ASSAYS CARRIED OUT USING THE THREE DIFFERENT MONOVALENT CATIONS.

NaCl monovalent cation		**KCl monovalent cation**		**NH_4Cl monovalent cation**	
Ct(var)-Ct(cons)		**Ct(var)-Ct(cons)**		**Ct(var)-Ct(cons)**	
SNV 1	2	SNV 1	1	SNV 1	1
SNV 2	4	SNV 2	4	SNV 2	4
SNV 3	2	SNV 3	2	SNV 3	2
SNV 4	5	SNV 4	5	SNV 4	5
SNV 5	3	SNV 5	2	SNV 5	1
SNV 6	6	SNV 6	5	SNV 6	3
SNV 7	1	SNV 7	2	SNV 7	1
SNV 8	5	SNV 8	4	SNV 8	3
Average Ct (cons)	24	Average Ct (cons)	25	Average Ct (cons)	25
Average Ct (var)	28	Average Ct (var)	28	Average Ct (var)	27
Average Ct	31	Average Ct	31	Average Ct	32

NaCl monovalent cation		KCl monovalent cation		NH_4Cl monovalent cation	
Ct(var)-Ct(cons)		Ct(var)-Ct(cons)		Ct(var)-Ct(cons)	
(bgd)		(bgd)		(bgd)	

As shown above in TABLE 18, optimal assay performance is observed with NaCl, however there are relatively minor differences between the three cations tested. This result was unexpected because according to Takahashi et al., *J. Biol Chem.* *259*(16):10041-10047 (1984), Na^+ inhibits Taq DNA ligase activity, while K^+ and NH_4^+ stimulate enzyme activity.

B. Determination of the effect of separating the annealing and ligation steps, with either (1) a shorter annealing time, (2) different ligation enzymes, (3) various ligation temperatures, or (4) various ligation concentrations, on the performance of the ligation-dependent genotyping assay

Rationale: The genotyping assays described in Examples 1 and 3 above were carried out with coupled annealing/ligation reactions in which the oligonucleotide reagents were added in the presence of thermostable ligase and subjected to conditions that allowed hybridization of the query oligonucleotides to the target templates. The following experiments were carried out to determine whether the annealing of the query oligonucleotides to the target template and subsequent ligation reaction in separate steps would improve the performance of the genotyping assay, and to test the effect of a shorter annealing time, different ligation enzymes, various ligation temperatures, and various ligase concentrations, on the performance of the genotyping assay.

Methods:

Annealing of templates and assay oligos was carried out as follows for each genotyping assay:

10 µl of 10 pM synthetic template (consensus or variant)

2 µl of 500 nM 5' consensus and variant ligation primers

2 µl of 250 nM kinased 3' common ligation primers

2 µl of 5 M NaCl

16 µl total (Note: the NaCl concentration in this annealing reaction is twice the concentration used in the monovalent comparison experiment described above)

The annealing mixtures were incubated in a thermal cycler across the following temperatures for the following time periods:

1. Standard protocol (total: 170 minutes)

95°C for 5 minutes;
75°C for 15 minutes;
70°C for 15 minutes;
65°C for 30 minutes;
60°C for 45 minutes;
55°C for 30 minutes;
50°C for 15 minutes;
45°C for 15 minutes;
4°C rest.

2. Rapid annealing protocol (total: 65 min)

95°C for 5 minutes;
75°C for 15 minutes;
70°C for 15 minutes;
65°C for 30 minutes;
4°C rest.

Ligations:

1. Taq DNA ligase Reactions

A ligation mix "cocktail" was prepared containing:

10 µl of 10X Taq DNA ligase buffer (NEB)
72 µl of H_2O
2 µl of Taq DNA ligase (40 U/µl, NEB)

84 µl total, which was added to each annealed reaction mixture (16 µl) for a total volume of 100 µl in each ligation reaction. For the Taq DNA ligase reactions, ligations were performed at 37°C, 45°C, 55°C, and 65°C for 30 minutes.

For the rapid annealing protocol described above, follow on ligation with Taq DNA ligase was performed at 45°C for 30 minutes.

2. T4 DNA ligase Reactions

A ligation mix "cocktail" was prepared containing:

10 µl of 10X T4 DNA ligase buffer (NEB)

72 µl of H_2O

2 µl of T4 DNA ligase (400 U/µl (NEB))

84 µl total, which was added to each annealed reaction mixture (16 µl) for a total volume of 100 µl in each ligation reaction.

For the T4 DNA ligase reactions, ligations were performed at 25°C, 30°C, and 37°C for 30 minutes.

Following the ligation reaction incubations at the indicated temperatures, each of the 100 µl ligation mixtures was diluted with 900 µl of TEzero (10 mM Tris pH 7.6, 0.1 mM EDTA) and 2 µl was assayed in quadruplicate by SYBR green qPCR as described above in Section 4.

Results:

The average raw Ct data from each of the qPCR assays was first determined across four wells of each quadruplicate assay. The results of the ligation with consensus templates were measured against a no template control to obtain a set of raw Ct data (data not shown). The scoring scheme of genotyping was then applied to the Ct data as described in Example 3.

TABLES 19 to 22 below show the genotyping results for all of the genotyping assays described in this Example.

"HC" stands for "homozygous consensus" genotyping calls, and is calculated as the Ct(variant)-Ct(consensus) for reactions with the consensus templates.

"HET" stands for "heterozygous" genotyping calls, and is calculated as the Ct(variant) for the variant template minus the Ct(consensus) for the consensus template.

"HV" stands for "homozygous variant" genotyping calls, and is calculated as the Ct(variant) - Ct(consensus) for reactions with the variant templates.

The symbol "Δ" represents the overall dynamic range of each assay set, which is calculated as the absolute value of "HC" - "HV."

The average values across the eight assays are shown for each condition in bold at the bottom of each table.

TABLE 19: HISTORICAL DATA (COUPLED ANNEALING/LIGATION REACTION WITH TAQ DNA LIGASE AS DESCRIBED IN EXAMPLE 3)

	HC	**HET**	**HV**	Δ
SNV 1	1	-2	-5	6
SNV 2	6	-2	-5	11
SNV 3	2	0	-6	8
SNV 4	4	1	-6	10
SNV 5	3	-2	-9	12
SNV 6	5	1	-2	7
SNV 7	2	-1	-4	6
SNV 8	5	0	-3	8
Average	**3**	**-1**	**-5**	**9**

TABLE 20: T4 DNA LIGASE REACTIONS AT 25°C, 30°C AND 37°C (SEPARATE ANNEALING, LIGATION STEPS)

	25°C ligation				**30°C ligation**				**37°C ligation**			
	HC	**HET**	**HV**	Δ	**HC**	**HET**	**HV**	Δ	**HC**	**HET**	**HV**	Δ
SNV 1	3	1	-2	4	3	1	-2	5	3	0	-2	5
SNV 2	2	0	-2	4	2	0	-2	4	2	0	-2	4
SNV 3	2	0	-3	5	2	0	-3	5	2	0	-3	5
SNV 4	3	2	-3	6	3	2	-3	6	4	2	-3	7
SNV 5	-1	-1	-4	3	0	-1	-4	4	1	-1	-3	4
SNV 6	3	1	-1	4	4	1	-1	4	4	1	-1	5
SNV 7	3	2	2	1	3	2	1	2	3	1	0	3
SNV 8	4	1	-2	6	4	1	-2	6	4	1	-3	6
Average	**2**	**1**	**-2**	**4**	**3**	**1**	**-2**	**5**	**3**	**0**	**-2**	**5**

TABLE 21: TAQ DNA LIGASE REACTIONS AT 37°C, 55°C AND 65°C (SEPARATE ANNEALING, LIGATION STEPS)

	37°C ligation				**55°C ligation**				**65°C ligation**			
	HC	**HET**	**HV**	Δ	**HC**	**HET**	**HV**	Δ	**HC**	**HET**	**HV**	Δ
SNV 1	6	0	-8	14	7	0	-8	14	6	0	-8	14
SNV 2	5	-1	-5	11	6	0	-6	12	6	0	-5	11
SNV 3	2	1	-6	8	2	1	-7	9	3	1	-6	8

	37°C ligation				55°C ligation				65°C ligation			
	HC	**HET**	**HV**	**Δ**	**HC**	**HET**	**HV**	**Δ**	**HC**	**HET**	**HV**	**Δ**
SNV 4	7	2	-4	11	8	2	-4	12	7	2	-4	11
SNV 5	7	-1	-11	18	7	-1	-12	19	7	0	-10	18
SNV 6	6	0	-5	12	7	0	-6	13	7	0	-6	13
SNV 7	2	-1	-6	8	2	-1	-7	9	2	-1	-6	8
SNV 8	7	1	-8	15	7	1	-8	15	7	1	-7	14
Average	**5**	**0**	**-7**	**12**	**6**	**0**	**-7**	**13**	**6**	**0**	**-7**	**12**

TABLE 22: TAQ DNA LIGASE REACTIONS AT 45°C AFTER RAPID ANNEALING (65 MINUTES) OR AFTER STANDARD ANNEALING (170 MINUTES) (SEPARATE ANNEALING, LIGATION STEPS)

	45°C ligation after standard annealing				**45°C ligation after rapid annealing**			
	HC	**HET**	**HV**	**Δ**	**HC**	**HET**	**HV**	**Δ**
SNV 1	7	0	-7	14	6	0	-8	14
SNV 2	6	-1	-6	12	6	-1	-5	11
SNV 3	2	1	-5	7	2	1	-6	8
SNV 4	8	2	-3	11	7	1	-4	11
SNV 5	7	-1	-11	18	6	-1	-11	17
SNV 6	8	0	-5	12	7	0	-5	12
SNV 7	2	-1	-7	9	2	-1	-7	9
SNV 8	8	1	-8	15	6	1	-8	14
Average	**6**	**0**	**-6**	**12**	**5**	**0**	**-7**	**12**

Discussion of Results:

Based on the results shown above in TABLES 19 and 20, the ligation-dependent genotyping assays carried out with T4 DNA ligase do not perform as well as those carried out with Taq DNA ligase. It is noted that the greater the Ct spreads between measurements of consensus versus variant genotypes, the better the accuracy in assigning genotypes. In this regard, the dynamic ranges of Taq ligated assays was far greater (i.e., average Δ value of 9) as compared to the dynamic range of the T4 DNA ligase assays (i.e., average Δ value of 4 to 5). It was determined, based on analysis of the raw Ct values, that the reason for this difference in dynamic

range is due to the fact that T4 ligase has a tendency to ligate mismatched oligos, therefore the background in the T4 ligase based assay is worse than in the Taq ligase based assay.

Importantly, as shown above in TABLES 19, 21, and 22, it was observed that the genotyping assays carried out with an annealing reaction followed by separate Taq DNA ligase reaction performed better than the coupled annealing/ligation assays with Taq DNA ligase at all ligation temperatures tested. For example, the average dynamic range of the coupled annealing/ligation genotyping assay with Taq DNA ligase had a dynamic range Δ value of 9, whereas the average dynamic range of the uncoupled assay (i.e., separate annealing and ligation steps) with Taq DNA ligase was increased (e.g., 37°C = Δ value of 12; 45°C = Δ value of 12; 55°C = Δ value of 13; and 65°C = Δ value of 12). Also, the distance between each of the genotyping calls (HC, HET, HV) was greater for the uncoupled Taq DNA ligase assays (e.g., average value for 37°C assay of 5, 0, -7, respectively), as compared to the distance between each genotyping call for the coupled Taq DNA ligase assays (e.g., average value of 3, 1, -5, respectively).

As shown in TABLES 20 and 21, the genotyping assays carried out with Taq DNA ligase under the various ligation temperatures tested in an uncoupled genotyping assay appear to be more or less equivalent. Therefore, a 45°C ligation temperature with Taq DNA ligase in an uncoupled annealing and ligation reaction was chosen for future experiments.

TABLE 22 shows the results of the comparison of a rapid annealing time (65 minutes total) to a standard annealing time (170 minutes) in an uncoupled genotyping assay with the ligation step carried out with Taq DNA ligase at 45°C. As shown in TABLE 22, the results are more or less equivalent, with the same dynamic range (Δ value of 12), and a good distance between each genotyping call (HC, HET, HV) for the rapid annealing assay (i.e., average value of 5, 0, -7, respectively), as compared to the distance between each genotyping call for the assay with the longer annealing time (i.e., average value of 6, 0, -6 respectively). These results demonstrate that oligonucleotide annealing times can be shortened from 170 minutes to 65 minutes or less, and the shorter annealing times were used in all subsequent experiments.

Therefore, based on the above results, it was concluded that the decoupled annealing and ligation reaction generally improved the results of the genotyping assays as compared to the coupled annealing/ligation reaction. In particular, it was observed that the optimal conditions for the ligation-dependent genotyping assay involved a rapid annealing step (approximately 60 minutes), followed by ligation with Taq DNA ligase at 45°C.

C. Determination of the effect of ligase enzyme concentration and incubation time on the performance of the ligation-dependent genotyping assay

In this series of experiments, the variables of Taq DNA ligase enzyme concentration and time of ligation were measured with respect to the genotyping assay performance. In order to determine the minimum ligase concentration required and the influence of time on ligation efficiency, the set of eight SNV query oligos described above in TABLE 15 were assayed against the consensus templates shown in TABLE 14 in a first experiment and the same query reagents were assayed in a second experiment with the variant templates shown in TABLE 14. The genotyping assays were carried out with the rapid annealing protocol followed by ligation with Taq DNA ligase at 45°C.

Annealing Reaction

For each assay reaction, the following reagents were combined:

10 µl of 10 pM pooled templates (variant or consensus)

2 µl of 500 nM pooled consensus and variant 5' ligation primers

2 µl of 250 nM kinased 3' common ligation primers

2 µl of 5 M NaCl

16 µl total volume

Annealing temperatures:

The rapid annealing protocol was carried out as follows:

95°C for 5 minutes;

75°C for 15 minutes;

70°C for 15 minutes;

65°C for 30 minutes;

4°C rest.

Taq DNA Ligation Reactions

A ligation mix "cocktail" was prepared containing:

10 µl of 10X Taq DNA ligase buffer (NEB)

74 µl of H_2O

2 µl, 1 µl, 0.5 µl, 0.1 µl, or 0.02 µl of Taq DNA ligase (40 U/µl, NEB)

85 µl total, which was added to each annealed reaction mixture (16 µl) for a total volume of 100 µl in each ligation reaction.

The ligation reactions were incubated at 45°C for 30 minutes, 20 minutes, 10 minutes, 5 minutes, or 1 minute. The ligation reactions were terminated by the addition of 900 µl of TE, and 2 µl of each ligation reaction was assayed in quadruplicate 10 µl qPCR reactions as described above in Section 4.

Results:

The average raw Ct data from each of the qPCR assays was first determined across four wells of each quadruplicate assays. The results of the ligation with consensus templates were measured against a no template control to obtain a set of raw Ct data (data not shown). The scoring scheme of genotyping was then applied to the Ct data as described in Example 3. The results are shown below in TABLE 23 and TABLE 24.

TABLE 23: RESULTS OF THE LIGATION REACTIONS CARRIED OUT WITH VARIOUS TIME AND ENZYME CONCENTRATIONS ON THE PERFORMANCE OF THE GENOTYPING ASSAYS USING THE CONSENSUS TEMPLATES. UNLESS OTHERWISE INDICATED, VALUES SHOWN ARE (CT(VARIANT)-CT(CONSENSUS).

	30 min ligation			**20 min ligation**			**10 min ligation**			**5 min ligation**		
ligase	2µl	1µl	0.5µl	2µl	1µl	0.5µl	2µl	1µl	0.5µl	2µl	1µl	0.5µl
SNV 1	8	9	9	8	9	8	7	9	8	8	9	8
SNV 2	7	6	7	6	6	5	6	6	6	6	6	6
SNV 3	2	2	2	2	2	2	2	2	2	2	2	3
SNV 4	8	7	7	8	7	8	7	7	7	7	8	8
SNV 5	8	8	9	8	7	7	7	7	7	7	8	8
SNV 6	9	8	8	8	8	8	8	8	9	9	7	8
SNV 7	2	3	3	2	2	2	2	3	3	2	2	3
SNV 8	7	7	7	7	6	6	7	6	7	6	7	7
Average Ct (cons)	**25**	**25**	**25**	**25**	**25**	**26**	**25**	**26**	**26**	**25**	**26**	**27**
Average Ct (var)	**31**	**31**	**32**	**31**	**31**	**31**	**31**	**32**	**32**	**31**	**32**	**33**
Average Ct (var)-Ct(cons)	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**

TABLE 24: RESULTS OF THE LIGATION REACTIONS CARRIED OUT WITH VARIOUS TIME AND ENZYME CONCENTRATIONS ON THE PERFORMANCE OF THE GENOTYPING ASSAYS USING THE VARIANT TEMPLATES.
UNLESS OTHERWISE INDICATED, VALUES SHOWN ARE (CT(CONSENSUS)-CT(VARIANT).

	5 min ligation			**1 min ligation**		
ligase	0.5µl	0.1µl	0.02µl	0.5µl	0.1µl	0.02µl
SNV 1	8	7	7	8	8	7
SNV 2	5	4	2	5	5	4
SNV 3	6	8	5	7	6	6
SNV 4	5	5	6	4	4	3
SNV 5	9	9	8	10	8	8
SNV 6	5	6	8	5	7	7
SNV 7	8	7	6	7	7	5
SNV 8	9	8	11	9	7	7
Average Ct (var)	**26**	**28**	**31**	**26**	**28**	**31**
Average Ct (cons)	**33**	**35**	**37**	**33**	**35**	**36**
Average Ct (cons)-Ct(var)	**7**	**7**	**7**	**7**	**6**	**6**

Discussion of Results:

As shown above in TABLE 23 and TABLE 24, the results of the genotyping assay with a ligation reaction carried out for 5 minutes is about equivalent to the results of the genotyping assay with a ligation reaction carried out for longer periods of time (i.e., 10, 20, or 30 minutes), both in terms of Ct(variant)-Ct(consensus) differences and with respect to the absolute Ct values for cognate versus mismatched templates.

As further shown in TABLE 23 and TABLE 24, low concentrations (0.5 µl to 1 µl of 40 U/µl) of Taq DNA ligase appear adequate for driving ligation to the same levels as observed with greater amounts of Taq DNA ligase enzyme.

Therefore, based on the above results, it was determined that the optimal conditions for the 100 µl ligation reaction in the ligation-dependent genotyping assay includes the use of a rapid

annealing step (approximately 60 minutes), followed by ligation with Taq DNA ligase at a concentration of from about 0.5 µl to about 1.0 µl of 40 U/µl for 5 minutes at 45°C.

EXAMPLE 6

This Example describes the manufacture of a 576-feature matrix of minimally interacting pairs of detection primers (also referred to as a "universal PCR decoding matrix") for use in decoding a multiplex assay, such as a multiplex ligation-dependent genotyping assay for genotyping a test sample at a plurality of SNV positions of interest.

PCR Primer Matrix Design

The goal of this Example was to design a matrix of minimally interacting primer pairs to manufacture a 576-feature matrix of detection primer pairs.

Rationale:

Since adenine residues cannot base pair with cytosine residues, these sequences were used as trinucleotides on the 3' ends of primers as the basis of a minimally interactive, non-primer-dimer forming primer matrix. Specifically, one set of 36 potential primers was designed to end in "CCC," and a second set of 36 primers was designed to end in "AAA" at each of their 3' ends.

Candidate 25 mer PCR primer sequences were chosen in the following way.

First, a 10,000 list of random 22-mer DNA sequences was generated. The only criterion was that these sequences were required to have at least four of each type of DNA base (A, G, C, T).

A list of 200 of the 10,000 sequences were chosen at random and screened for the presence of either "TTT" or "GGG" in the 3' terminal 6 nucleotides, which were then removed from the list of candidate primers. The rationale for removal of these primers is that "TTT" can pair with "AAA" and "GGG" can pair with "CCC," therefore, primers with these 3' terminal sequences would be susceptible to primer-dimer formation. Approximately 15% of the randomly selected sequences were removed from the list of candidate PCR primers via this filtering process, leaving a total of 170 candidate sequences.

72 of the 170 remaining candidate sequences were randomly chosen as the candidate PCR primer sequences. The 3' terminal sequence of "CCC" was added to the first set of 36 of these sequences ("row primers"), and the 3' terminal sequence of "AAA" was added to the second set of 36 of these sequences ("column primers"), thereby creating a 36 by 36 primer matrix, as shown below in TABLE 25.

TABLE 25: SET OF 36 "CCC" AND 36 "AAA" PRIMERS

Reference Number	Sequence (5' to 3')	SEQ ID NO:	Used in final matrix
AAA1	GATCTGGCTAGGTGCCACAACAAAA	434	+
AAA2	GACATGCTAACCACGTTGCAGGAAA	435	+
AAA3	GACCTCGTAAAAGGGGGTATAGAAA	436	-
AAA4	AAAATACCATCTTGGCCATTATAAA	437	+
AAA5	GAGTGACTGCAACTAAAATGCTAAA	438	+
AAA6	TGTATCAGAGGATTGCGTTCGAAAA	439	+
AAA7	GTTCGGGGATACATTCTGAGTAAAA	440	-
AAA8	TGCAACTAGATTGAGGCCTCTAAAA	441	+
AAA9	CTATATGTAGGGGCTCTAACCGAAA	442	-
AAA10	CATCTGCTGCGTTTGGAATACGAAA	443	+
AAA11	ATACCAGCCGGCTGATGATCGTAAA	444	+
AAA12	AGCCACTCTGTAGCACTGATGGAAA	445	+
AAA13	TACCCTAGTTGGCAGTTCATCGAAA	446	+
AAA14	ATAATAGTCGCTGGTATGGTACAAA	447	+
AAA15	ATTTGGAACACCGCAGCTCGGTAAA	448	+
AAA16	GACCCCGTGCACGGATGCATGAAAA	449	+
AAA17	GTCGGGCAGCACCCAAGTTCTGAAA	450	+
AAA18	AGCTGTGGTTAAGGATAGTTCGAAA	451	+
AAA19	AGTGCAAATTCGACACTTGACGAAA	452	+
AAA20	GGCCCTCCTTATTAAACATCCGAAA	453	+
AAA21	ACTCACTCTGGGCAGACGCAGAAAA	454	-
AAA22	TTCGGGCGTTCTGAAGACCTGTAAA	455	-
AAA23	CCGGGGGAGTCATTGTATTACGAAA	456	-
AAA24	GTAGACCGTAGCGAACACCGGAAAA	457	+
AAA25	AGTCTCGGTTCCGCATGCGTCGAAA	458	-
AAA26	CATACCGTCAACTAATATTCTCAAA	459	+
AAA27	GTGGGATGGAGTCCACGAAATTAAA	460	+

Reference Number	Sequence (5' to 3')	SEQ ID NO:	Used in final matrix
AAA28	TTGGAGTTTAGCGACACGCATTAAA	461	+
AAA29	ATCTATCTTGAACCCGGGCGATAAA	462	+
AAA30	GGCACTCGGGTCTTATCCGTTGAAA	463	-
AAA31	GCTTATACGCAACTGTGTCTGGAAA	464	-
AAA32	CAAAAGAGGTTGTCGTAGCTCGAAA	465	+
AAA33	GTGGATGTCCAGGTTAACTCAAAAA	466	+
AAA34	AAGGTGCTTGAGCCATGGGATCAAA	467	-
AAA35	GTAACTTCATACACTCCACATTAAA	468	+
AAA36	TTTCGTGCAAGTCAACAATTGAAAA	469	+
CCC1	GCATGAGGCCCTGATGCAGTGACCC	470	+
CCC2	AACGGTGATGTCGTCAAAGATTCCC	471	-
CCC3	CAGACATCTCCTAGCGAGTCAGCCC	472	-
CCC4	ATTGGTGTCTCCCCGAGCTGTACCC	473	+
CCC5	TCGCCATATCGTACCGATGTCTCCC	474	-
CCC6	CGTAGACTAACCGACTCATCGACCC	475	+
CCC7	ACTGACCGTTTAAGGGTCCAAGCCC	476	+
CCC8	TACGTTCACCATCGTCAATAGGCCC	477	+
CCC9	GTCGCCACGAACGCTGAAGAAGCCC	478	+
CCC10	GCTGCACGTTGTCTCACAGCTTCCC	479	-
CCC11	GCTACGCGTCCTCCAATATGCGCCC	480	+
CCC12	GAGTAGGGTAATACGTTCTACACCC	481	-
CCC13	GAACCCTTTAGCTCCACAATTGCCC	482	-
CCC14	AATAACGCATGCGTTATCCCACCCC	483	+
CCC15	AATGATCAACGAACGTCGCTGGCCC	484	+
CCC16	TATAGCAATGAGGGCCAGTGATCCC	485	+
CCC17	TAGCTAAGCTTGTGCTAGATTACCC	486	+
CCC18	ACGGCGTCAGTTGTAAGGATATCCC	487	+
CCC19	TATGATAACCCACTTCCAAGTTCCC	488	-
CCC20	CCATAACCTTAGTATGTAGTCGCCC	489	+

Reference Number	Sequence (5' to 3')	SEQ ID NO:	Used in final matrix
CCC21	CGTCTGTGGCAATAACGCTTCACCC	490	+
CCC22	TATGCTTCCTGGAGCTGCAAGCCCC	491	+
CCC23	CGGCATTCTGAACAACTATATGCCC	492	+
CCC24	CGCATCTGCACGTAAAACGGCGCCC	493	-
CCC25	TCAGGGCTACGCGACCTCGTACCCC	494	+
CCC26	ATGCCGAGATTCGAATATCGGACCC	495	+
CCC27	CCAAATTCCGCGGGCCTTGAACCCC	496	+
CCC28	ACTTGCGTACCCATACATGTATCCC	497	-
CCC29	TAGAAGCGCGAAGTATAGGATGCCC	498	-
CCC30	TAGTACCGGCAATTCCTTGTTGCCC	499	+
CCC31	AACCACGAGTCGTCACTGACCGCCC	500	+
CCC32	GTAAATGGTCTAGAGGTTACGGCCC	501	-
CCC33	CGTCGGATTGTGCTATGTAAAACCC	502	+
CCC34	GACAGTTCATCTACACATTGCACCC	503	+
CCC35	AAGGGAACCGGCACGAATCAGTCCC	504	+
CCC36	TCATTGCTAGCACCTACCAGACCCC	505	-

Screening for minimally interacting primers for use in the 24 x 24 primer matrix

The 72 candidate PCR primers shown above in TABLE 25 were screened as described below in order to identify a subset of 24 column primers and 24 row primers that would collectively define a primer matrix with low levels of primer-dimer formation.

The 72 candidate PCR primers for use in a primer matrix were resuspended to a working concentration of 10 µM in water. A grid of 36 by 36 wells containing PCR master mix was prepared by aliquoting 25 µl of 2X power SYBR master mix (Applied Biosystems, Foster City, CA) and 17 µl of water in each well of a set of 384 well optical PCR plates as follows.

4 µl of column primers ("AAA":SEQ ID NOS:434-469) were added to each well along each column and 4 µl of row primers ("CCC":SEQ ID NOS:470-505) were added to each well along each row. Following mixing, four 10 µl aliquots from each well were distributed in quadruplicate into 384 optical PCR plates as shown below in TABLE 26 and analyzed for qPCR using 40 cycles on an ABI 7900 qPCR instrument following the standard cycling protocol. The

average Ct of each set of quadruplicate wells was then calculated and wells with Ct values of less than 38 were identified (as shown in bold in TABLE 26) as wells in which primer pairs were able to interact to form detectable PCR products at unacceptably low Cts. The primer pairs with Ct values > 38 are indicated with a "+" symbol, indicating that the primer pairs are useful for inclusion in the final matrix.

TABLE 26: PRIMER PAIRS THAT PRODUCED CT VALUES LESS THAN 38 ARE SHOWN IN BOLD

Ct values	AAA1	AAA2	AAA3	AAA4	AAA5	AAA6	AAA7	AAA8	AAA9	AAA10	AAA11	AAA12	AAA13	AAA14	AAA15	AAA16	AAA17	AAA18	AAA19	AAA20	AAA21	AAA22	AAA23	AAA24	AAA25	AAA26	AAA27	AAA28	AAA29	AAA30	AAA31	AAA32	AAA33	AAA34	AAA35	AAA36
CCC1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC2	+	**32**	+	+	+	+	+	+	+	+	+	**33**	+	+	+	+	+	+	**36**	+	+	+	+	+	+	+	+	+	**34**	+	+	+	+	+	+	+
CCC3	+	+	+	+	+	+	+	+	+	+	**35**	+	+	+	+	+	**36**	+	+	+	**27**	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC4	+	+	**38**	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC5	+	+	+	+	+	+	**36**	+	+	+	+	+	+	+	+	+	+	+	+	+	**37**	+	**37**	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC6	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	**37**	+	+	+	+	**37**	+	+	+	+	+	+
CCC7	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	**38**	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC8	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	**38**	+	+	+	+	+
CCC9	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	**35**	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC10	+	**25**	+	+	+	+	+	+	+	+	+	**27**	+	+	+	+	**33**	+	+	+	**37**	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC11	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	**37**	+	+	+	+	**38**	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC12	+	+	**35**	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC13	+	+	+	+	**36**	+	+	+	+	+	+	+	+	+	+	+	**37**	**36**	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC14	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC15	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	**34**	+	+	+	**37**	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC16	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	**38**	+	+	**30**	**30**	+	+
CCC17	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC18	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC19	+	**34**	+	+	+	+	+	+	+	+	**35**	+	+	+	+	+	+	+	+	+	+	**36**	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC20	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	**34**	**35**	+	+	+	+	+	+	**26**	+	+	+	+	+	+	+
CCC21	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC22	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	**36**	+	+	+	+	+
CCC23	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC24	+	+	+	+	+	+	**36**	+	**38**	+	**37**	+	+	+	**37**	+	+	+	+	+	**36**	**35**	+	+	+	+	+	+	**34**	+	**36**	+	+	+	+	+
CCC25	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC26	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

AAA36	+	+	+	+	+	+	+	+	+	+
AAA35	+	+	+	+	+	+	+	+	+	+
AAA34	**34**	+	+	+	+	+	+	+	+	+
AAA33	+	+	+	+	+	+	+	+	+	+
AAA32	+	+	+	+	+	+	+	+	+	+
AAA31	+	+	+	+	+	+	**34**	+	+	**38**
AAA30	+	+	+	**37**	**30**	+	+	+	+	+
AAA29	+	+	+	+	+	+	+	+	+	+
AAA28	+	+	+	+	+	+	+	+	+	+
AAA27	+	+	+	+	+	+	+	+	+	+
AAA26	+	+	+	+	+	+	+	+	+	+
AAA25	+	+	+	+	+	+	+	+	+	+
AAA24	+	+	+	+	+	+	+	+	+	+
AAA23	+	+	+	+	+	+	+	+	+	+
AAA22	+	+	**35**	+	**33**	**38**	+	+	+	+
AAA21	+	+	**29**	+	**34**	+	+	+	+	+
AAA20	+	+	+	+	+	+	+	+	+	+
AAA19	+	+	+	+	+	+	+	+	+	+
AAA18	+	+	+	+	+	+	+	+	+	+
AAA17	+	+	**32**	**35**	**36**	**37**	+	+	+	+
AAA16	+	+	+	+	+	+	+	+	+	+
AAA15	+	+	+	+	+	+	+	+	+	+
AAA14	+	+	+	+	+	+	+	+	+	+
AAA13	+	+	+	+	+	+	+	+	+	+
AAA12	+	+	+	+	+	+	+	+	+	+
AAA11	+	+	+	+	+	+	+	+	+	+
AAA10	+	+	+	+	+	+	+	+	+	+
AAA9	+	+	+	+	+	**30**	+	+	+	+
AAA8	+	+	+	+	+	+	+	+	+	+
AAA7	**38**	**30**	+	+	+	+	+	+	+	+
AAA6	+	+	+	+	+	+	+	+	+	+
AAA5	+	+	+	+	+	+	+	+	+	+
AAA4	+	+	+	+	+	+	+	+	+	+
AAA3	+	+	+	+	+	+	+	+	+	**33**
AAA2	+	+	+	+	+	+	+	+	+	+
AAA1	+	+	+	+	+	+	+	+	+	+
Ct values	CCC27	CCC28	CCC29	CCC30	CCC31	CCC32	CCC33	CCC34	CCC35	CCC36

TABLE 27: FINAL CONFIGURATION OF 24 X 24 PRIMER MATRIX"

"+" symbol indicates Ct value is 38 or higher

	AAA1	AAA2	AAA4	AAA5	AAA6	AAA8	AAA10	AAA11	AAA12	AAA13	AAA15	AAA16	AAA18	AAA19	AAA20	AAA24	AAA26	AAA27	AAA28	AAA29	AAA32	AAA33	AAA35	AAA36
CCC1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC6	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC7	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC8	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC9	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC11	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC14	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC15	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC16	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC17	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC18	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC20	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC21	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC22	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC23	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC25	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC26	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC27	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC30	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC31	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC33	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC34	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CCC35	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

As indicated above in TABLE 27, the final 24 primer by 24 primer matrix used for the qPCR amplification of the ligation-dependent genotyping assay carries no primer pairs that produced a Ct value of less than 38, and therefore all the primer pairs contained in this primer matrix are minimally interacting primer pairs suitable for use in the genotyping assays described herein.

The 24 by 24 primer grid provides 576 unique primer pairs (i.e., features) that can be used to perform consensus versus variant genotyping assays on 288 putative SNV positions (288 consensus plus 288 variant assays = 576 PCR reactions). Therefore, the matrix can be used with sets of 288 assays, as demonstrated below in EXAMPLE 7.

EXAMPLE 7

This Example demonstrates the use of the 24 by 24 primer matrix described in Example 6 for use in the ligation-dependent genotyping assay for genotyping 799 putative SNV locations identified during DNA sequencing of 14 *Pichia pastoris* yeast strains.

Rationale: High throughput sequencing of 14 *Pichia pastoris* yeast strains indicated that as many as 799 SNVs that differed from the *Pichia pastoris* reference sequence may be present in one or more strains that were examined. In order to further examine these putative SNV locations, we generated 799 consensus and variant genotyping assays with synthetic consensus and variant DNA templates.

Methods:

1. Preparation of assay oligos:

A set of 799 genotyping reagents was generated for the 799 SNV positions of interest, including 5' ligation oligos (consensus and variant), 3' common ligation oligos and synthetic consensus and variant templates for each SNV position of interest, using the same design criteria as described above in Example 5 (oligo sequences not shown).

2. Pooling of oligos:

To perform the genotyping assays, the 799 genotyping oligos were divided into two sets of 288 assays and one set of the remaining 223 assays.

For each set of assays (i.e., the first set of 288 assays, the second set of 288 assays and the third set of 223 assays), consensus and variant 5' ligation oligos were pooled and diluted to 500 nM (860 pM in each unique oligo).

Similarly, the common 3' ligation oligos for each set of assays was pooled, treated with kinase to add a 5' terminal phosphate as described in Example 5, and diluted to a final working concentration of 250 nM (860 pM in individual oligos).

Finally, for each set of assays, pools of 288 or 233 consensus template oligos, and pools of 288 or 233 variant template oligos were pooled and diluted to 100 pM.

3. Annealing

The ligation-dependent genotyping assays were performed by the decoupled annealing and ligation method, as follows.

Annealing Reaction

For each assay reaction, the following reagents were combined:

10 µl of 100 pM pooled templates (variant or consensus)

10 µl of 500 nM pooled consensus and variant 5' ligation oligos

10 µl of 250 nM kinased 3' common ligation oligos

2 µl of 5 M NaCl

32 µl total volume

Annealing:

The rapid annealing protocol was carried out as follows:

95°C for 5 minutes;

75°C for 15 minutes;

70°C for 15 minutes;

65°C for 30 minutes;

25°C rest.

4. Ligation

For each assay, 68 µl of a ligation cocktail was added to the 32 µl annealed mixture, the ligation cocktail containing:

10 µl of 10X Taq DNA ligase buffer (NEB)

57 µl of H_2O

1 µl of 40 U/µl Taq DNA ligase (NEB)

100 µl total volume

Note: 3 different reaction mixtures were prepared, one for each assay: the first set of 288 assays, the second set of 288 assays and the third set of 223 assays.

The ligation mixtures were incubated at 45°C for 5 minutes and diluted to 1 ml with 900 µl of TEzero (10 mM Tris pH 7.6, 0.1 mM EDTA). Six such identical reactions were run for each

set of 288 consensus or 288 variant assays in order to provide enough material to assay on PCR plates.

5. qPCR Assay

For the qPCR assay readout, 2 µl of the ligation mixture was assayed in a 10 µl reaction volume containing 5 µl of 2X power SYBR master mix (Applied Biosystems, Foster City, CA), 1.4 µl H_2O, 0.8 µl column matrix primer and 0.8 µl row matrix primer.

Each mixture was aliquoted in quadruplicate across a matrix of 24 by 24 separate PCR reactions, as described in Example 6, which translated into 6 independent 384 well optical PCR plates per set of 288 genotyping assays.

The PCR plates were run on an Applied Biosystems 7900 qPCR instrument according to the manufacturer's instructions.

Results:

The average Cts across quadruplicate wells were calculated for each consensus and variant pair of the 799 SNV assay set. It was determined that all of the assays involving the column primer AAA29 (SEQ ID NO:461: 5'ATCTATCTTGAACCCGGGCGATAAA 3') yielded Ct values greater than 35, whereas the standard genotyping Ct readout was below 30 for all the other primer pairs. This indicated that the AAA29 primer (SEQ ID NO:461) does not support robust PCR amplification, and therefore all the assays (32 total assays) using this primer SEQ ID NO:461 were not evaluated further. To avoid inclusion of poor performing primers such as the AAA29 primer, in the future, matrix primer screening, as described in Example 6, will also include a positive test against synthetic templates for functional PCR amplification performance. In the present example, a 36 x 36 matrix of primers was screened, using the methods described in Example 6, and it was determined that only about 5 to 6 row primers and 5 to 6 column primers were poor performers (i.e., high background, low Cts). In the process of choosing primers from this screen for use in the 24 x 24 matrix, many primers were excluded that would fit the criteria of good performers. One of these good but previously excluded primers was substituted for primer SEQ ID NO:461 in the matrix and the assay worked well with the substituted primer (data not shown).

For the remaining 767 assays, the Ct(variant)-Ct(consensus) = Δ consensus for consensus templates and the Ct(consensus)-Ct(variant) = Δ variant for variant templates was calculated. The performance of the ligation-dependent genotyping assay was evaluated based on the sum of Δ consensus + Δ variant. It was empirically determined that if the sum of Δ consensus + Δ variant is greater than 3, then genotyping calls can be made with confidence in diploid organisms. This was

established in separate experiments by genotyping of two inbred mouse strains and their F1 progeny at known SNPs. In this system, the parental strains were uniformly homozygous and the progeny were uniformly heterozygous at every SNP location. A survey of 576 independent SNP assays in this system revealed the greatest accuracy when only the genotyping assays were considered that had a Δ consensus + Δ variant value of greater than 3 (data not shown).

For haploid organisms such as *P. pastoris*, the genotyping results are expected to be even more accurate, because only two genotypes are possible (consensus or variant), in contrast to the case in diploid species where three genotypes are possible (consensus, variant, or heterozygote). Hence, in haploid organisms, the expected genotype will only be consensus or variant, and not potentially a heterozygous blend of the two as is found in a diploid organism such as a human. Therefore, for haploid organisms, the value of Ct(variant) - Ct(consensus) is predicted to resemble either Δ consensus or -Δ variant.

The Δ consensus + Δ variant values for all 767 functional assays were calculated. Of these, 730 (95%) had values greater than or equal to 3, indicating that the genotyping calls can be made with confidence. Of the 37 failed assays (values below 3), it is interesting to note that 19 of them shared overlapping DNA sequences in two groups of 7 assays and 12 assays, respectively. Subsequent in-house comparisons of two independently generated draft genome sequences of *Pichia pastoris* revealed almost perfect identity except in these regions, where the assembled sequences disagreed. While not wishing to be bound by any particular theory, this suggests that these regions are generally difficult to sequence and that the sequences that were genotyped may not exist in *P. pastoris*. If the DNA sequences of the genotyping primers do not match those of the target region, then the genotyping assays would be expected to fail. The remaining 18 failed assays occurred across unrelated sequence groups.

In summary, this Example demonstrates that of the 767 ligation-dependent genotyping assays carried out that were designed to query random SNVs, 95% of the assays returned useful data. This percent of discovered SNVs that can be assayed in a particular technology platform with high confidence, otherwise referred to as "conversion rate" in the genotyping field, is very high and comparable to other commercially available platforms such as the Affymetrix SNP array or the Illumina Bead array.

Unlike commercially available genotyping solutions, which are fixed and can only monitor known SNPS, the ligation-dependent genotyping assays described herein combine the advantages of a highly successful conversion rate and the flexibility to monitor novel single-nucleotide

variants. The ligation-dependent genotyping assays as described herein are therefore a unique, low cost solution to the validation of putative sequence variants that are suggested by high-throughput resequencing technologies.

While the preferred embodiment of the invention has been illustrated and described, it will be appreciated that various changes can be made therein without departing from the spirit and scope of the invention.

What is claimed:

1. A method of determining the genotype of a test sample at one or more polymorphic loci of interest, the method comprising:

a) contacting, in a reaction mixture, one or more set(s) of query oligonucleotides with the test sample having one or more polymorphic loci of interest within one or more target nucleic acid region(s) of interest, wherein each set of query oligonucleotides comprises:

(i) at least one 5' ligation oligonucleotide having, from the 5' to 3' end, a first PCR primer binding region, a target-specific binding region selected to hybridize 5' of a polymorphic locus of interest, and a 3' region chosen to hybridize to either a consensus or variant nucleotide sequence at the polymorphic loci of interest; and

(ii) a phosphorylated 3' ligation oligonucleotide having, from the 5' to 3' end, a target-specific binding region selected to hybridize 3' of the polymorphic loci of interest and a second PCR primer binding region;

under conditions that allow hybridization between the one or more set(s) of query oligonucleotides and the target nucleic acid region(s) of interest such that the 5' ligation oligonucleotides and the phosphorylated 3' ligation oligonucleotides hybridize adjacent to each other on the target nucleic acid region(s) of interest;

b) contacting the reaction mixture of step (a) with a DNA ligase under conditions suitable to ligate the 5' ligation oligonucleotides and the adjacent 3' phosphorylated ligation oligonucleotides, thereby generating a plurality of ligation products indicative of the genotype of the test sample at the one or more polymorphic loci of interest; and

c) measuring the amount of the plurality of ligation products in the reaction mixture of step (b).

2. The method of Claim 1, further comprising the step of comparing the amount of plurality of ligation products measured according to step (c) with at least one reference standard that is indicative of the presence or absence of the consensus or variant nucleotide at each polymorphic loci of interest.

3. The method of Claim 1, wherein in the reaction mixture of step a) the test sample is contacted with the one or more set(s) of query oligonucleotides and the DNA ligase under conditions that allow hybridization between the one or more set(s) of query oligonucleotides and the target nucleic acid region(s) of interest and to allow the query oligonucleotides to hybridize adjacent to each other on the target nucleic acid region of interest, and to allow ligation of the 5' ligation oligonucleotide and the adjacent 3' phosphorylated ligation oligonucleotides, thereby generating a plurality of ligation products, so as to couple hybridization and ligation in the reaction mixture.

4. The method of Claim 1, wherein the test sample comprises a haploid or diploid genome.

5. The method of Claim 1, wherein the test sample comprises non-amplified target nucleic acid region(s) of interest.

6. The method of Claim 1, wherein each set of query oligonucleotides according to step (a) comprises a pair of allele-specific 5' ligation oligonucleotides for each polymorphic loci of interest, the pair comprising (i) a first 5' ligation oligonucleotide comprising a 3' region chosen to hybridize to a consensus nucleotide sequence at the polymorphic loci of interest and (ii) a second 5' ligation oligonucleotide comprising a 3' region chosen to hybridize to a variant nucleotide sequence at the polymorphic loci of interest.

7. The method of Claim 1, wherein the 5' ligation oligonucleotides comprise the first PCR primer binding region having different nucleotide sequences.

8. The method of Claim 1, wherein step (a) comprises contacting, in the single reaction mixture, the test sample with at least 10 sets of query oligonucleotides for genotyping at least 10 different polymorphic loci positions of interest.

9. The method of Claim 1, wherein the DNA ligase is thermostable.

10. The method of Claim 1, wherein the measuring in step (c) comprises amplifying the plurality of ligation products with one or more pair(s) of detection primers, each detection primer pair having (i) a forward PCR primer that binds to the first PCR primer binding region in the 5' ligation oligonucleotide and (ii) a reverse PCR primer that binds to the second PCR primer binding region in the 3' ligation oligonucleotide.

11. The method of Claim 1, wherein the measuring in step (c) comprises amplifying the plurality ligation products with: (i) a first pair of detection primers having a forward PCR primer that binds to the PCR binding region of the 5' ligation oligonucleotide comprising the consensus binding region; and with (ii) a second pair of detection primers comprising a forward PCR primer that binds to the PCR binding region of the 5' ligation oligonucleotide comprising the variant binding region.

12. The method of claim 10, wherein the penultimate 2 or 3 nucleotides at the 3' end of the first pair or second pair of detection primers are selected to reduce primer-dimer formation by selecting 2 or 3 nucleotide that reduce annealing between the first and second pair of detection primers or that reduce self-annealing of the first and second pair of detection primers.

13. The method of claim 1, wherein the measuring in step (c) comprises measuring fluorescence.

14. The method of claim 1, wherein the measuring in step (c) includes contacting the ligation product with a dye that intercalates double-stranded DNA.

15. The method of claim 10, wherein the one or more pair(s) of detection primers comprise a fluorescent label.

16. The method of claim 11, wherein the first or second pair of detection primers comprise a fluorescent label.

17. The method of Claim 1, wherein the test sample is enriched for the one or more target region(s) of interest prior to the contacting of step (a).

18. The method of Claim 1, wherein the query oligonucleotides each have a length of about 40 nucleotides to about 200 nucleotides.

19. The method of Claim 1, wherein the target-specific binding region of the query oligonucleotides have a length of about 10 nucleotides to about 150 nucleotides in length.

20. A method of genotyping a test sample at one or more single nucleotide variant(s) (SNVs) position(s) of interest, the method comprising: for each SNV position of interest,

a) contacting in three separate reaction mixtures:

(i) a synthetic template comprising the target region of interest having a consensus nucleotide at the SNV position of interest;

(ii) a synthetic template comprising the target region of interest having a variant nucleotide at the SNV position of interest; and

(iii) a test sample comprising the target region of interest comprising the SNV position of interest to be genotyped;

with one or more set(s) of SNV query oligonucleotides, each set comprising (i) a pair of allele-specific 5' ligation oligonucleotides, the pair comprising a first 5' ligation oligonucleotide comprising, from the 5' to 3' end, a first PCR primer binding region, a target-specific binding region selected to hybridize 5' of the SNV nucleotide position of interest, and a 3' region chosen to hybridize to the consensus nucleotide sequence at the SNV position of interest and a second 5' ligation oligonucleotide comprising, from the 5' to 3' end, a first PCR primer binding region, a target-specific binding region selected to hybridize 5' of the SNV nucleotide position of interest, and a 3' region chosen to hybridize to the variant nucleotide sequence at the SNV position of interest and (ii) a phosphorylated 3' ligation oligonucleotide comprising from the 5' to 3' end, a target-specific binding region selected to hybridize 3' of the SNV position of interest and a second PCR primer binding region, under conditions that allow hybridization between the one or more sets of SNV query oligonucleotides and the target regions of interest having the consensus nucleotide, the variant nucleotide, and the SNV position of interest, such that the 5' ligation oligonucleotides and the phosphorylated 3' ligation oligonucleotides hybridize adjacent to each other on the target region of interest;

b) contacting the three separate reaction mixtures of step (a) with a DNA ligase under conditions suitable to ligate the 5' ligation oligonucleotides and the adjacent 3' phosphorylated ligation oligonucleotides, thereby generating three separate mixtures each having a plurality of ligation products; and

c) measuring the amount of the plurality of ligation products in each of the three separate mixtures of step (b).

21. The method of Claim 20, wherein in at least one of the three separate reaction mixtures the of step a) the test sample is contacted with the one or more set(s) of query

oligonucleotides and the DNA ligase under conditions that allow hybridization between the one or more set(s) of query oligonucleotides and the target nucleic acid region(s) of interest to allow the query oligonucleotides to hybridize adjacent to each other on the target nucleic acid region of interest, and to allow ligation of the 5' ligation oligonucleotide and the adjacent 3' phosphorylated ligation oligonucleotides, thereby generating a plurality of ligation products.

22. The method of Claim 20, wherein the test sample comprises a haploid or diploid genome.

23. The method of Claim 20, wherein the test sample comprises a non-amplified target region of interest.

24. The method of Claim 20, wherein the first 5' ligation oligonucleotides comprise the first PCR primer binding regions having different nucleotide sequences.

25. The method of Claim 20, wherein step (a) comprises contacting the test sample with at least 10 sets of SNV query oligonucleotides for genotyping at least 10 different SNV positions of interest.

26. The method of Claim 20, wherein the DNA ligase is thermostable.

27. The method of Claim 20, wherein the measuring in step (c) comprises amplifying the plurality ligation products with (i) a set of detection primers comprising forward PCR primers that bind to the first PCR binding region of the first 5' ligation oligonucleotide comprising the consensus binding region, (ii) a set of detection primers comprising forward PCR primers that bind to the first PCR binding region of the second 5' ligation oligonucleotide comprising the variant binding region, and (iii) a set of detection primers comprising reverse PCR primers that bind to the second PCR primer binding region in the 3' ligation oligonucleotide.

28. The method of claim 27, wherein the penultimate 2 or 3 nucleotides at the 3' end of the forward or reverse PCR primers are selected to reduce primer-dimer formation by selecting 2 or 3 nucleotide that reduce annealing between the first and second pair of detection primers or that reduce self-annealing of the first and second pair of detection primers.

29. The method of claim 20, wherein the measuring in step (c) comprises measuring fluorescence.

30. The method of claim 20, wherein the measuring in step (c) comprises contacting the plurality of ligation products with a dye that intercalates double-stranded DNA.

31. The method of claim 27, wherein the forward PCR primer or the reverse PCR primer comprises a fluorescent label.

32. The method of Claim 20, wherein the test sample is enriched for the one or more target region(s) of interest prior to the contacting in step (a).

33. The method of Claim 20, wherein the SNV query oligonucleotides have a length of about 40 nucleotides to about 200 nucleotides.

34. The method of Claim 20, wherein the target-specific binding region of the SNV query oligonucleotides have a length of about 10 nucleotides to about 150 nucleotides in length.

35. A two-dimensional nucleic acid matrix comprising forward and reverse primer pairs and ligation products distributed into positionally addressable wells, wherein the wells include:

a) the forward PCR primers each having
   (i) a 5' region that hybridizes to a 5' primer binding region of a target nucleic acid molecule of interest and
   (ii) a 3' region selected to avoid primer-dimer formation with the reverse primer

b) the reverse PCR primers each having
   (i) a 5' region that hybridizes to a 3' primer binding region of the target nucleic acid molecule of interest and
   (ii) a 3' region selected to avoid primer-dimer formation with the forward primer; and

c) ligation products generated by annealing the target nucleic acid molecule of interest with
   (i) a 5' ligation oligonucleotide having from the 5' to 3' end, the reverse PCR primer binding region, a target-specific binding region selected to

hybridize 5' of a polymorphic locus of interest, and a 3' region chosen to hybridize to either a consensus or variant nucleotide sequence at the polymorphic locus of interest and

(ii) an adjacent phosphorylated 3' ligation oligonucleotide having from the 5' to 3' end, a target-specific binding region selected to hybridize 3' of the polymorphic locus of interest and a forward PCR primer binding region and

(iii) ligating the 5' ligation oligonucleotides and the adjacent 3' phosphorylated ligation oligonucleotides so as to generate the ligation products.

36. The matrix of Claim 35, wherein the 5' ligation oligonucleotides comprise the reverse PCR primer binding region having different sequences.

37. The matrix of Claim 35, wherein the penultimate 2 or 3 nucleotides of the 3' region in the forward and reverse PCR primers are selected to reduce primer-dimer formation by selecting 2 or 3 nucleotide that reduce annealing between the first and second pair of detection primers or that reduce self-annealing of the first and second pair of detection primers.

38. The matrix of Claim 37, wherein the 3' region selected to avoid primer-dimer formation in the forward PCR primers comprises the nucleotide sequence "CT" and the 3' region selected to avoid primer-dimer formation in the reverse primers comprises the nucleotide sequence "GA."

39. The matrix of Claim 37, wherein the 3' region selected to avoid primer-dimer formation in the forward PCR primers comprises the nucleotide sequence "ACA" and the 3' region selected to avoid primer-dimer formation in the reverse primers comprises of the nucleotide sequence "CAC."

40. The matrix of Claim 37, wherein the 3' region selected to avoid primer-dimer formation excludes "TTT" and "GGG" sequences.

41. The matrix of Claim 37, wherein the 3' region selected to avoid primer-dimer formation in the forward PCR primer comprises a terminal sequence of "CCC" and the 3' region

selected to avoid primer-dimer formation in the reverse PCR primer comprises a terminal sequence of terminal sequence of "AAA".

42. The matrix of Claim 37, wherein the 3' region selected to avoid primer-dimer formation in the forward PCR primer comprises a terminal sequence of "AAA" and the 3' region selected to avoid primer-dimer formation in the reverse PCR primer comprises a terminal sequence of terminal sequence of "CCC".

43. The matrix of Claim 37, wherein the last nine nucleotides of the forward and reverse PCR primer sequences are selected to exclude the sequence "ACA" or "TGT."

44. The matrix of Claim 35, wherein the total length of the forward and reverse PCR primers comprises about 15 to 35 nucleotides.

45. The matrix of Claim 35, wherein the 3' region selected to avoid primer-dimer formation in the forward and reverse PCR primers comprises 6 nucleotides.

46. The matrix of Claim 35, further comprising an enzyme reaction mixture for PCR amplification.

47. A kit for genotyping a test sample at one or more polymorphic loci of interest, the kit comprising:

a) at least one set of query oligonucleotides for genotyping a polymorphic loci of interest, the set including: (i) at least one 5' ligation oligonucleotide having, from the 5' to 3' end, a first PCR primer binding region, a target-specific binding region selected to hybridize 5' of the polymorphic loci of interest, and a 3' region chosen to hybridize to either a consensus or variant nucleotide sequence at the polymorphic loci of interest, and (ii) a phosphorylated 3' ligation oligonucleotide having, from the 5' to 3' end, a target-specific binding region selected to hybridize 3' of the polymorphic loci of interest and a second PCR primer binding region; and

b) one or more pair(s) of detection primers, each detection primer pair having (i) a forward PCR primer that binds to the first PCR primer binding region in the 5' ligation oligonucleotide and (ii) a reverse PCR primer that binds to the second PCR primer binding region in the 3' ligation oligonucleotide.

48. The kit of Claim 47, wherein the 5' ligation oligonucleotide comprises the 5' PCR primer binding region having different nucleotide sequences.

49. The kit of Claim 47, wherein the forward and reverse PCR primers comprise penultimate 2 or 3 nucleotides at the 3' end that are selected to reduce primer-dimer formation by selecting 2 or 3 nucleotide that reduce annealing between the first and second pair of detection primers or that reduce self-annealing of the first and second pair of detection primers.

50. The kit of Claim 47, further comprising a DNA ligase.

51. The kit of Claim 50, wherein the ligase is thermostable.

52. The kit of Claim 47, further comprising at least one nucleic acid sample having a consensus nucleotide sequence or a variant nucleotide sequence at the polymorphic locus of interest.

53. The kit of Claim 47, wherein the one or more pair(s) of detection primers are disposed in a multi-well container.

Fig.1.

IDT Exhibit 1020.2615

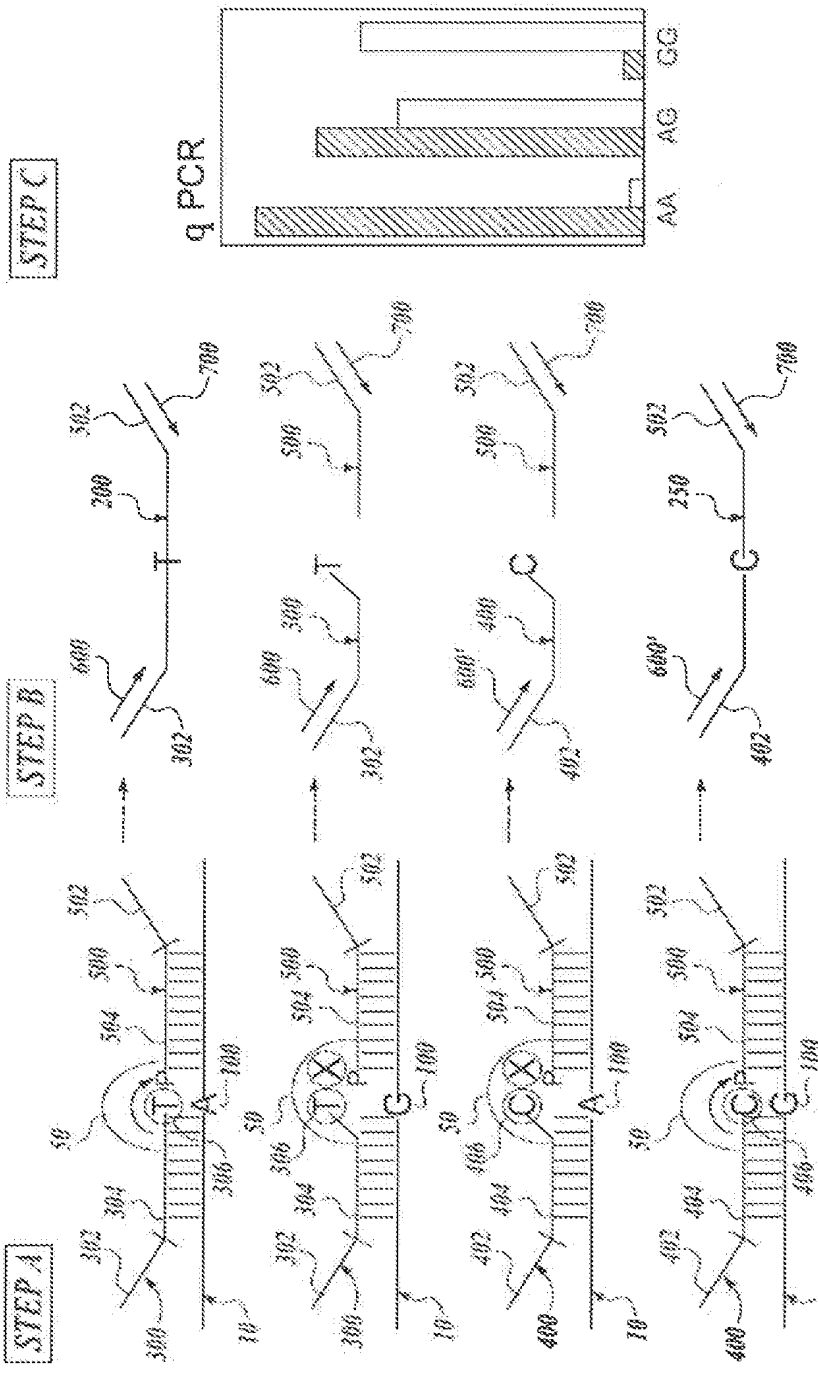


Fig.2.

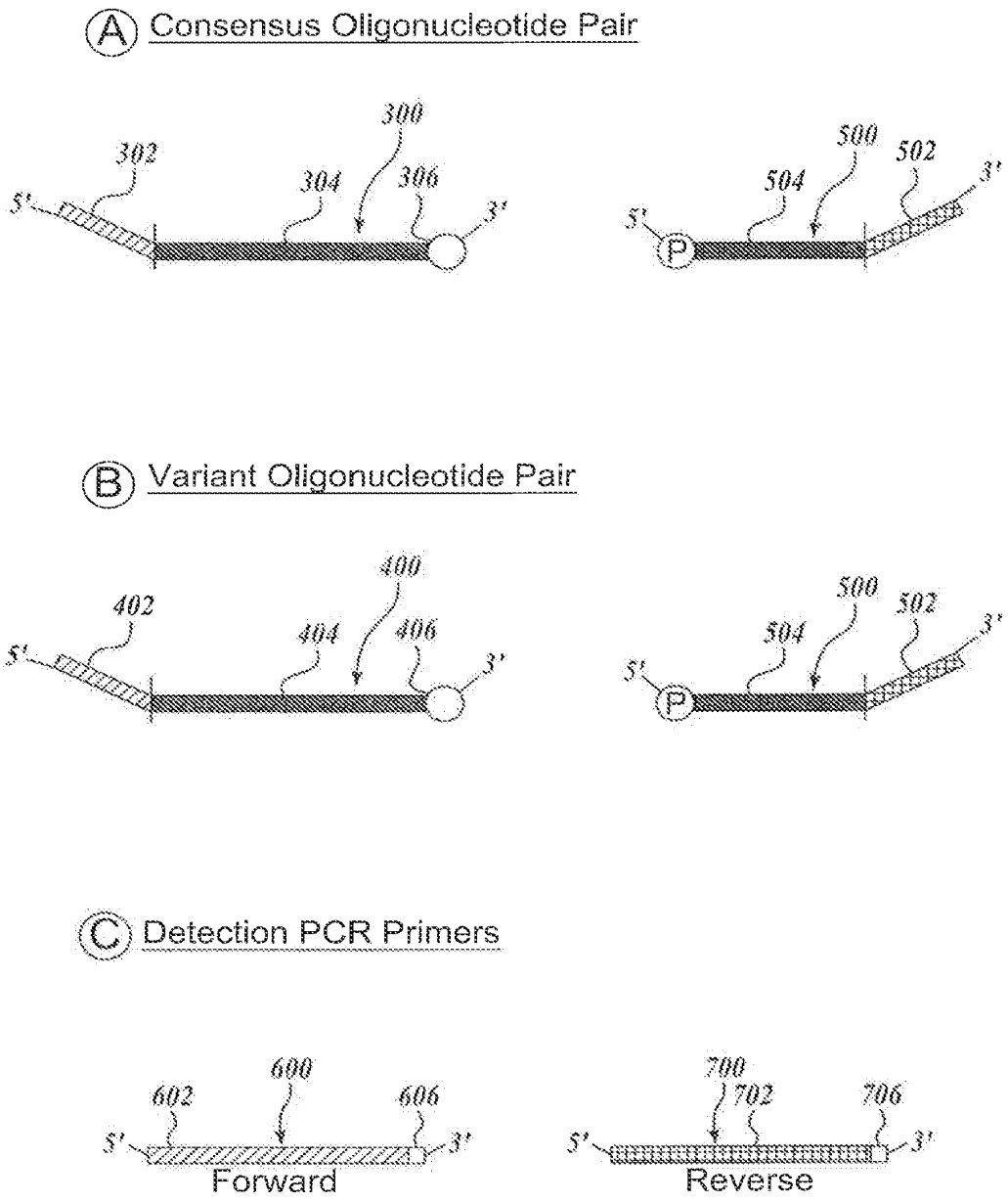

A Consensus Oligonucleotide Pair
300
302
304
306
5'
3'
500
502
504
5'
3'
P
B Variant Oligonucleotide Pair
400
402
404
406
5'
3'
500
502
504
5'
3'
P
C Detection PCR Primers
600
602
606
5'
3'
Forward
700
702
706
5'
3'
Reverse

Fig.3.

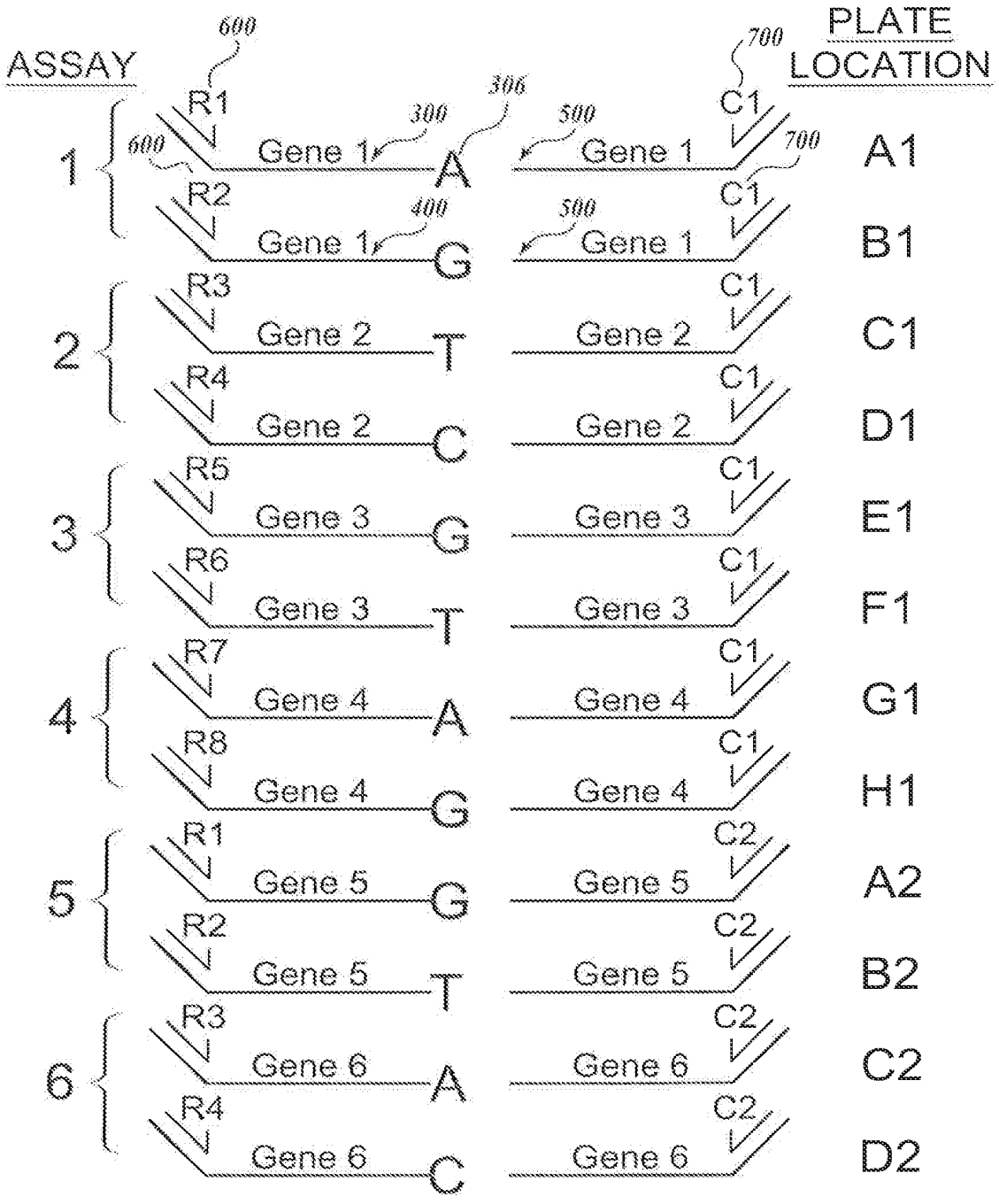

ASSAY
PLATE LOCATION
600
700
R1
C1
300
306
500
600
700
Gene 1
A
Gene 1
A1
1
R2
C1
400
500
Gene 1
G
Gene 1
B1
R3
C1
2
Gene 2
T
Gene 2
C1
R4
C1
Gene 2
C
Gene 2
D1
R5
C1
3
Gene 3
G
Gene 3
E1
R6
C1
Gene 3
T
Gene 3
F1
R7
C1
4
Gene 4
A
Gene 4
G1
R8
C1
Gene 4
G
Gene 4
H1
R1
C2
5
Gene 5
G
Gene 5
A2
R2
C2
Gene 5
T
Gene 5
B2
R3
C2
6
Gene 6
A
Gene 6
C2
R4
C2
Gene 6
C
Gene 6
D2

Fig.4.

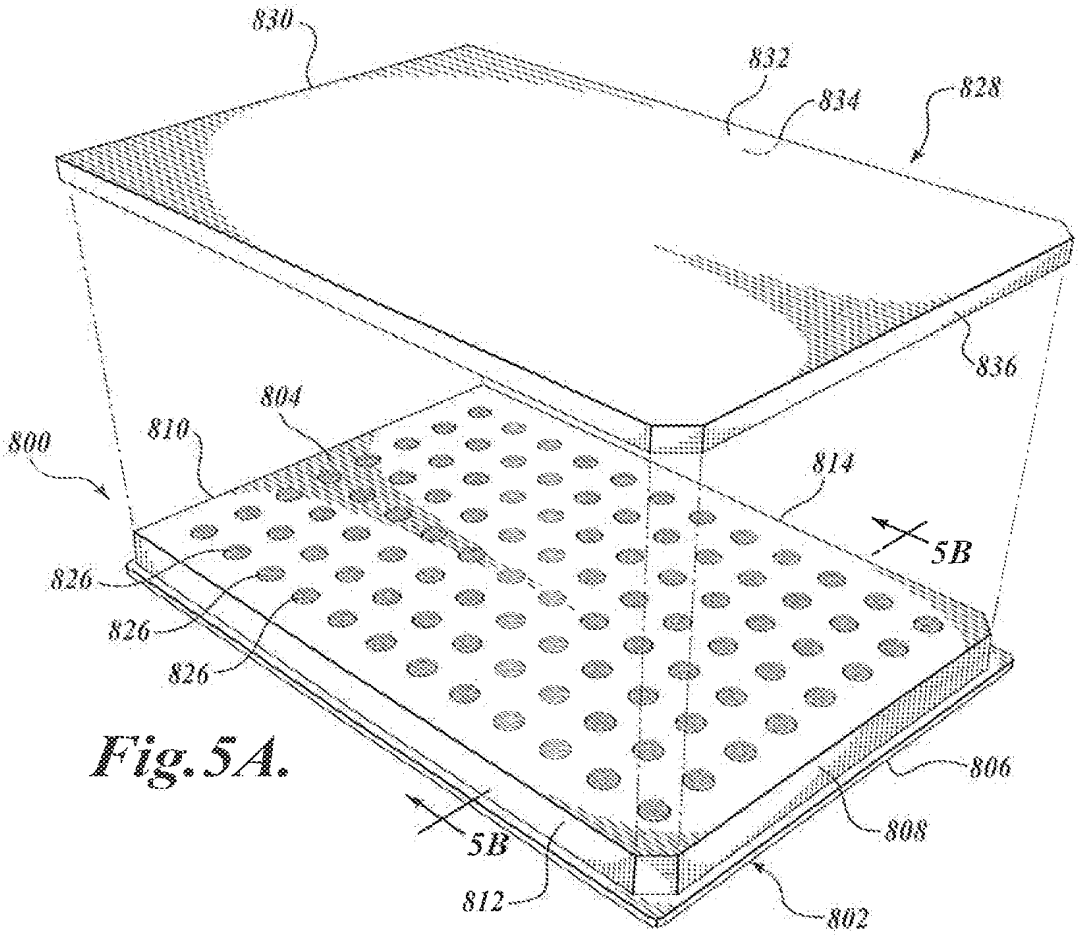

830
832
834
828
836
804
810
800
814
5B
826
826
826
806
808
5B
812
802

Fig.5A.

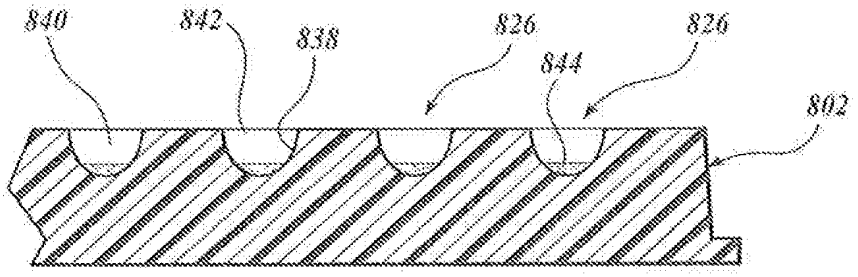

840
842
838
826
826
844
802

Fig.5B.

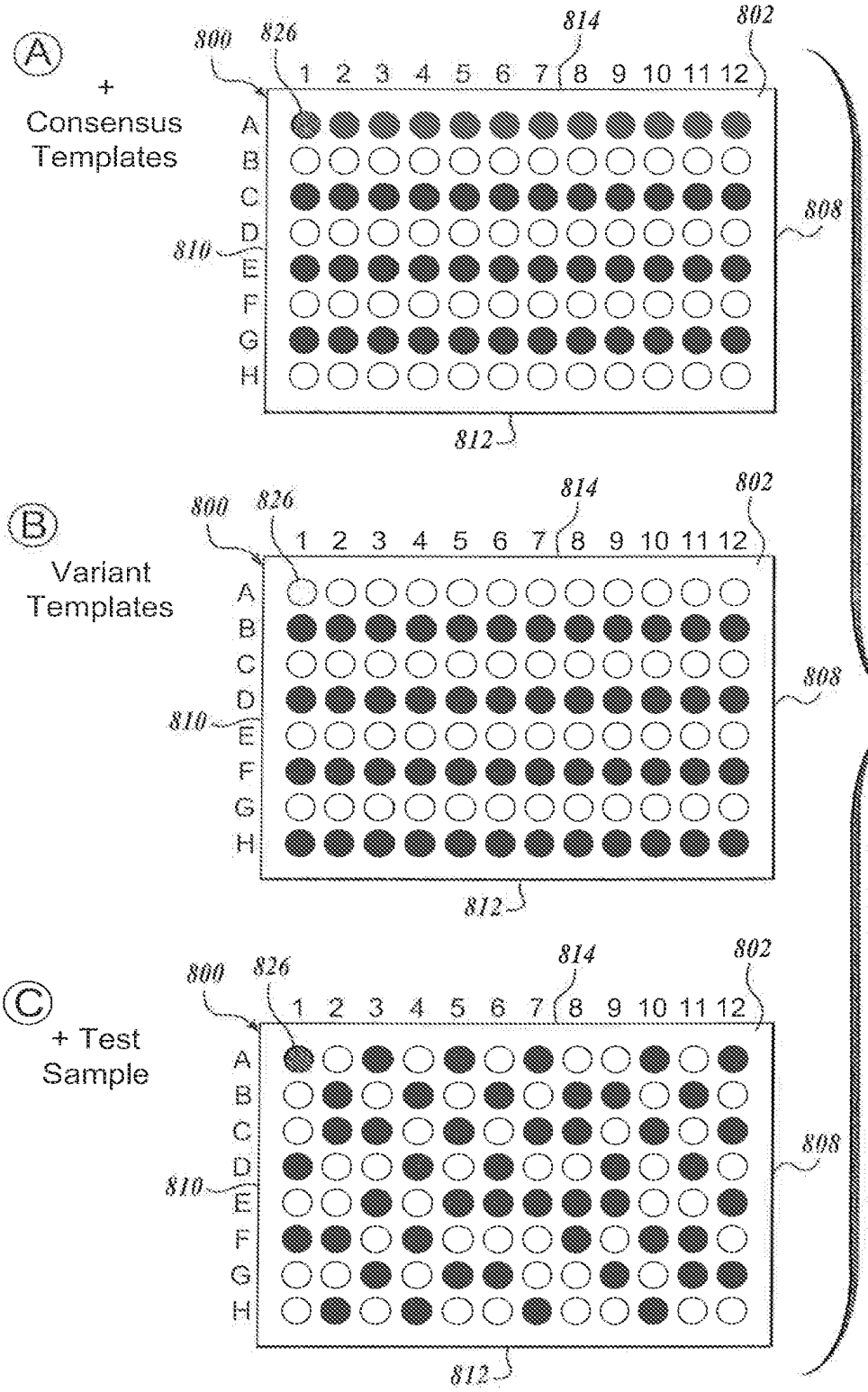

A
+
Consensus
Templates
B
Variant
Templates
C
+ Test
Sample
800
826
814
802
808
810
812
1 2 3 4 5 6 7 8 9 10 11 12
A
B
C
D
E
F
G
H

Fig. 6.

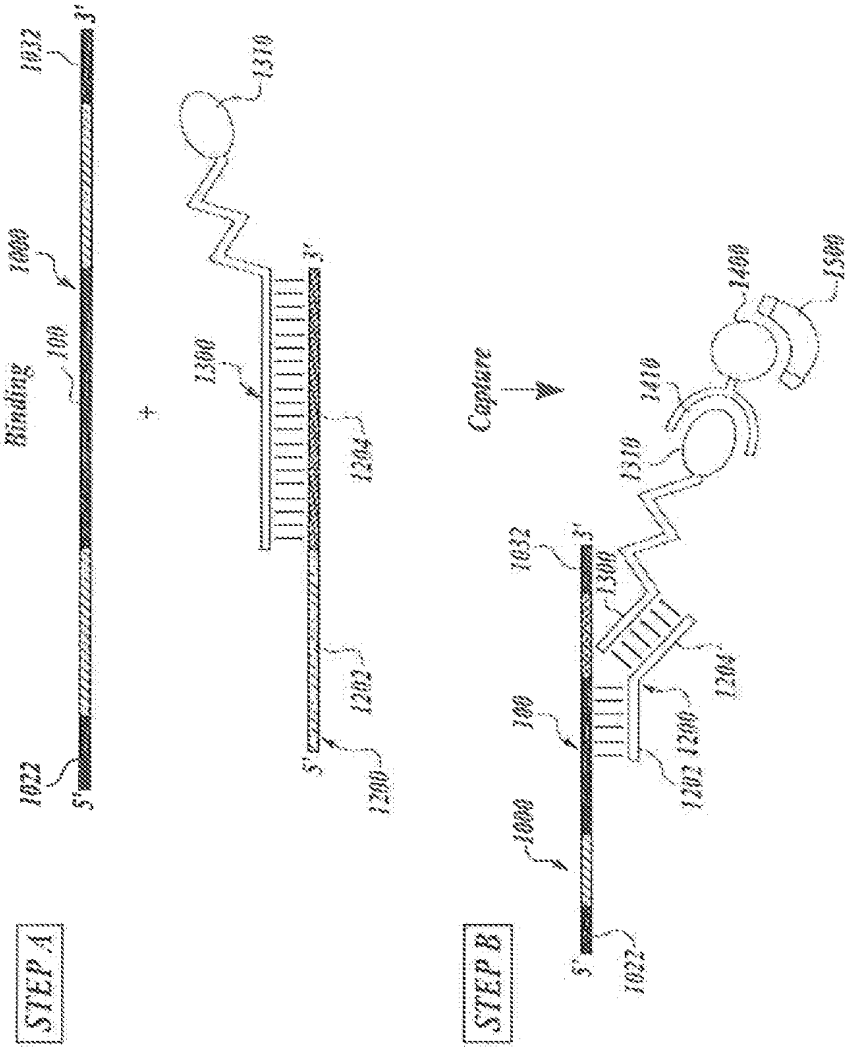

STEP A
Binding
1822
5'
100
1000
1032
3'
+
1310
1300
3'
1204
5'
1202
1200
STEP B
Capture
1000
1822
5'
100
1032
3'
1300
1310
1410
1400
1500
1204
1202
1200

Fig. 7.

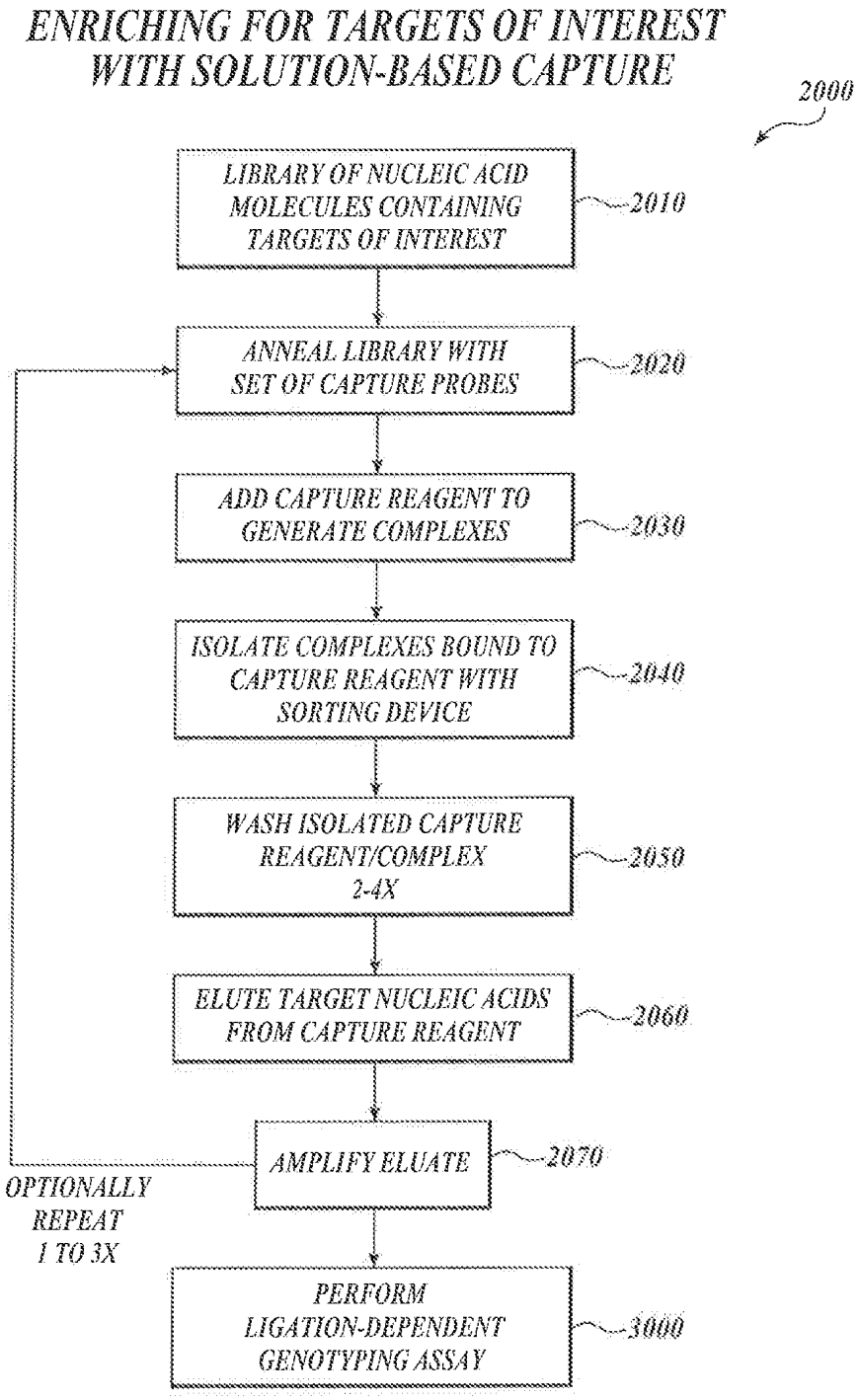

ENRICHING FOR TARGETS OF INTEREST WITH SOLUTION-BASED CAPTURE
2000
LIBRARY OF NUCLEIC ACID MOLECULES CONTAINING TARGETS OF INTEREST
2010
ANNEAL LIBRARY WITH SET OF CAPTURE PROBES
2020
ADD CAPTURE REAGENT TO GENERATE COMPLEXES
2030
ISOLATE COMPLEXES BOUND TO CAPTURE REAGENT WITH SORTING DEVICE
2040
WASH ISOLATED CAPTURE REAGENT/COMPLEX 2-4X
2050
ELUTE TARGET NUCLEIC ACIDS FROM CAPTURE REAGENT
2060
AMPLIFY ELUATE
2070
OPTIONALLY REPEAT 1 TO 3X
PERFORM LIGATION-DEPENDENT GENOTYPING ASSAY
3000

Fig. 8.

LIGATION - DEPENDENT GENOTYPING ASSAY

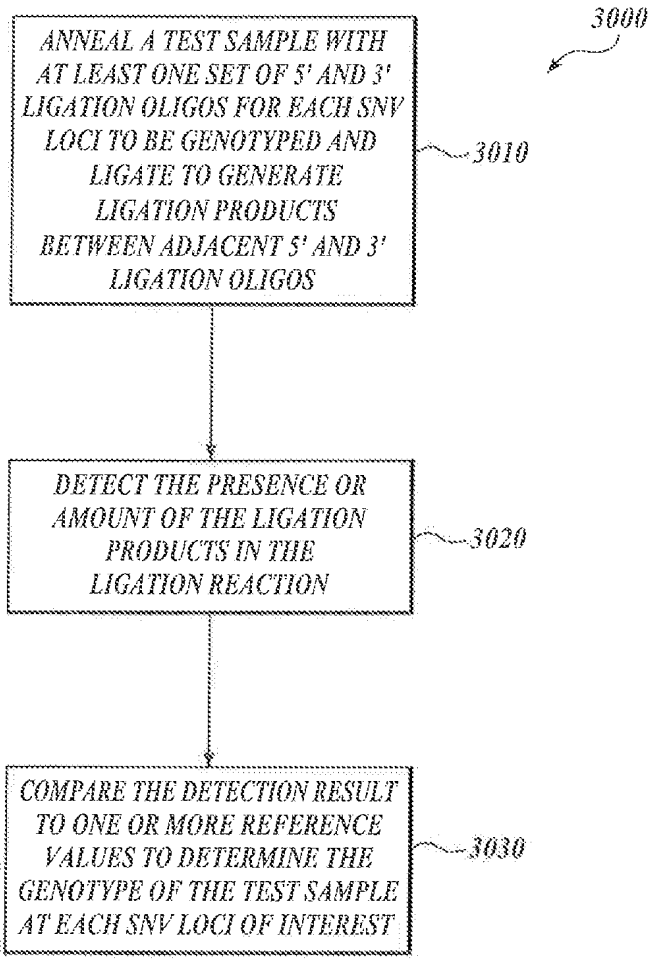


Fig. 9.

LIGATION-DEPENDENT MULTIPLEXED GENOTYPING ASSAY

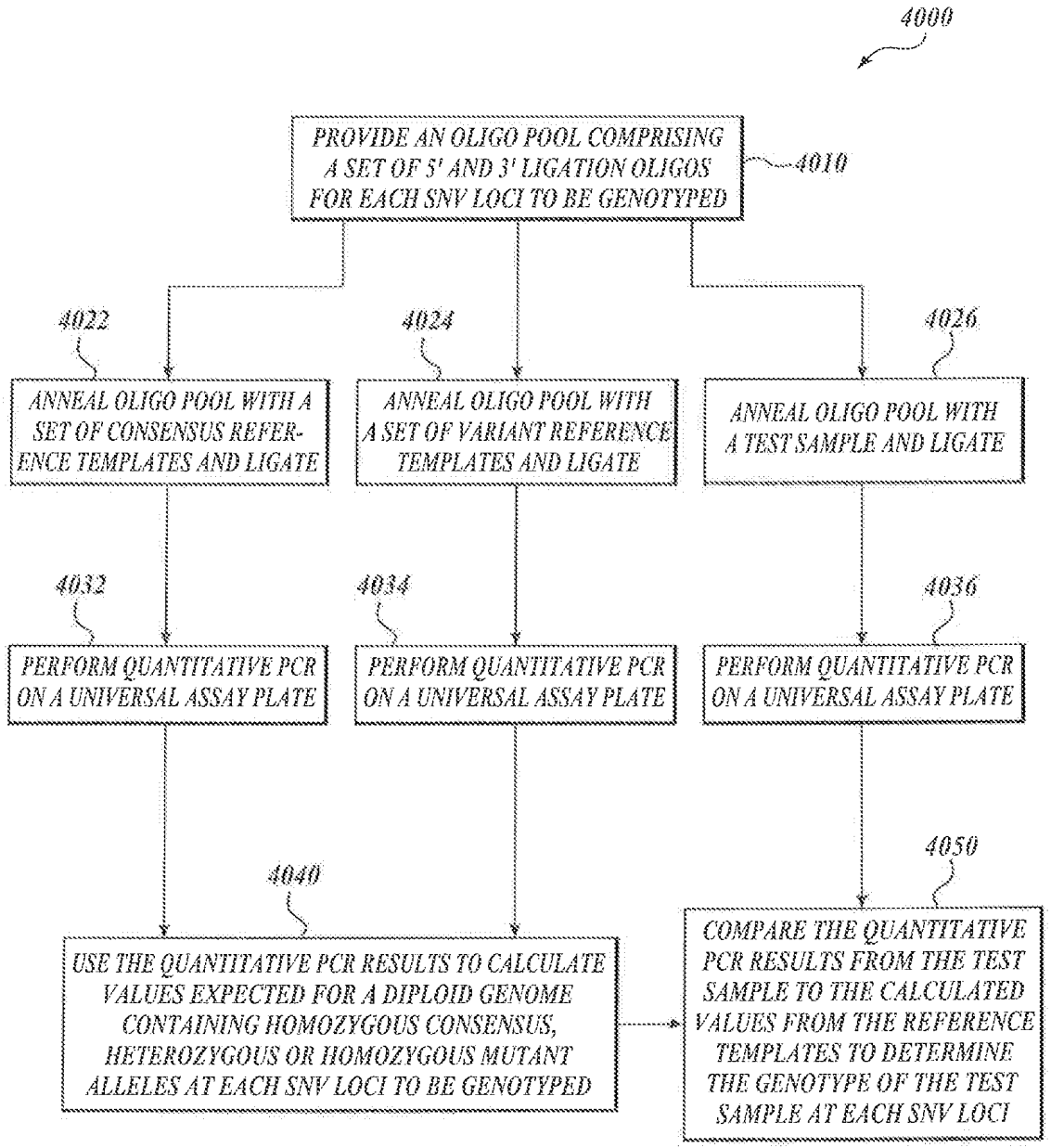


Fig. 10.

Espacenet

Bibliographic data: WO2010115154 (A1) — 2010-10-07

MULTI-PRIMER AMPLIFICATION METHOD FOR BARCODING OF TARGET NUCLEIC ACIDS

Inventor(s): MAY ANDREW [US]; CHEN PEILIN [US]; WANG JUN [US]; KAPER FIONA [US]; ANDERSON MEGAN [US] ± (MAY, ANDREW, ; CHEN, PEILIN, ; WANG, JUN, ; KAPER, FIONA, ; ANDERSON, MEGAN)

Applicant(s): FLUIDIGM CORP [US]; MAY ANDREW [US]; CHEN PEILIN [US]; WANG JUN [US]; KAPER FIONA [US]; ANDERSON MEGAN [US] ± (FLUIDIGM CORPORATION, ; MAY, ANDREW, ; CHEN, PEILIN, ; WANG, JUN, ; KAPER, FIONA, ; ANDERSON, MEGAN)

Classification: **- international:** ***C12Q1/68***
- cooperative: **B01L3/502738**; **C12Q1/6806**; **C12Q1/686**; B01L2300/0816; B01L2300/0864; B01L2300/0867; B01L2300/087; B01L2400/0487; B01L2400/0655; B01L3/50273; B01L7/52 more

Application number: WO2010US29854 20100402 Global Dossier

Priority number(s): US20090166181P 20090402 ; US20090166105P 20090402 ; US20090186327P 20090611 ; US20100305907P 20100218

Also published as: US2014272952 (A1) US2010273219 (A1) US8691509 (B2) SG10201402770Y (A) SG174617 (A1) more

Abstract of WO2010115154 (A1)

In certain embodiments, the present invention provides amplification methods in which nucleotide tag(s) and, optionally, a barcode nucleotide sequence are added to target nucleotide sequences. In other embodiments, the present invention provides a microfluidic device that includes a plurality of first input lines and a plurality of second input lines. The microfluidic device also includes a plurality of sets of first chambers and a plurality of sets of second chambers. Each set of first chambers is in fluid communication with one of the plurality of first input lines. Each set of second chambers is in fluid communication with one of the plurality of second input lines. The microfluidic device further includes a plurality of first pump elements in fluid communication with a first portion of the plurality of second input lines and a plurality of

IDT Exhibit 1020.2625

second pump elements in fluid communication with a second portion of the plurality of second input lines.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 October 2010 (07.10.2010)

PCT

(10) International Publication Number
WO 2010/115154 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2010/029854

(22) International Filing Date:
2 April 2010 (02.04.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

61/166,181	2 April 2009 (02.04.2009)	US
61/166,105	2 April 2009 (02.04.2009)	US
61/186,327	11 June 2009 (11.06.2009)	US
61/305,907	18 February 2010 (18.02.2010)	US

(71) Applicant *(for all designated States except US)*: **FLUIDIGM CORPORATION** [US/US]; 7000 Shoreline Court, Suite 100, South San Francisco, California 94080 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **MAY, Andrew** [GB/US]; 2087 Golden Gate Avenue, San Francisco, California 94115 (US). **CHEN, Peilin** [US/US]; 1023 Landmark Circle, Richmond, California 94806 (US). **WANG, Jun** [CA/US]; 1005 Moffett Circle, Palo Alto, California 94303 (US). **KAPER, Fiona** [NL/US]; 1808 Parkwood Drive, San Mateo, California 94403 (US). **ANDERSON, Megan** [US/US]; 2400 M Street NW, Apt. 439, Washington, District of Columbia 20037 (US).

(74) Agents: **HALIDAY, Emily, M.** et al.; WEAVER AUSTIN VILLENEUVE & SAMPSON LLP, P.O. Box 70250, Oakland, California 94612-0250 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

(54) **Title**: MULTI-PRIMER AMPLIFICATION METHOD FOR BARCODING OF TARGET NUCLEIC ACIDS

(57) **Abstract**: In certain embodiments, the present invention provides amplification methods in which nucleotide tag(s) and, optionally, a barcode nucleotide sequence are added to target nucleotide sequences. In other embodiments, the present invention provides a microfluidic device that includes a plurality of first input lines and a plurality of second input lines. The microfluidic device also includes a plurality of sets of first chambers and a plurality of sets of second chambers. Each set of first chambers is in fluid communication with one of the plurality of first input lines. Each set of second chambers is in fluid communication with one of the plurality of second input lines. The microfluidic device further includes a plurality of first pump elements in fluid communication with a first portion of the plurality of second input lines and a plurality of second pump elements in fluid communication with a second portion of the plurality of second input lines.

WO 2010/115154 A1

IDT Exhibit 1020.2627

MULTI-PRIMER AMPLIFICATION METHOD FOR BARCODING OF TARGET NUCLEIC ACIDS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of prior U.S. provisional application no. 61/166,181, filed April 2, 2009; prior U.S. provisional application no. 61/166,105, filed April 2, 2009; prior U.S. provisional application no. 61/186,327, filed June 11, 2009; and prior U.S. provisional application no. 61/305,907, filed February 18, 2010, which are all hereby incorporated by reference in their entireties.

FIELD OF THE INVENTION

[0002] The present invention relates generally to the area of high-throughput assays for detection and/ or sequencing of particular target nucleic acids. In certain embodiments, the present invention provides amplification methods in which nucleotide tag(s) and a barcode nucleotide sequence are added to target nucleotide sequences.

BACKGROUND OF THE INVENTION

[0003] The ability to detect specific nucleic acid sequences in a sample has resulted in new approaches in diagnostic and predictive medicine, environmental, food and agricultural monitoring, molecular biology research, and many other fields. In addition, new sequencing methodologies provide the means for rapid high-throughput nucleic acid sequencing.

[0004] Additional methods, especially methods that facilitate analysis of many targets and/or the analysis of many samples simultaneously across a broad range of concentrations in a sample would be of great benefit.

[0005] Microfluidic devices can be used for analytical, preparative, metering, and other manipulative functions on a scale not imagined until recently. The advantages of microfluidic devices include conservation of precious reagents and samples, high density and throughput of sample analysis or synthesis, fluidic precision and accuracy at a level scarcely visible to the unaided eye, and a space reduction accompanying the replacement of counterpart equipment operating at the

IDT Exhibit 1020.2628

macrofluidic scale. Associated with the reduction in size and the increased density of microfluidic devices is increased complexity and higher engineering and fabrication costs associated with increasingly intricate device architecture.

[0006] Recently, there have been concerted efforts to develop and manufacture microfluidic systems to perform various chemical and biochemical analyses and syntheses. Additionally, microfluidic devices have the potential to be adapted for use with automated systems, thereby providing the additional benefits of further cost reductions and decreased operator errors because of the reduction in human involvement. Microfluidic devices have been proposed for use in a variety of applications, including, for instance, capillary electrophoresis, gas chromatography, and cell separations.

[0007] However, realization of these benefits has often been thwarted because of various complications associated with the microfluidic devices that have thus far been manufactured. For instance, many of the current microfluidic devices are manufactured from silica-based substrates, which are difficult and complicated to machine. As a result, many devices made from such materials are fragile. Furthermore, transport of fluid through many existing microfluidic devices requires regulation of complicated electrical fields to transport fluids in a controlled fashion through the device.

[0008] Thus, in view of the foregoing benefits that can be achieved with microfluidic devices but the current limitations of existing devices, there remains a need for microfluidic devices designed for use in conducting a variety of chemical and biochemical analyses. Because of its importance in modern biochemistry, there is a particular need for devices that can be utilized to conduct a variety of nucleic acid amplification reactions, while having sufficient versatility for use in other types of analyses as well.

[0009] Devices with the ability to conduct nucleic acid amplifications would have diverse utilities. For example, such devices could be used as an analytical tool to determine whether a particular target nucleic acid of interest is present or absent in a sample. Thus, the devices could be utilized to test for the presence of particular pathogens (e.g., viruses, bacteria, or fungi), and for identification purposes (e.g., paternity and forensic applications). Such devices could also be utilized to detect or

IDT Exhibit 1020.2629

characterize specific nucleic acids previously correlated with particular diseases or genetic disorders. When used as analytical tools, the devices could also be utilized to conduct genotyping analyses and gene expression analyses (e.g., differential gene expression studies). Alternatively, the devices can be used in a preparative fashion to amplify sufficient nucleic acid for further analysis such as sequencing of amplified product, cell-typing, DNA fingerprinting, and the like. Amplified products can also be used in various genetic engineering applications, such as insertion into a vector that can then be used to transform cells for the production of a desired protein product.

[0010] Despite these advances in microfluidic design and use, it would be useful to reduce the complexity of microfluidic chips and simplify their operation. Additionally, a need exists for an increased ability to recover reaction products from microfluidic devices. Thus, there is a need in the art for improved methods and systems related to microfluidic devices.

SUMMARY OF THE INVENTION

[0011] In certain embodiments, the invention provides a method for amplifying, tagging, and barcoding a plurality of target nucleic acids in a plurality of samples. The method entails preparing an amplification mixture for each target nucleic acid. Each amplification mixture includes:

a forward primer comprising a target-specific portion;

a reverse primer comprising a target-specific portion, wherein the forward primer additionally comprises a first nucleotide tag and/or the reverse primer additionally comprises a second nucleotide tag; and

at least one barcode primer including a barcode nucleotide sequence and a first and/or second nucleotide tag-specific portion, wherein the barcode primer is in excess of the forward and/or reverse primer(s).

Each amplification mixture is subjected to amplification to produce a plurality of target amplicons, wherein each target amplicon includes a tagged target nucleotide sequence, with first and/or second nucleotide tags flanking the target nucleotide sequence, and at least one barcode nucleotide sequence at the 5' or 3' end of the target amplicon.

IDT Exhibit 1020.2630

[0012] In specific embodiments, the forward primer additionally includes a first nucleotide tag. If desired, the reverse primer can additionally include a second nucleotide tag.

[0013] In certain embodiments of the tagging/barcoding method, the concentration of the barcode primer in the amplification mixtures is at least 4-fold the concentration of the forward and/or reverse primer(s). In variations of such embodiments, the concentration of the barcode primer in the amplification mixtures is at least 50-fold the concentration of the forward and/or reverse primer(s).

[0014] In particular embodiments of the barcoding/tagging method, the first and/or second nucleotide tags and/or the barcode nucleotide sequence are selected so as to avoid substantial annealing to the target nucleic acids. In illustrative embodiments, the barcode nucleotide sequence identifies a particular sample. Where the barcode primer includes a barcode nucleotide sequence and a first nucleotide tag-specific portion, in certain embodiments, a plurality of forward primers include the same first nucleotide tag. For, example, where multiple targets are to be amplified in different samples, the set of forward primers corresponding to the set of targets can all have the same first nucleotide tag.

[0015] In particular embodiments of the barcoding/tagging method, the forward and reverse primers for each target are initially combined separately from the sample, and each barcode primer is initially combined with its corresponding sample. For example, where T targets are to be amplified in S samples, T and S being integers greater than one, the method can additionally include preparing S×T amplification mixtures wherein the initially combined forward and reverse primers are added to the initially combined samples and barcode primers.

[0016] In certain embodiments of the barcoding/tagging method, the amplification is carried out for at least 3 cycles to introduce the first and second nucleotide tags and the barcode nucleotide sequence. In variations of these embodiments, the amplification is carried out for between 5 and 50 cycles. In particular embodiments, the amplification is carried out for a sufficient number of cycles to normalize target amplicon copy number across targets and across samples.

[0017] In certain embodiments of the barcoding/tagging method, at least 50 percent of the target amplicons produced upon amplification are present at greater

IDT Exhibit 1020.2631

than 50 percent of the average number of copies of target amplicons and less than 2-fold the average number of copies of target amplicons.

[0018] In other embodiments, the invention provides a method in which barcoding is, optionally, omitted and the target nucleotide sequences are tagged after amplification. This method entails amplifying a plurality of target nucleic acids, typically, in a plurality of samples. An amplification mixture is prepared for each target nucleic acid, wherein each amplification mixture includes:

a forward primer including a target-specific sequence; and

a reverse primer including a target-specific sequence;

Each amplification mixture is subjected to amplification to produce a plurality of target nucleotide sequences. The target nucleotide sequences are then tagged (e.g., by ligation of nucleotide tags unto one or both ends of the target nucleotide sequences) to produce a plurality of target amplicons. Each target amplicon includes first and/or second nucleotide tags flanking the target nucleotide sequence. In particular embodiments, at least 50 percent of the target amplicons are present at greater than 50 percent of the average number of copies of target amplicons and less than 2-fold the average number of copies of target amplicons.

[0019] In certain embodiments of the amplification methods described herein, at least 70 percent of the target amplicons are present at greater than 50 percent of the average number of copies of target amplicons and less than 2-fold the average number of copies of target amplicons. In illustrative embodiments, at least 90 percent of the target amplicons are present at greater than 50 percent of the average number of copies of target amplicons and less than 2-fold the average number of copies of target amplicons.

[0020] In various embodiments, the average length of the target amplicons is at least 25 bases, 50 bases, 100 bases, 200 bases, 500 bases, and 750 bases. Longer average lengths, such as 1 kilobase or more are also possible, as, for example, when amplification is carried out by long-range PCR. In such embodiments, amplification may yield target amplicons wherein at least 70 percent of the target amplicons are present at greater than 50 percent of the average number of copies of target amplicons and less than 2-fold the average number of copies of target amplicons.

IDT Exhibit 1020.2632

[0021] An advantage of the methods described herein is that amplification can (but need not) be carried out in small reaction volumes. In particular embodiments, the volume of the amplification mixtures is in the range of about 1 picoliter to about 50 nanoliters. In certain embodiments, the volume of the amplification mixtures is in the range of about 5 picoliters to about 25 nanoliters.

[0022] The methods described herein can, optionally, include recovering the target amplicons from the amplification mixtures. In certain embodiments, the target amplicons are recovered in a volume and/or copy number that varies less than about 50% among the recovered target amplicons. The recovered amplicons can be employed for further amplification and/or analysis (e.g., DNA sequencing). In some embodiments, at least one target amplicon can be subjected to amplification using primers specific for the first and second nucleotide tags to produce a target amplicon lacking the barcode nucleotide sequence, if such is desired.

[0023] In particular embodiments of the methods described herein, the target nucleic acids include genomic DNA. In variations of these embodiments, the genomic DNA can be DNA intended for DNA sequencing, e.g., automated DNA sequencing

[0024] In certain embodiments, one or more of the forward primer, reverse primer, and barcode primer can include at least one additional primer binding site. For example, if a barcode primer is employed, the barcode primer can include at least a first additional primer binding site upstream of the barcode nucleotide sequence, which is upstream of the first nucleotide tag. In such an embodiment, the reverse primer can include at least a second additional primer binding site downstream of the second nucleotide tag. In particular embodiments, where the target nucleotide sequences are to be sequenced by automated DNA sequencing, the first and second additional primer binding sites are capable of being bound by DNA sequencing primers.

[0025] If a barcode primer is not employed, and the target nucleotide sequences are tagged after amplification, the first and second nucleotide tags can be capable of being bound by DNA sequencing primers.

[0026] Thus, the methods described herein can, optionally, include subjecting at least one target amplicon to DNA sequencing.

IDT Exhibit 1020.2633

[0027] In certain embodiments, the method can, optionally, include quantifying the amount of target amplicons in the amplification mixtures. This step may be carried out, for example, prior to automated DNA sequencing. In particular embodiments, quantification includes recovering the target amplicons and subjecting them to digital amplification. Digital amplification includes, in particular embodiments,

distributing the preamplified target amplicons into discrete reaction mixtures, wherein each reaction mixture, on average, includes no more than one amplicon per reaction mixture; and

subjecting the reaction mixtures to amplification.

Quantification in digital amplification may be carried out by real-time PCR and/or endpoint PCR.

[0028] The amplification methods described herein can, optionally, include determining the amount of each target nucleic acid present in each sample. In certain embodiments, the methods can be performed in determining the copy numbers of the target nucleic acids in each sample. In particular embodiments, the methods can be performed in determining the genotypes at loci corresponding to the target nucleic acids. In other embodiments, the methods can be performed in determining the expression levels of the target nucleic acids.

[0029] In particular embodiments, the present invention relates to microfluidic devices. More particularly, the present invention relates to a microfluidic device that provides for recovery of reaction products. Merely by way of example, the method and apparatus has been applied to a PCR sample preparation system used to prepare libraries for next generation sequencing. However, it would be recognized that the invention has a much broader range of applicability.

[0030] According to an embodiment of the present invention, a microfluidic device is provided. The microfluidic device includes a plurality of first input lines and a plurality of second input lines. The microfluidic device also includes a plurality of sets of first chambers and a plurality of sets of second chambers. Each set of first chambers is in fluid communication with one of the plurality of first input lines and each set of second chambers is in fluid communication with one of the plurality of second input lines. The microfluidic device further includes a plurality of first pump

IDT Exhibit 1020.2634

elements in fluid communication with a first portion of the plurality of second input lines and a plurality of second pump elements in fluid communication with a second portion of the plurality of second input lines.

[0031] According to another embodiment of the present invention, a method of operating a microfluidic device having an assay chamber, a sample chamber, and a harvesting port is provided. The method includes closing a fluid line between the assay chamber and the sample chamber, flowing a sample into the sample chamber via a sample input line, and flowing an assay into the assay chamber via an assay input line. The method also includes opening the fluid line between the assay chamber and the sample chamber, combining at least a portion of the sample and at least a portion of the assay to form a mixture, and reacting the mixture to form a reaction product. The method further includes closing the fluid line between the assay chamber and the sample chamber, flowing a harvesting reagent from the harvesting port to the sample chamber, and removing the reaction product from the microfluidic device.

[0032] According to a particular embodiment of the present invention, a method of preparing reaction products is provided. The method includes providing M samples and providing N assays. The method also includes mixing the M samples and N assays to form MxN pairwise combinations. Each of the MxN pairwise combinations are contained in a closed volume. The method further includes forming MxN reaction products from the MxN pairwise combinations and recovering the MxN reaction products.

[0033] Many benefits are achieved by way of the present invention over conventional techniques. For example, embodiments of the present invention provide for mixing and reaction of MxN samples and assays followed by recovery of the reaction products in sample-by-sample pools. Additionally, dilation pumping is utilized to remove substantially all of the reaction products from the microfluidic device, providing uniformity between the various reaction product pools. Utilizing the systems and methods described herein, the time and labor required to prepare libraries is reduced in comparison with conventional techniques. These and other embodiments of the invention along with many of its advantages and features are described in more detail in conjunction with the text below and attached figures.

IDT Exhibit 1020.2635

BRIEF DESCRIPTION OF THE DRAWINGS

[0034] The invention may be understood by reference to the following description taken in conjunction with the accompanying drawings that illustrate certain specific embodiments of the present invention.

[0035] Figure 1 depicts an illustrative matrix-type microfluidic device in plan view.

[0036] Figure 2 is a simplified perspective illustration of a carrier and a microfluidic device that permits recovery of reaction products.

[0037] Figure 3 is a simplified schematic diagram of a microfluidic device that permits recovery of reaction products.

[0038] Figure 4 is a simplified schematic diagram of several unit cells of the microfluidic device illustrated in Fig. 3.

[0039] Figure 5A is simplified schematic diagram of a microfluidic device that permits recovery of reaction products.

[0040] Figure 5B is a simplified schematic diagram of portions of the microfluidic device illustrated in Fig. 5A.

[0041] Figure 6 is a simplified schematic diagram of several unit cells of the microfluidic device illustrated in Fig. 5A.

[0042] Figure 7 is a simplified flowchart of a method of operating a microfluidic device that permits recovery of reaction products.

[0043] Figures 8A-8D are simplified schematic diagrams illustrating fluid flow through unit cells of a microfluidic device that permits recovery of reaction products during operation.

[0044] Figures 9A-9D are simplified schematic diagrams illustrating fluid flow through a microfluidic device that permits recovery of reaction products during operation.

[0045] Figure 10 illustrates an embodiment of a 3-primer amplification method for barcoding target nucleic acids prior to sequencing.

[0046] Figure 11 shows a photograph of the gel described in Example 1. The lanes are as follows: (2) molecular markers, (4) sample amplified with 454 tails; (5)

IDT Exhibit 1020.2636

sample (NTC) amplified with 454 tails; (7) sample amplified with A5 primer pair; (8) sample (NTC) amplified with A5 primer pair; (10) sample amplified with 3 primers; and (11) sample (NTC) amplified with 3 primers.

[0047] Figure 12 shows gel-view electropherograms obtained from 4 Agilent 1K Biolanalyzer chips for each of the individual samples run on the Access Array IFC (Integrated Fluidic Circuit) in Example 4. Each column in the figure shows the size distribution of DNA products in each sample. All samples produce similar distributions of products

[0048] Figure 13A-13B shows results from Example 4. A) Predicted sizes of all PCR products for this set of target specific primers. B) Electropherogram of one of the sample pools obtained from the Access Array IFC. Distribution of product size within a single product pool. All products fall within the predicted size range shown in (B).

[0049] Figure 14A-14C shows results from Example 4. A) Number of sequences counted per barcode on the 454 sequence run. Upper horizontal line represents 2x average number of counts per barcode. Lower horizontal line represents 50% of average number of counts per barcode. B) Number of sequences counted per amplicon. Each point on the plot represents the number of times the sequence for an individual chamber on the Access Array IFC were measured on the sequencer. Triangular points represent PCR reactions with greater than 2x the average representation. Dark grey points represent PCR reactions with less than 0.5x the average representation. C) Frequency distribution of amplicon representation. The dark grey line represents the number of amplicons present at a given representation. The light grey line represents the number of reads that would be measured at a given coverage (e.g. 98% at 20x coverage). Percentage of amplicons within 2-fold of average: 95.8%; percentage of amplicons within 5-fold of average: 99.7%.

[0050] Figure 15A-15B shows an example of a multi-primer reaction set-up using 4 outer primers with different combinations of primer binding site and nucleotide tags. (Example 5.) A) Two forward barcode primers (454B-BC-Tag8, 454A-BC-Tag8 and two reverse barcode primers (454A-BC-Tag5, 454B-BC-Tag8) are combined with one inner primer pair (Tag8-TSF and Tag5-TSR). B) The two major PCR products formed from this PCR reaction. PCR products containing 454-A

IDT Exhibit 1020.2637

Tag8 and 454A-Tag5 at each end or 454B-Tag8 and 454B-Tag5 at each end do not produce significant PCR products due to PCR suppression

[0051] Figure 16A-16B shows the representation of each of the primer sequences in each of the samples for each of the amplicons in Figure 15B. The number of sequences counted per amplicon were normalized to the average number of counts per amplicon within a sample. The normalized counts for an individual amplicon were summed between the A and B emulsions (A) for Tag5 amplicons in Emulsion A plus Tag 8 amplicons in Emulsion B and (B) for Tag5 amplicons in Emulsion B plus Tag 8 amplicons in Emulsion A. The middle dark grey line represents the average representation of each amplicon. The upper light grey line represents 2x average coverage. The lower light grey line represents 50% of average representation.

[0052] Figure 17 shows the results from Example 6: Successful amplication of a PCR product using the 4-primer strategy designed for use on the Illumina GA II sequencer. The barcode primers listed in Table 14 are labelled as Outer Short.

[0053] Figure 18 shows results from Example 8: PCR reactions of three pools of 10 sets of PCR primers (A, B, C) when the PCR reactions were run for template-specific primers only and in 4-primer mode. The presence of higher molecular weight products in the 4-primer strategy demonstrates successful 4-primer assembly.

[0054] Figure 19 shows results from Example 8: Changing the ratio of inner and outer primers impacts yield in multiplex 4-primer PCR using inner and outer primers.

DETAILED DESCRIPTION

[0055] In certain embodiments, the present invention provides amplification methods in which nucleotide tag(s) and a barcode nucleotide sequence are added to target nucleotide sequences. The added sequences can then serve as primer and/or probe-binding sites. The barcode nucleotide sequence can encode information, such as, e.g., sample origin, about the target nucleotide sequence to which it is attached. Tagging and/or barcoding target nucleotide sequences can increase the number of samples that can be analyzed for one or multiple targets in a single assay, while

IDT Exhibit 1020.2638

minimizing increases in assay cost. The methods are particularly well-suited for increasing the efficiency of assays performed on microfluidic devices.

[0056] In particular embodiments, the methods are used to prepare nucleic acids for DNA sequencing by, e.g., adding binding sites for DNA sequencing primers, optionally followed by sample calibration for DNA sequencing. In specific, illustrative embodiments, the method can be employed to add binding sites for DNA sequencing primers in a microfluidic device that permits recovery of reaction products. In illustrative devices of this type, dilation pumping can utilized to remove substantially all of the reaction products from the microfluidic device, providing uniformity between the various reaction product pools. Thus, it is possible to produce pools of barcoded reaction products that are uniform with respect to volume and copy number. In various embodiments, the volume and/or copy number uniformity is such that the variability, with respect to volume and/or copy number, of each pool recovered from the device is less than about 100 percent, less than about 90 percent, less than about 80 percent, less than about 70 percent, less than about 60 percent, less than about 50 percent, less than about 40 percent, less than about 30 percent, less than about 20 percent, less than about 17 percent, or less than about 15, 12, 10, 9, 8, 7, 6, 5, 4.5, 4, 3.5, 3, 2.5, 2, 1.5, 1, or 0.5 percent. Those of skill in the art appreciate that the volume and/or copy number variability may fall within any range bounded by any of these values (e.g., about 2 to about 7 percent). In an illustrative embodiment, the volume samples recovered from a microfluidic device vary by no more than approximately 10%. Standard pipetting error is on the order of between 5 and 10%. Thus, the observed variability in volumes is largely attributable to pipetting error. Utilizing the systems and methods described herein, the time and labor required to prepare sequencing libraries is reduced in comparison with conventional techniques.

[0057] It is understood that the invention is not limited to the particular methodology, protocols, and reagents, etc., described herein, as these can be varied by the skilled artisan. It is also understood that the terminology used herein is used for the purpose of describing particular illustrative embodiments only, and is not intended to limit the scope of the invention. It also noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include the plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a

IDT Exhibit 1020.2639

cell" is a reference to one or more cells and equivalents thereof known to those skilled in the art.

[0058] The embodiments of the invention and the various features and advantageous details thereof are explained more fully with reference to the non-limiting embodiments and examples that are described and/or illustrated in the accompanying drawings and detailed in the following description. It should be noted that the features illustrated in the drawings are not necessarily drawn to scale, and features of one embodiment may be employed with other embodiments as the skilled artisan would recognize, even if not explicitly stated herein. Descriptions of well-known components and processing techniques may be omitted so as to not unnecessarily obscure the embodiments of the invention.

Definitions

[0059] Terms used in the claims and specification are defined as set forth below unless otherwise specified. These terms are defined specifically for clarity, but all of the definitions are consistent with how a skilled artisan would understand these terms.

[0060] The term "adjacent," when used herein to refer two nucleotide sequences in a nucleic acid, can refer to nucleotide sequences separated by 0 to about 20 nucleotides, more specifically, in a range of about 1 to about 10 nucleotides, or sequences that directly abut one another.

[0061] The term "nucleic acid" refers to a nucleotide polymer, and unless otherwise limited, includes known analogs of natural nucleotides that can function in a similar manner (e.g., hybridize) to naturally occurring nucleotides.

[0062] The term nucleic acid includes any form of DNA or RNA, including, for example, genomic DNA; complementary DNA (cDNA), which is a DNA representation of mRNA, usually obtained by reverse transcription of messenger RNA (mRNA) or by amplification; DNA molecules produced synthetically or by amplification; and mRNA.

[0063] The term nucleic acid encompasses double- or triple-stranded nucleic acids, as well as single-stranded molecules. In double- or triple-stranded nucleic

IDT Exhibit 1020.2640

acids, the nucleic acid strands need not be coextensive (i.e, a double-stranded nucleic acid need not be double-stranded along the entire length of both strands).

[0064] The term nucleic acid also encompasses any chemical modification thereof, such as by methylation and/or by capping. Nucleic acid modifications can include addition of chemical groups that incorporate additional charge, polarizability, hydrogen bonding, electrostatic interaction, and functionality to the individual nucleic acid bases or to the nucleic acid as a whole. Such modifications may include base modifications such as 2'-position sugar modifications, 5-position pyrimidine modifications, 8-position purine modifications, modifications at cytosine exocyclic amines, substitutions of 5-bromo-uracil, backbone modifications, unusual base pairing combinations such as the isobases isocytidine and isoguanidine, and the like.

[0065] More particularly, in certain embodiments, nucleic acids, can include polydeoxyribonucleotides (containing 2-deoxy-D-ribose), polyribonucleotides (containing D-ribose), and any other type of nucleic acid that is an N- or C-glycoside of a purine or pyrimidine base, as well as other polymers containing nonnucleotidic backbones, for example, polyamide (e.g., peptide nucleic acids (PNAs)) and polymorpholino (commercially available from the Anti-Virals, Inc., Corvallis, Oregon, as Neugene) polymers, and other synthetic sequence-specific nucleic acid polymers providing that the polymers contain nucleobases in a configuration which allows for base pairing and base stacking, such as is found in DNA and RNA. The term nucleic acid also encompasses linked nucleic acids (LNAs), which are described in U.S. Patent Nos. 6,794,499, 6,670,461, 6,262,490, and 6,770,748, which are incorporated herein by reference in their entirety for their disclosure of LNAs.

[0066] The nucleic acid(s) can be derived from a completely chemical synthesis process, such as a solid phase-mediated chemical synthesis, from a biological source, such as through isolation from any species that produces nucleic acid, or from processes that involve the manipulation of nucleic acids by molecular biology tools, such as DNA replication, PCR amplification, reverse transcription, or from a combination of those processes.

[0067] The term "target nucleic acids" is used herein to refer to particular nucleic acids to be detected in the methods of the invention.

IDT Exhibit 1020.2641

[0068] As used herein the term "target nucleotide sequence" refers to a molecule that includes the nucleotide sequence of a target nucleic acid, such as, for example, the amplification product obtained by amplifying a target nucleic acid or the cDNA produced upon reverse transcription of an RNA target nucleic acid.

[0069] As used herein, the term "complementary" refers to the capacity for precise pairing between two nucleotides. I.e., if a nucleotide at a given position of a nucleic acid is capable of hydrogen bonding with a nucleotide of another nucleic acid, then the two nucleic acids are considered to be complementary to one another at that position. Complementarity between two single-stranded nucleic acid molecules may be "partial," in which only some of the nucleotides bind, or it may be complete when total complementarity exists between the single-stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands.

[0070] "Specific hybridization" refers to the binding of a nucleic acid to a target nucleotide sequence in the absence of substantial binding to other nucleotide sequences present in the hybridization mixture under defined stringency conditions. Those of skill in the art recognize that relaxing the stringency of the hybridization conditions allows sequence mismatches to be tolerated.

[0071] In particular embodiments, hybridizations are carried out under stringent hybridization conditions. The phrase "stringent hybridization conditions" generally refers to a temperature in a range from about 5°C to about 20°C or 25°C below than the melting temperature (T_m) for a specific sequence at a defined ionic strength and pH. As used herein, the T_m is the temperature at which a population of double-stranded nucleic acid molecules becomes half-dissociated into single strands. Methods for calculating the T_m of nucleic acids are well known in the art (see, e.g., Berger and Kimmel (1987) METHODS IN ENZYMOLOGY, VOL.152: GUIDE TO MOLECULAR CLONING TECHNIQUES, San Diego: Academic Press, Inc. and Sambrook et al. (1989) MOLECULAR CLONING: A LABORATORY MANUAL, 2ND ED., VOLS. 1-3, Cold Spring Harbor Laboratory), both incorporated herein by reference). As indicated by standard references, a simple estimate of the T_m value may be calculated by the equation: $T_m = 81.5+0.41(\%\ G+C)$, when a nucleic acid is in aqueous solution at 1 M NaCl (see, e.g., Anderson and Young, Quantitative Filter Hybridization in NUCLEIC ACID HYBRIDIZATION (1985)). The melting

IDT Exhibit 1020.2642

temperature of a hybrid (and thus the conditions for stringent hybridization) is affected by various factors such as the length and nature (DNA, RNA, base composition) of the primer or probe and nature of the target nucleic acid (DNA, RNA, base composition, present in solution or immobilized, and the like), as well as the concentration of salts and other components (e.g., the presence or absence of formamide, dextran sulfate, polyethylene glycol). The effects of these factors are well known and are discussed in standard references in the art. Illustrative stringent conditions suitable for achieving specific hybridization of most sequences are: a temperature of at least about 60°C and a salt concentration of about 0.2 molar at pH7.

[0072] The term "oligonucleotide" is used to refer to a nucleic acid that is relatively short, generally shorter than 200 nucleotides, more particularly, shorter than 100 nucleotides, most particularly, shorter than 50 nucleotides. Typically, oligonucleotides are single-stranded DNA molecules.

[0073] The term "primer" refers to an oligonucleotide that is capable of hybridizing (also termed "annealing") with a nucleic acid and serving as an initiation site for nucleotide (RNA or DNA) polymerization under appropriate conditions (i.e., in the presence of four different nucleoside triphosphates and an agent for polymerization, such as DNA or RNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. The appropriate length of a primer depends on the intended use of the primer, but primers are typically at least 7 nucleotides long and, more typically range from 10 to 30 nucleotides, or even more typically from 15 to 30 nucleotides, in length. Other primers can be somewhat longer, e.g., 30 to 50 nucleotides long. In this context, "primer length" refers to the portion of an oligonucleotide or nucleic acid that hybridizes to a complementary "target" sequence and primes nucleotide synthesis. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template but must be sufficiently complementary to hybridize with a template. The term "primer site" or "primer binding site" refers to the segment of the target nucleic acid to which a primer hybridizes.

[0074] A primer is said to anneal to another nucleic acid if the primer, or a portion thereof, hybridizes to a nucleotide sequence within the nucleic acid. The statement that a primer hybridizes to a particular nucleotide sequence is not intended

IDT Exhibit 1020.2643

to imply that the primer hybridizes either completely or exclusively to that nucleotide sequence. For example, in certain embodiments, amplification primers used herein are said to "anneal to a nucleotide tag." This description encompasses primers that anneal wholly to the nucleotide tag, as well as primers that anneal partially to the nucleotide tag and partially to an adjacent nucleotide sequence, e.g., a target nucleotide sequence. Such hybrid primers can increase the specificity of the amplification reaction.

[0075] As used herein, the selection of primers "so as to avoid substantial annealing to the target nucleic acids" means that primers are selected so that the majority of the amplicons detected after amplification are "full-length" in the sense that they result from priming at the expected sites at each end of the target nucleic acid, as opposed to amplicons resulting from priming within the target nucleic acid, which produces shorter-than-expected amplicons. In various embodiments, primers are selected to that at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% are full-length.

[0076] The term "primer pair" refers to a set of primers including a 5' "upstream primer" or "forward primer" that hybridizes with the complement of the 5' end of the DNA sequence to be amplified and a 3' "downstream primer" or "reverse primer" that hybridizes with the 3' end of the sequence to be amplified. As will be recognized by those of skill in the art, the terms "upstream" and "downstream" or "forward" and "reverse" are not intended to be limiting, but rather provide illustrative orientation in particular embodiments.

[0077] A "probe" is a nucleic acid capable of binding to a target nucleic acid of complementary sequence through one or more types of chemical bonds, generally through complementary base pairing, usually through hydrogen bond formation, thus forming a duplex structure. The probe binds or hybridizes to a "probe binding site." The probe can be labeled with a detectable label to permit facile detection of the probe, particularly once the probe has hybridized to its complementary target. Alternatively, however, the probe may be unlabeled, but may be detectable by specific binding with a ligand that is labeled, either directly or indirectly. Probes can vary significantly in size. Generally, probes are at least 7 to 15 nucleotides in length. Other probes are at least 20, 30, or 40 nucleotides long. Still other probes are

IDT Exhibit 1020.2644

somewhat longer, being at least 50, 60, 70, 80, or 90 nucleotides long. Yet other probes are longer still, and are at least 100, 150, 200 or more nucleotides long. Probes can also be of any length that is within any range bounded by any of the above values (e.g., 15-20 nucleotides in length).

[0078] The primer or probe can be perfectly complementary to the target nucleic acid sequence or can be less than perfectly complementary. In certain embodiments, the primer has at least 65% identity to the complement of the target nucleic acid sequence over a sequence of at least 7 nucleotides, more typically over a sequence in the range of 10-30 nucleotides, and often over a sequence of at least 14-25 nucleotides, and more often has at least 75% identity, at least 85% identity, at least 90% identity, or at least 95%, 96%, 97%. 98%, or 99% identity. It will be understood that certain bases (e.g., the 3' base of a primer) are generally desirably perfectly complementary to corresponding bases of the target nucleic acid sequence. Primer and probes typically anneal to the target sequence under stringent hybridization conditions.

[0079] The term "nucleotide tag" is used herein to refer to a predetermined nucleotide sequence that is added to a target nucleotide sequence. The nucleotide tag can encode an item of information about the target nucleotide sequence, such the identity of the target nucleotide sequence or the identity of the sample from which the target nucleotide sequence was derived. In certain embodiments, such information may be encoded in one or more nucleotide tags, e.g., a combination of two nucleotide tags, one on either end of a target nucleotide sequence, can encode the identity of the target nucleotide sequence.

[0080] As used herein the term "barcode primer" refers to a primer that includes a specific barcode nucleotide sequence that encodes information about the amplicon produced when the barcode primer is employed in an amplification reaction. For example, a different barcode primer can be employed to amplify one or more target sequences from each of a number of different samples, such that the barcode nucleotide sequence indicates the sample origin of the resulting amplicons.

[0081] As used herein, the term "encoding reaction" refers to reaction in which at least one nucleotide tag is added to a target nucleotide sequence. Nucleotide tags can be added, for example, by an "encoding PCR" in which the at least one

IDT Exhibit 1020.2645

primer comprises a target-specific portion and a nucleotide tag located on the 5' end of the target-specific portion, and a second primer that comprises only a target-specific portion or a target-specific portion and a nucleotide tag located on the 5' end of the target-specific portion. For illustrative examples of PCR protocols applicable to encoding PCR, see pending WO Application US03/37808 as well as U.S. Pat. No.6,605,451. Nucleotide tags can also be added by an "encoding ligation" reaction that can comprise a ligation reaction in which at least one primer comprises a target-specific portion and nucleotide tag located on the 5' end of the target-specific portion, and a second primer that comprises a target-specific portion only or a target-specific portion and a nucleotide tag located on the 5' end of the target specific portion. Illustrative encoding ligation reactions are described, for example, in U.S. Patent Publication No. 2005/0260640, which is hereby incorporated by reference in its entirety, and in particular for ligation reactions.

[0082] As used herein an "encoding reaction" produces a "tagged target nucleotide sequence," which includes a nucleotide tag linked to a target nucleotide sequence.

[0083] As used herein with reference to a portion of a primer, the term "target-specific" nucleotide sequence refers to a sequence that can specifically anneal to a target nucleic acid or a target nucleotide sequence under suitable annealing conditions.

[0084] As used herein with reference to a portion of a primer, the term "nucleotide tag-specific nucleotide sequence" refers to a sequence that can specifically anneal to a nucleotide tag under suitable annealing conditions.

[0085] Amplification according to the present teachings encompasses any means by which at least a part of at least one target nucleic acid is reproduced, typically in a template-dependent manner, including without limitation, a broad range of techniques for amplifying nucleic acid sequences, either linearly or exponentially. Illustrative means for performing an amplifying step include ligase chain reaction (LCR), ligase detection reaction (LDR), ligation followed by Q-replicase amplification, PCR, primer extension, strand displacement amplification (SDA), hyperbranched strand displacement amplification, multiple displacement amplification (MDA), nucleic acid strand-based amplification (NASBA), two-step

IDT Exhibit 1020.2646

multiplexed amplifications, rolling circle amplification (RCA), and the like, including multiplex versions and combinations thereof, for example but not limited to, OLA/PCR, PCR/OLA, LDR/PCR, PCR/PCR/LDR, PCR/LDR, LCR/PCR, PCR/LCR (also known as combined chain reaction--CCR), and the like. Descriptions of such techniques can be found in, among other sources, Ausbel et al.; PCR Primer: A Laboratory Manual, Diffenbach, Ed., Cold Spring Harbor Press (1995); The Electronic Protocol Book, Chang Bioscience (2002); Msuih et al., J. Clin. Micro. 34:501-07 (1996); The Nucleic Acid Protocols Handbook, R. Rapley, ed., Humana Press, Totowa, N.J. (2002); Abramson et al., Curr Opin Biotechnol. 1993 Feb.;4(1):41-7, U.S. Pat. No. 6,027,998; U.S. Pat. No. 6,605,451, Barany et al., PCT Publication No. WO 97/31256; Wenz et al., PCT Publication No. WO 01/92579; Day et al., Genomics, 29(1): 152-162 (1995), Ehrlich et al., Science 252:1643-50 (1991); Innis et al., PCR Protocols: A Guide to Methods and Applications, Academic Press (1990); Favis et al., Nature Biotechnology 18:561-64 (2000); and Rabenau et al., Infection 28:97-102 (2000); Belgrader, Barany, and Lubin, Development of a Multiplex Ligation Detection Reaction DNA Typing Assay, Sixth International Symposium on Human Identification, 1995 (available on the world wide web at: promega.com/geneticidproc/ussymp6proc/blegrad.html-); LCR Kit Instruction Manual, Cat. #200520, Rev. #050002, Stratagene, 2002; Barany, Proc. Natl. Acad. Sci. USA 88:188-93 (1991); Bi and Sambrook, Nucl. Acids Res. 25:2924-2951 (1997); Zirvi et al., Nucl. Acid Res. 27:e40i-viii (1999); Dean et al., Proc Natl Acad Sci USA 99:5261-66 (2002); Barany and Gelfand, Gene 109:1-11 (1991); Walker et al., Nucl. Acid Res. 20:1691-96 (1992); Polstra et al., BMC Inf. Dis. 2:18- (2002); Lage et al., Genome Res. 2003 Feb.;13(2):294-307, and Landegren et al., Science 241:1077-80 (1988), Demidov, V., Expert Rev Mol Diagn. 2002 Nov.;2(6):542-8., Cook et al., J Microbiol Methods. 2003 May;53(2):165-74, Schweitzer et al., Curr Opin Biotechnol. 2001 Feb.;12(1):21-7, U.S. Pat. No. 5,830,711, U.S. Pat. No. 6,027,889, U.S. Pat. No. 5,686,243, PCT Publication No. WO0056927A3, and PCT Publication No. WO9803673A1.

[0086] In some embodiments, amplification comprises at least one cycle of the sequential procedures of: annealing at least one primer with complementary or substantially complementary sequences in at least one target nucleic acid; synthesizing at least one strand of nucleotides in a template-dependent manner using a

IDT Exhibit 1020.2647

polymerase; and denaturing the newly-formed nucleic acid duplex to separate the strands. The cycle may or may not be repeated. Amplification can comprise thermocycling or can be performed isothermally.

[0087] The term "qPCR" is used herein to refer to quantitative real-time polymerase chain reaction (PCR), which is also known as "real-time PCR" or "kinetic polymerase chain reaction."

[0088] A "reagent" refers broadly to any agent used in a reaction, other than the analyte (e.g., nucleic acid being analyzed). Illustrative reagents for a nucleic acid amplification reaction include, but are not limited to, buffer, metal ions, polymerase, reverse transcriptase, primers, template nucleic acid, nucleotides, labels, dyes, nucleases, and the like. Reagents for enzyme reactions include, for example, substrates, cofactors, buffer, metal ions, inhibitors, and activators.

[0089] The term "universal detection probe" is used herein to refer to any probe that identifies the presence of an amplification product, regardless of the identity of the target nucleotide sequence present in the product.

[0090] The term "universal qPCR probe" is used herein to refer to any such probe that identifies the presence of an amplification product during qPCR. In particular embodiments, nucleotide tags according to the invention can comprise a nucleotide sequence to which a detection probe, such as a universal qPCR probe binds. Where a tag is added to both ends of a target nucleotide sequence, each tag can, if desired, include a sequence recognized by a detection probe. The combination of such sequences can encode information about the identity or sample source of the tagged target nucleotide sequence. In other embodiments, one or more amplification primers can comprise a nucleotide sequence to which a detection probe, such as a universal qPCR probe binds. In this manner, one, two, or more probe binding sites can be added to an amplification product during the amplification step of the methods of the invention. Those of skill in the art recognize that the possibility of introducing multiple probe binding sites during preamplification (if carried out) and amplification facilitates multiplex detection, wherein two or more different amplification products can be detected in a given amplification mixture or aliquot thereof.

[0091] The term "universal detection probe" is also intended to encompass primers labeled with a detectable label (e.g., a fluorescent label), as well as non-

IDT Exhibit 1020.2648

sequence-specific probes, such as DNA binding dyes, including double-stranded DNA (dsDNA) dyes, such as SYBR Green.

[0092] The term "target-specific qPCR probe" is used herein to refer to a qPCR probe that identifies the presence of an amplification product during qPCR, based on hybridization of the qPCR probe to a target nucleotide sequence present in the product.

[0093] "Hydrolysis probes" are generally described in U.S. Patent No. 5,210,015, which is incorporated herein by reference in its entirety for its description of hydrolysis probes. Hydrolysis probes take advantage of the 5'-nuclease activity present in the thermostable Taq polymerase enzyme typically used in the PCR reaction (TaqMan® probe technology, Applied Biosystems, Foster City CA). The hydrolysis probe is labeled with a fluorescent detector dye such as fluorescin, and an acceptor dye or quencher. In general, the fluorescent dye is covalently attached to the 5' end of the probe and the quencher is attached to the 3' end of the probe, and when the probe is intact, the fluorescence of the detector dye is quenched by fluorescence resonance energy transfer (FRET). The probe anneals downstream of one of the primers that defines one end of the target nucleic acid in a PCR reaction. Using the polymerase activity of the Taq enzyme, amplification of the target nucleic acid is directed by one primer that is upstream of the probe and a second primer that is downstream of the probe but anneals to the opposite strand of the target nucleic acid. As the upstream primer is extended, the Taq polymerase reaches the region where the labeled probe is annealed, recognizes the probe-template hybrid as a substrate, and hydrolyzes phosphodiester bonds of the probe. The hydrolysis reaction irrevocably releases the quenching effect of the quencher dye on the reporter dye, thus resulting in increasing detector fluorescence with each successive PCR cycle. In particular, hydrolysis probes suitable for use in the invention can be capable of detecting 8-mer or 9-mer motifs that are common in the human and other genomes and/or transcriptomes and can have a high T_m of about 70°C enabled by the use of linked nucleic acid (LNA) analogs.

[0094] The term "label," as used herein, refers to any atom or molecule that can be used to provide a detectable and/or quantifiable signal. In particular, the label can be attached, directly or indirectly, to a nucleic acid or protein. Suitable labels that can be attached to probes include, but are not limited to, radioisotopes, fluorophores,

IDT Exhibit 1020.2649

chromophores, mass labels, electron dense particles, magnetic particles, spin labels, molecules that emit chemiluminescence, electrochemically active molecules, enzymes, cofactors, and enzyme substrates.

[0095] The term "dye," as used herein, generally refers to any organic or inorganic molecule that absorbs electromagnetic radiation at a wavelength greater than or equal 340 nm.

[0096] The term "fluorescent dye," as used herein, generally refers to any dye that emits electromagnetic radiation of longer wavelength by a fluorescent mechanism upon irradiation by a source of electromagnetic radiation, such as a lamp, a photodiode, or a laser.

[0097] The term "elastomer" has the general meaning used in the art. Thus, for example, Allcock et al. (Contemporary Polymer Chemistry, 2nd Ed.) describes elastomers in general as polymers existing at a temperature between their glass transition temperature and liquefaction temperature. Elastomeric materials exhibit elastic properties because the polymer chains readily undergo torsional motion to permit uncoiling of the backbone chains in response to a force, with the backbone chains recoiling to assume the prior shape in the absence of the force. In general, elastomers deform when force is applied, but then return to their original shape when the force is removed.

[0098] A "polymorphic marker" or "polymorphic site" is a locus at which nucleotide sequence divergence occurs. Illustrative markers have at least two alleles, each occurring at frequency of greater than 1%, and more typically greater than 10% or 20% of a selected population. A polymorphic site may be as small as one base pair. Polymorphic markers include restriction fragment length polymorphism (RFLPs), variable number of tandem repeats (VNTR's), hypervariable regions, minisatellites, dinucleotide repeats, trinucleotide repeats, tetranucleotide repeats, simple sequence repeats, deletions, and insertion elements such as Alu. The first identified allelic form is arbitrarily designated as the reference form and other allelic forms are designated as alternative or variant alleles. The allelic form occurring most frequently in a selected population is sometimes referred to as the wildtype form. Diploid organisms may be homozygous or heterozygous for allelic forms. A diallelic polymorphism has two forms. A triallelic polymorphism has three forms.

IDT Exhibit 1020.2650

[0099] A "single nucleotide polymorphism" (SNP) occurs at a polymorphic site occupied by a single nucleotide, which is the site of variation between allelic sequences. The site is usually preceded by and followed by highly conserved sequences of the allele (e.g., sequences that vary in less than 1/100 or 1/1000 members of the populations). A SNP usually arises due to substitution of one nucleotide for another at the polymorphic site. A transition is the replacement of one purine by another purine or one pyrimidine by another pyrimidine. A transversion is the replacement of a purine by a pyrimidine or vice versa. SNPs can also arise from a deletion of a nucleotide or an insertion of a nucleotide relative to a reference allele.

Amplification Methods

In General

[0100] In particular embodiments, the invention provides an amplification method for introducing a plurality (e.g., at least three) of selected nucleotide sequences into one or more target nucleic acid(s). The method entails amplifying a plurality of target nucleic acids, typically, in a plurality of samples. In illustrative embodiments, the same set of target nucleic acids can be amplified in each of two or more different samples. The samples can differ from one another in any way, e.g., the samples can be from different tissues, subjects, environmental sources, etc. At least three primers can be used to amplify each target nucleic acid, namely: forward and reverse amplification primers, each primer including a target-specific portion and one or both primers including a nucleotide tag. The target-specific portions can specifically anneal to a target under suitable annealing conditions. The nucleotide tag for the forward primer can have a sequence that is the same as, or different from, the nucleotide tag for the reverse primer. Generally, the nucleotide tags are 5' of the target-specific portions. The third primer is a barcode primer comprising a barcode nucleotide sequence and a first and/or second nucleotide tag-specific portion. The barcode nucleotide sequence is a sequence selected to encode information about the amplicon produced when the barcode primer is employed in an amplification reaction. The tag-specific portion can specifically anneal to the one or both nucleotide tags in the forward and reverse primers. The barcode primer is generally 5' of the tag-specific portion.

IDT Exhibit 1020.2651

[0101] The barcode primer is typically present in the amplification mixture in excess of the forward and/or reverse primer(s). More specifically, if the barcode primer anneals to the nucleotide tag in the forward primer, the barcode primer is generally present in excess of the forward primer. If the barcode primer anneals to the nucleotide tag in the reverse primer, the barcode primer is generally present in excess of the reverse primer. In each instance the third primer in the amplification mixture, i.e., the reverse primer or the forward primer, respectively, can be present, in illustrative embodiments, at a concentration approximately similar to that of the barcode primer. Generally the barcode primer is present in substantial excess. For example, the concentration of the barcode primer in the amplification mixtures can be at least 2-fold, at least 4-fold, at least 5-fold, at least 10-fold, at least 15-fold, at least 20-fold, at least 25-fold, at least 30-fold, at least 35-fold, at least 40-fold, at least 45-fold, at least 50-fold, at least 100-fold, at least 500-fold, at least 10^3-fold, at least 5 $\times 10^3$-fold, at least 10^4-fold, at least 5 $\times 10^4$-fold, at least 10^5-fold, at least 5 $\times 10^5$-fold, at least 10^6-fold, or higher, relative to the concentration of the forward and/or reverse primer(s). In addition, the concentration excess of the barcode primer can fall within any range having any of the above values as endpoints (e.g., 2-fold to 10^5-fold). In illustrative embodiments, where the barcode primer has a tag-specific portion that is specific for the nucleotide tag on the forward primer, the forward primer can be present in picomolar to nanomolar concentrations, e.g., about 5 pM to 500 nM, about 5 pM to 100 nM, about 5 pM to 50 nM, about 5 pM to 10 nM, about 5 pM to 5 nM, about 10 pM to 1 nM, about 50 pM to about 500 pM, about 100 pM or any other range having any of these values as endpoints (e.g., 10 pM to 50 pM). Suitable, illustrative concentrations of barcode primer that could be used on combination with any of these concentrations of forward primer include about 10 nM to about 10 μM, about 25 nM to about 7.5 μM, about 50 nM to about 5 μM, about 75 nM to about 2.5 μM, about 100 nM to about 1 μM, about 250 nM to about 750 nM, about 500 nM or any other range having any of these values as endpoints (e.g., 100 nM to 500 nM). In amplification reactions using such concentrations of forward and barcode primers, the reverse primer have a concentration on the same order as the barcode primer (e.g. within about 10-fold, within about 5-fold, or equal).

[0102] Each amplification mixture can be subjected to amplification to produce target amplicons comprising tagged target nucleotide sequences, each

IDT Exhibit 1020.2652

comprising first and second nucleotide tags flanking the target nucleotide sequence, and at least one barcode nucleotide sequence at the 5' or 3' end of the target amplicon (relative to one strand of the target amplicon). In certain embodiments, the first and second nucleotide tags and/or the barcode nucleotide sequence are selected so as to avoid substantial annealing to the target nucleic acids. In such embodiments, the tagged target nucleotide sequences can include molecules having the following elements: 5'-(barcode nucleotide sequence)-(first nucleotide tag from the forward primer)-(target nucleotide sequence)-(second nucleotide tag sequence from the reverse primer)-3' or 5'-(first nucleotide tag from the forward primer)-(target nucleotide sequence)-(second nucleotide tag sequence from the reverse primer) -(barcode nucleotide sequence)-3'.

[0103] In illustrative embodiments, the barcode nucleotide sequence identifies a particular sample. Thus, for example, a set of T target nucleic acids can be amplified in each of S samples, where S and T are integers, typically greater than one. In such embodiments, amplification can be performed separately for each sample, wherein the same set of forward and reverse primers is used for each sample and the set of forward and reverse primers has at least one nucleotide tag that is common to all primers in the set. A different barcode primer can be used for each sample, wherein the bar code primers have different barcode nucleotide sequences, but the same tag-specific portion that can anneal to the common nucleotide tag. This embodiment has the advantage of reducing the number of different primers that would need to be synthesized to encode sample origin in amplicons produced for a plurality of target sequences. Alternatively, different sets of forward and reverse primers can be employed for each sample, wherein each set has a nucleotide tag that is different from the primers in the other set, and different barcode primers are used for each sample, wherein the barcode primers have different barcode nucleotide sequences and different tag-specific portions. In either case, the amplification produces a set of T amplicons from each sample that bear sample-specific barcodes.

[0104] In embodiments, wherein the same set of forward and reverse primers is used for each sample, the forward and reverse primers for each target can be initially combined separately from the sample, and each barcode primer can be initially combined with its corresponding sample. Aliquots of the initially combined forward and reverse primers can then be added to aliquots of the initially combined

IDT Exhibit 1020.2653

sample and barcode primer to produce S×T amplification mixtures. These amplification mixtures can be formed in any article that can be subjected to conditions suitable for amplification. For example, the amplification mixtures can be formed in, or distributed into, separate compartments of a microfluidic device prior to amplification. Suitable microfluidic devices include, in illustrative embodiments, matrix-type microfluidic devices, such as those described below.

[0105] Any amplification method can be employed to produce amplicons from the amplification mixtures. In illustrative embodiments, PCR is employed. The amplification is generally carried out for at least three cycles to introduce the first and second nucleotide tags and the barcode nucleotide sequence. In various embodiments, amplification is carried out for 5, 10, 15, 20, 25, 30, 35, 40, 45, or 50 cycles, or for any number of cycles falling within a range having any of these values as endpoints (e.g. 5-10 cycles). In particular embodiments, amplification is carried out for a sufficient number of cycles to normalize target amplicon copy number across targets and across samples (e.g., 15, 20, 25, 30, 35, 40, 45, or 50 cycles, or for any number of cycles falling within a range having any of these values as endpoints).

[0106] Particular embodiments of the above-described method provide substantially uniform amplification, yielding a plurality of target amplicons wherein the majority of amplicons are present at a level relatively close to the average copy number calculated for the plurality of target amplicons. Thus, in various embodiments, at least 50, at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least 91, at least 92, at least 93, at least 94, at least 95, at least 96, at least 97, at least 98, or at least 99 percent of the target amplicons are present at greater than 50 percent of the average number of copies of target amplicons and less than 2-fold the average number of copies of target amplicons.

[0107] The invention also provides, in certain embodiments, a method for amplifying a plurality of target nucleotides in which barcoding is, optionally, omitted and the target nucleotide sequences are tagged after amplification. More specifically, the invention provides a method for amplifying a plurality of target nucleic acids, typically, in a plurality of samples, that entails preparing an amplification mixture for each target nucleic acid. Each amplification mixture includes a forward primer including a target-specific sequence and a reverse primer including a target-specific sequence. The amplification mixtures are subjected to amplification to produce a

IDT Exhibit 1020.2654

plurality of target nucleotide sequences. The target nucleotide sequences are then tagged to produce a plurality of target amplicons, each including first and/or second nucleotide tags flanking the target nucleotide sequence. This method produces a plurality of target amplicons, wherein at least 50 percent of the target amplicons are present at greater than 50 percent of the average number of copies of target amplicons and less than 2-fold the average number of copies of target amplicons. In various embodiments of this method at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least 91, at least 92, at least 93, at least 94, at least 95, at least 96, at least 97, at least 98, or at least 99 percent of the target amplicons are present at greater than 50 percent of the average number of copies of target amplicons and less than 2-fold the average number of copies of target amplicons.

[0108] In various embodiments, the target nucleotide sequence amplified can be, e.g., 25 bases, 50 bases, 100 bases, 200 bases, 500 bases, or 750 bases. In certain embodiments of the above-described methods, a long-range amplification method, such as long-range PCR can be employed to produce amplicons from the amplification mixtures. Long-range PCR permits the amplification of target nucleotide sequences ranging from one or a few kilobases (kb) to over 50 kb. In various embodiments, the target nucleotide sequences that are amplified by long-range PCR are at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, or 50 kb in length. Target nucleotide sequences can also fall within any range having any of these values as endpoints (e.g., 25 bases to 100 bases or 5-15 kb). The use of long-range PCR in the above-described methods can, in some embodiments, yield a plurality of target amplicons wherein at least 50, at least 55, at least 60, at least 65, or at least 70 percent of the target amplicons are present at greater than 50 percent of the average number of copies of target amplicons and less than 2-fold the average number of copies of target amplicons.

[0109] Long-range PCR is well known in the art. See, e.g., Cheng S, Fockler C, Barnes WM, Higuchi R (June 1994). "Effective amplification of long targets from cloned inserts and human genomic DNA". Proc. Natl. Acad. Sci. U.S.A. 91 (12): 5695–9. Enzymes, protocols, and kits for long-range PCR that are suitable for use in the methods described here are commercially available; examples include: SequalPrep™ Long PCR Kit (Invitrogen, USA), PfuUltra® II Fusion HS DNA

IDT Exhibit 1020.2655

polymerase (Stratagene), Phusion® DNA polymerases, Phusion® Flash High Fidelity PCR Master Mix (Finnzymes).

[0110] In certain embodiments, the target amplicons can be recovered from the amplification mixtures. For example, a matrix-type microfluidic device that is adapted to permit recovery of the contents of each reaction chamber (see below) can be employed for the amplification to generate the target amplicons. In variations of these embodiments, the target amplicons can be subjected to further amplification and/or analysis. For example, one or more target amplicon(s) can be subjected to amplification using primers specific for the first and second nucleotide tags to produce a target amplicon lacking the barcode nucleotide sequence. In certain embodiments, the amount of target amplicons produced in the amplification mixtures can be quantified during amplication, e.g., by quantitative real-time PCR, or after.

[0111] In particular embodiments, the above-described amplification methods are employed to produce amplicons suitable for automated DNA sequencing. In particular, the ability of the methods to provide substantially uniform amplification, as described above, of target nucleotide sequences is helpful in preparing DNA sequencing libraries having good coverage. In the context of automated DNA sequencing, the term "coverage" refers to the number of times the sequence is measured upon sequencing. A DNA sequencing library that has substantially uniform coverage can yield sequence data where the coverage is also substantially uniform. Thus, in various embodiments, upon performing automated sequencing of a plurality of target amplicons prepared as described herein, the sequences of at least 50 percent of the target amplicons are present at greater than 50 percent of the average number of copies of target amplicon sequences and less than 2-fold the average number of copies of target amplicon sequences. In various embodiments of this method at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least 91, at least 92, at least 93, at least 94, at least 95, at least 96, at least 97, at least 98, or at least 99 percent of the target amplicon sequences are present at greater than 50 percent of the average number of copies of target amplicon sequences and less than 2-fold the average number of copies of target amplicon sequences.

IDT Exhibit 1020.2656

Preparation of Nucleic Acids for DNA Sequencing

[0112] Many current DNA sequencing techniques rely on "sequencing by synthesis." These techniques entail library creation, massively parallel PCR amplification of library molecules, and sequencing. Library creation starts with conversion of sample nucleic acids to appropriately sized fragments, ligation of adaptor sequences onto the ends of the fragments, and selection for molecules properly appended with adaptors. The presence of the adaptor sequences on the ends of the library molecules enables amplification of random-sequence inserts. The above-described methods for tagging nucleotide sequences can be substituted for ligation, to introduce adaptor sequences, as described in greater detail below.

[0113] In particular embodiments, the number of library DNA molecules produced in the massively parallel PCR step is low enough that the chance of two molecules associating with the same substrate, e.g. the same bead (in 454 DNA sequencing) or the same surface patch (in Solexa DNA sequencing) is low, but high enough so that the yield of amplified sequences is sufficient to provide a high throughput. As discussed further below, after suitable adaptor sequences are introduced, digital PCR can be employed to calibrate the number of library DNA molecules prior to sequencing by synthesis.

Addition of DNA Sequencing Primers to Nucleic Acids

[0114] The DNA to be sequenced can be any type of DNA. In particular embodiments, the DNA is genomic DNA from an organism. In variations of such embodiments, total genomic DNA obtained from a sample taken from an organism or from a DNA library is prepared for sequencing.

[0115] As described above, at least three primers are employed to prepare the DNA for sequencing: forward, reverse, and barcode primers. However, one or more of the forward primer, reverse primer, and barcode primer includes at least one additional primer binding site. In specific embodiments, the barcode primer includes at least a first additional primer binding site upstream of the barcode nucleotide sequence, which is upstream of the first nucleotide tag-specific portion. In certain embodiments, two of the forward primer, reverse primer, and barcode primer include at least one additional primer binding site (i.e, such that the amplicon produced upon amplification includes the nucleotide tag sequences, the barcode nucleotide sequence,

IDT Exhibit 1020.2657

and the two additional binding sites). For example, if the barcode primer includes a first additional primer binding site upstream of the barcode nucleotide sequence, in specific embodiments, the reverse primer can include at least a second additional primer binding site downstream of the second nucleotide tag. Amplification then yields a molecule having the following elements: 5'-(first additional primer binding site)-(barcode nucleotide sequence)-(first nucleotide tag from the forward primer)-(target nucleotide sequence)-(second nucleotide tag from the reverse primer)-(second additional primer binding site)-3'. In specific embodiments, the first and second additional primer binding sites are capable of being bound by DNA sequencing primers, to facilitate sequencing of the entire amplicon, including the barcode, which can, as discussed above, indicate sample origin.

[0116] In some embodiments, more than three primers can be employed to add desired elements to a target nucleotide sequence. For example, four primers can be employed to produce molecules having the same five elements discussed above, plus an optional additional barcode e.g., 5'-(first additional primer binding site)-(barcode nucleotide sequence)-(first nucleotide tag from the forward primer)-(target nucleotide sequence)-(second nucleotide tag from the reverse primer)-(additional barcode nucleotide sequence)-(second additional primer binding site)-3'. In an illustrative four-primer embodiment, the forward primer includes a target-specific portion and first nucleotide tag, and the reverse primer includes a target-specific portion and a second nucleotide tag. Together, these two primers constitute the "inner primers." The remaining two primers are the "outer primers," which anneal to the first and second nucleotide tags present in the inner primers. One outer primer is the barcode primer, which can contain at least a first additional primer binding site upstream of the barcode nucleotide sequence, which is upstream of the first nucleotide tag-specific portion (i.e., the same barcode primer discussed in the previous paragraph). The second outer primer can include a second tag-specific portion, an additional barcode nucleotide sequence and, downstream of this, a second additional primer binding site.

[0117] Amplification to incorporate elements from more than three primers can be carried out in one or multiple amplification reactions. For example, a four-primer amplification can be carried out in one amplification reaction, in which all four primers are present. Alternatively, a four-primer amplification can be carried out,

IDT Exhibit 1020.2658

e.g., in two amplification reactions: one to incorporate the inner primers and a separate amplification reaction to incorporate the outer primers. Where all four primers are present in one amplification reaction, the outer primers are generally present in the reaction mixture in excess. The relative concentration values give above for the barcode primer relative to the forward and/or reverse primers also apply to the relative concentrations of the outer primers relative to inner primers in a one-step, four-primer amplification reaction.

[0118] In an illustrative embodiment of the four-primer amplification reaction, each of the outer primers contains a unique barcode. For example, one barcode primer would be constructed of the elements 5'-(first additional primer binding site)-(first barcode nucleotide sequence)-(first nucleotide tag)-3', and the second barcode primer would be constructed of the elements 5'-(second additional primer binding site)-(second barcode nucleotide sequence)-(second nucleotide tag)-3'. In this embodiment, a number (J) of first barcode primers can be combined with a number (K) of second barcode primers to create JxK unique amplification products.

[0119] In a further illustrative embodiment of the invention, more than four primers can be combined in a single reaction to append different combinations of additional primer binding sites, barcode sequences, and nucleotide tags. For example, outer barcode primers containing the following elements: 5'-(first additional primer binding site)-(first barcode nucleotide sequence)-(first nucleotide tag)-3', 5'-(first additional primer binding site)-(first barcode nucleotide sequence)-(second nucleotide tag)-3', 5'-(second additional primer binding site)-(first barcode nucleotide sequence)-(first nucleotide tag)-3', 5'-(second additional primer binding site)-(first barcode nucleotide sequence)-(second nucleotide tag)-3', can be combined with inner target-specific primers as described above to produce amplification product pools containing all combinations of the barcode primers with the desired amplicon sequence.

[0120] In other illustrative embodiments of the invention, outer barcode primers in any of the combinations described above, or other combinations that would be obvious to one of skill in the art, can be combined with more than one pair of target primer sequences bearing the same first and second nucleotide tag sequences. For example, inner primers containing up to ten different target-specific forward primer sequences combined with the same first nucleotide tag and up to ten different

IDT Exhibit 1020.2659

target-specific reverse primer sequences combined with the same second nucleotide tag can be combined with the up to 2 or up to 4 outer barcode primers to generate multiple amplification products as described above. In various embodiments, at least 10, at least 20, at least 50, at least 100, at least 200, at least 500, at least 1000, at least 2000, at least 5000 or at least 10000 different target-specific primer pairs bearing the same first nucleotide tag and second nucleotide tag would be combined with the up to 2 or up to 4 outer barcode primers to generate multiple amplification products.

[0121] The methods of the invention can include subjecting at least one target amplicon to DNA sequencing using any available DNA sequencing method. In particular embodiments, a plurality of target amplicons is sequenced using a high throughput sequencing method. Such methods typically use an in vitro cloning step to amplify individual DNA molecules. Emulsion PCR (emPCR) isolates individual DNA molecules along with primer-coated beads in aqueous droplets within an oil phase. PCR produces copies of the DNA molecule, which bind to primers on the bead, followed by immobilization for later sequencing. emPCR is used in the methods by Marguilis et al. (commercialized by 454 Life Sciences, Branford, CT), Shendure and Porreca et al. (also known as "polony sequencing") and SOLiD sequencing, (Applied Biosystems Inc., Foster City, CA). See M. Margulies, et al. (2005) "Genome sequencing in microfabricated high-density picolitre reactors" Nature 437: 376–380; J. Shendure, et al. (2005) "Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome" Science 309 (5741): 1728–1732. In vitro clonal amplification can also be carried out by "bridge PCR," where fragments are amplified upon primers attached to a solid surface. Braslavsky et al. developed a single-molecule method (commercialized by Helicos Biosciences Corp., Cambridge, MA) that omits this amplification step, directly fixing DNA molecules to a surface. I. Braslavsky, et al. (2003) "Sequence information can be obtained from single DNA molecules" Proceedings of the National Academy of Sciences of the United States of America 100: 3960–3964.

[0122] DNA molecules that are physically bound to a surface can be sequenced in parallel. "Sequencing by synthesis," like dye-termination electrophoretic sequencing, uses a DNA polymerase to determine the base sequence. Reversible terminator methods (commercialized by Illumina, Inc., San Diego, CA and Helicos Biosciences Corp., Cambridge, MA) use reversible versions of dye-

IDT Exhibit 1020.2660

terminators, adding one nucleotide at a time, and detect fluorescence at each position in real time, by repeated removal of the blocking group to allow polymerization of another nucleotide. "Pyrosequencing" also uses DNA polymerization, adding one nucleotide at a time and detecting and quantifying the number of nucleotides added to a given location through the light emitted by the release of attached pyrophosphates (commercialized by 454 Life Sciences, Branford, CT). See M. Ronaghi, et al. (1996). "Real-time DNA sequencing using detection of pyrophosphate release" Analytical Biochemistry 242: 84-89.

Sample Preparation by Digital PCR

[0123] In some embodiments, samples are loaded into an amplification device, for example, a PCR plate or a microfluidic device, at sample concentrations containing on average less than one amplification template per well or chamber. Each well or chamber in the device is prepared such that it contains suitable tagged target-specific primers and a unique combination of forward and reverse barcode primers. For example, one well can contain barcode primers containing the elements 5'-(first additional primer binding site)-(first barcode sequence)-(first nucleotide tag)-3', 5'-(second additional primer binding site)-(second barcode sequence)-(second nucleotide tag)-3'. A second well or chamber can contain barcode primers containing the elements 5'-(first additional primer binding site)-(third barcode sequence)-(first nucleotide tag)-3', 5'-(second additional primer binding site)-(fourth barcode sequence)-(second nucleotide tag)-3'. Amplification products produced in each well would be labeled uniquely with the combinations of barcode sequences loaded into these wells.

Sample Calibration by Digital PCR

[0124] In particular embodiments, the number of target amplicons produced, e.g. from a DNA library, using the above-described methods can be calibrated using a digital amplification method. The step is finds particular application in preparing DNA for sequencing by synthesis. For discussions of "digital PCR" see, for example, Vogelstein and Kinzler, 1999, *Proc Natl Acad Sci USA* 96:9236-41; McBride et al., U.S Patent Application Publication No. 20050252773, especially Example 5 (each of these publications are hereby incorporated by reference in their entirety, and in particular for their disclosures of digital amplification). Digital amplification methods

IDT Exhibit 1020.2661

can make use of certain-high-throughput devices suitable for digital PCR, such as microfluidic devices typically including a large number and/or high density of small-volume reaction sites (e.g., nano-volume reaction sites or reaction chambers). In illustrative embodiments, digital amplification is performed using a microfluidic device, such as the Digital Array microfluidic devices described below. Digital amplification can entail distributing or partitioning a sample among hundreds to thousands of reaction mixtures disposed in a reaction/assay platform or microfluidic device. In such embodiments, a limiting dilution of the sample is made across a large number of separate amplification reactions such that most of the reactions have no template molecules and give a negative amplification result. In counting the number of positive amplification results, e.g, at the reaction endpoint, one is counting the individual template molecules present in the input sample one-by-one. A major advantage of digital amplification is that the quantification is independent of variations in the amplification efficiency – successful amplifications are counted as one molecule, independent of the actual amount of product.

[0125] In certain embodiments, digital amplification can be carried out after preamplification of sample nucleic acids. Typically, preamplification prior to digital amplification is performed for a limited number of thermal cycles (e.g., 5 cycles, or 10 cycles). In certain embodiments, the number of thermal cycles during preamplification can range from about 4 to 15 thermal cycles, or about 4-10 thermal cycles. In certain embodiments the number of thermal cycles can be 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or more than 15. The above-described amplification to produce adaptor sequence-containing amplicons for DNA sequencing can be substituted for the typical preamplification step.

[0126] Digital amplication methods are described in U.S. Publication No. 20090239308, published September 24, 2009, which is hereby incorporated by reference in its entirety and, in particular, for its disclosure of digital amplification methods and devices. Generally, in digital amplification, identical (or substantially similar) amplification reactions are run on a nucleic acid sample, such as genomic DNA. The number of individual reactions for a given nucleic acid sample may vary from about 2 to over 1,000,000. Typically, the number of reactions performed on a sample is about 100 or greater, more typically about 200 or greater, and even more typically about 300 or greater. Larger scale digital amplification can also be

IDT Exhibit 1020.2662

performed in which the number of reactions performed on a sample is about 500 or greater, about 700 or greater, about 765 or greater, about 1,000 or greater, about 2,500 or greater, about 5,000 or greater, about 7,500 or greater, or about 10,000or greater. The number of reactions performed may also be significantly higher, such up to about 25,000, up to about 50,000, up to about 75,000, up to about 100,000, up to about 250,000, up to about 500,000, up to about 750,000, up to about 1,000,000, or even greater than 1,000,000 assays per genomic sample.

[0127] In particular embodiments, the quantity of nucleic acid subjected to digital amplification is generally selected such that, when distributed into discrete reaction mixtures, each individual amplification reaction is expected to include one or fewer amplifiable nucleic acids. One of skill in the art can determine the concentration of target amplicon(s) produced as described above and calculate an appropriate amount for use in digital amplification. More conveniently, a set of serial dilutions of the target amplicon(s) can be tested. For example, a device that is commercially available from Fluidigm Corp. as the 12.765 Digital Array microfluidic device allows 12 different dilutions to be tested simultaneously. Optionally, a suitable dilution can be determined by generating a linear regression plot. For the optimal dilution, the line should be straight and pass through the origin. Subsequently the concentration of the original samples can be calculated from the plot.

[0128] The appropriate quantity of target amplicon(s) can be distributed into discrete locations or reaction wells or chambers such that each reaction includes, for example, an average of no more than about one amplicon per volume. The target amplicon(s) can be combined with reagents selected for quantitative or nonquantitative amplification, prior to distribution or after.

[0129] Following distribution, the reaction mixtures are subjected to amplification to identify those reaction mixtures that contained a target amplicon. Any amplification method can be employed, but conveniently, PCR is used, e.g., real-time PCR or endpoint PCR. This amplification can employ any primers capable of amplifying the target amplicon(s). Thus, in particular embodiments, the primers can be DNA sequencing primers that anneal to the primer binding sites introduced in the previous amplification step.

IDT Exhibit 1020.2663

[0130] The concentration of any target amplicon (copies/μL) is correlated with the number of positive (i.e., amplification product-containing) reaction mixtures. See copending U.S. Application No. 12/170,414, entitled "Method and Apparatus for Determining Copy Number Variation Using Digital PCR," which is incorporated by reference for all purposes, and, in particular, for analysis of digital PCR results. Also see Dube et al., 2008, "Mathematical Analysis of Copy Number Variation in a DNA Sample Using Digital PCR on a Nanofluidic Device" PLoS ONE 3(8): e2876. doi:10.1371/journal.pone.0002876, which is incorporated by reference for all purposes and, in particular, for analysis of digital PCR results.

[0131] In an illustrative embodiment of sample calibration for DNA sequencing by digital PCR, a PCR reaction mix containing roughly 100-360 amplicons per μl can be loaded onto a Digital Array microfluidic device, such as Fluidigm Corporation's (South San Francisco, CA) 12.765 Digital Array microfluidic device, described below. The microfluidic chip has 12 panels and each panel contains 765 chambers. Replicate panels on the digital chip can be assayed in order to obtain absolute quantification of the initial concentration of library. The diluted samples having typical relative coefficients of variation (between replicates) within 9-12% (or lower) can be used for sequencing. See. e.g., White III RA, Blainey PC, Fan CH, Quake SR. "Digital PCR provides sensitive and absolute calibration for high throughput sequencing" BMC Genomics 10:116 doi:10.1186/1471-2164-10-116.

Sample Nucleic Acids

[0132] Preparations of nucleic acids ("samples") can be obtained from biological sources and prepared using conventional methods known in the art. In particular, DNA or RNA useful in the methods described herein can be extracted and/or amplified from any source, including bacteria, protozoa, fungi, viruses, organelles, as well higher organisms such as plants or animals, particularly mammals, and more particularly humans. Suitable nucleic acids can also be obtained from environmental sources (e.g., pond water), from man-made products (e.g., food), from forensic samples, and the like. Nucleic acids can be extracted or amplified from cells, bodily fluids (e.g., blood, a blood fraction, urine, etc.), or tissue samples by any of a variety of standard techniques. Illustrative samples include samples of plasma, serum, spinal fluid, lymph fluid, peritoneal fluid, pleural fluid, oral fluid, and external

IDT Exhibit 1020.2664

sections of the skin; samples from the respiratory, intestinal genital, and urinary tracts; samples of tears, saliva, blood cells, stem cells, or tumors. For example, samples of fetal DNA can be obtained from an embryo or from maternal blood. Samples can be obtained from live or dead organisms or from in vitro cultures. Illustrative samples can include single cells, paraffin-embedded tissue samples, and needle biopsies. Nucleic acids useful in the invention can also be derived from one or more nucleic acid libraries, including cDNA, cosmid, YAC, BAC, P1, PAC libraries, and the like.

[0133] Nucleic acids of interest can be isolated using methods well known in the art, with the choice of a specific method depending on the source, the nature of nucleic acid, and similar factors. The sample nucleic acids need not be in pure form, but are typically sufficiently pure to allow the amplification steps of the methods of the invention to be performed. Where the target nucleic acids are RNA, the RNA can be reversed transcribed into cDNA by standard methods known in the art and as described in Sambrook, J., Fritsch, E.F., and Maniatis, T., *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, NY, Vol. 1, 2, 3 (1989), for example. The cDNA can then be analyzed according to the methods of the invention.

Target Nucleic Acids

[0134] Any target nucleic acid that can be tagged in an encoding reaction of the invention (described herein) can be detected using the methods of the invention. In typical embodiments, at least some nucleotide sequence information will be known for the target nucleic acids. For example, if the encoding reaction employed is PCR, sufficient sequence information is generally available for each end of a given target nucleic acid to permit design of suitable amplification primers. In an alternative embodiment, the target-specific sequences in primers could be replaced by random or degenerate nucleotide sequences.

[0135] The targets can include, for example, nucleic acids associated with pathogens, such as viruses, bacteria, protozoa, or fungi; RNAs, e.g., those for which over- or under-expression is indicative of disease, those that are expressed in a tissue- or developmental-specific manner; or those that are induced by particular stimuli; genomic DNA, which can be analyzed for specific polymorphisms (such as SNPs), alleles, or haplotypes, e.g., in genotyping. Of particular interest are genomic DNAs

IDT Exhibit 1020.2665

that are altered (e.g., amplified, deleted, and/or mutated) in genetic diseases or other pathologies; sequences that are associated with desirable or undesirable traits; and/or sequences that uniquely identify an individual (e.g., in forensic or paternity determinations).

Primer Design

[0136] Primers suitable for nucleic acid amplification are sufficiently long to prime the synthesis of extension products in the presence of the agent for polymerization. The exact length and composition of the primer will depend on many factors, including, for example, temperature of the annealing reaction, source and composition of the primer, and where a probe is employed, proximity of the probe annealing site to the primer annealing site and ratio of primer:probe concentration. For example, depending on the complexity of the target nucleic acid sequence, an oligonucleotide primer typically contains in the range of about 15 to about 30 nucleotides, although it may contain more or fewer nucleotides. The primers should be sufficiently complementary to selectively anneal to their respective strands and form stable duplexes. One skilled in the art knows how to select appropriate primer pairs to amplify the target nucleic acid of interest.

[0137] For example, PCR primers can be designed by using any commercially available software or open source software, such as Primer3 (*see, e.g.,* Rozen and Skaletsky (2000) *Meth. Mol. Biol.,* 132: 365-386; www.broad.mit.edu/node/1060, and the like) or by accessing the Roche UPL website. The amplicon sequences are input into the Primer3 program with the UPL probe sequences in brackets to ensure that the Primer3 program will design primers on either side of the bracketed probe sequence.

[0138] Primers may be prepared by any suitable method, including, for example, cloning and restriction of appropriate sequences or direct chemical synthesis by methods such as the phosphotriester method of Narang *et al.* (1979) *Meth. Enzymol.* 68: 90-99; the phosphodiester method of Brown *et al.* (1979) *Meth. Enzymol.* 68: 109-151; the diethylphosphoramidite method of Beaucage *et al.* (1981) *Tetra. Lett.*, 22: 1859-1862; the solid support method of U.S. Patent No. 4,458,066 and the like, or can be provided from a commercial source.

IDT Exhibit 1020.2666

[0139] Primers may be purified by using a Sephadex column (Amersham Biosciences, Inc., Piscataway, NJ) or other methods known to those skilled in the art. Primer purification may improve the sensitivity of the methods of the invention.

Microfluidic Devices

[0140] In certain embodiments, any of the methods of the invention can be carried out using a microfluidic device. In illustrative embodiments, the device is a matrix-type microfluidic device is one that allows the simultaneous combination of a plurality of substrate solutions with reagent solutions in separate isolated reaction chambers. It will be recognized, that a substrate solution can comprise one or a plurality of substrates and a reagent solution can comprise one or a plurality of reagents. For example, the microfluidic device can allow the simultaneous pair-wise combination of a plurality of different amplification primers and samples. In certain embodiments, the device is configured to contain a different combination of primers and samples in each of the different chambers. In various embodiments, the number of separate reaction chambers can be greater than 50, usually greater than 100, more often greater than 500, even more often greater than 1000, and sometimes greater than 5000, or greater than 10,000.

[0141] In particular embodiments, the matrix-type microfluidic device is a Dynamic Array ("DA") microfluidic device, an example of which is shown in Fig. 1. A DA microfluidic device is a matrix-type microfluidic device designed to isolate pair-wise combinations of samples and reagents (*e.g.*, amplification primers, detection probes, *etc.*) and suited for carrying out qualitative and quantitative PCR reactions including real-time quantitative PCR analysis. In some embodiments, the DA microfluidic device is fabricated, at least in part, from an elastomer. DA microfluidic devices are described in PCT publication WO05107938A2 (Thermal Reaction Device and Method For Using The Same) and US Pat. Publication US20050252773A1, both incorporated herein by reference in their entireties for their descriptions of DA microfluidic devices. DA microfluidic devices may incorporate high-density matrix designs that utilize fluid communication vias between layers of the microfluidic device to weave control lines and fluid lines through the device and between layers. By virtue of fluid lines in multiple layers of an elastomeric block, high density reaction cell arrangements are possible. Alternatively DA microfluidic devices may

IDT Exhibit 1020.2667

be designed so that all of the reagent and sample channels are in the same elastomeric layer, with control channels in a different layer.

[0142] U.S. Patent Publication No. 2008/0223721 and PCT Publication No. WO 05/107938A2 describe illustrative matrix-type devices that can be used to practice the methods described herein. Figure 21 of the latter is reproduced as Fig. 1 and shows an illustrative matrix design having a first elastomeric layer 2110 (1st layer) and a second elastomeric layer 2120 (2d layer) each having fluid channels formed therein. For example, a reagent fluid channel in the first layer 2110 is connected to a reagent fluid channel in the second layer 2120 through a via 2130, while the second layer 2120 also has sample channels therein, the sample channels and the reagent channels terminating in sample and reagent chambers 2180, respectively. The sample and reagent chambers 2180 are in fluid communication with each other through an interface channel 2150 that has an interface valve 2140 associated therewith to control fluid communication between each of the chambers 2180 of a reaction cell 2160. In use, the interface is first closed, then reagent is introduced into the reagent channel from the reagent inlet and sample is introduced into the sample channel through the sample inlet; containment valves 2170 are then closed to isolate each reaction cell 2160 from other reaction cells 2160. Once the reaction cells 2160 are isolated, the interface valve 2140 is opened to cause the sample chamber and the reagent chamber to be in fluid communication with each other so that a desired reaction may take place. It will be apparent from this (and the description in WO 05/107938A2) that the DA microfluidic device may be used for reacting M number of different samples with N number of different reagents.

[0143] Although the DA microfluidic devices described above in WO 05/107938 are well suited for conducting the methods described herein, the invention is not limited to any particular device or design. Any device that partitions a sample and/or allows independent pair-wise combinations of reagents and sample may be used. U.S. Patent Publication No. 20080108063 (which is hereby incorporated by reference it its entirety) includes a diagram illustrating the 48.48 Dynamic Array IFC (Integrated Fluidic Circuit), a commercially available device available from Fluidigm Corp. (South San Francisco Calif.). It will be understood that other configurations are possible and contemplated such as, for example, 48×96; 96×96; 30×120; *etc.*

IDT Exhibit 1020.2668

[0144] In specific embodiments, the microfluidic device can be a Digital Array microfluidic device, which is adapted to perform digital amplification. Such devices can have integrated channels and valves that partition mixtures of sample and reagents into nanolitre volume reaction chambers. In some embodiments, the Digital Array microfluidic device is fabricated, at least in part, from an elastomer. Illustrative Digital Array microfluidic devices are described in copending U.S. Applications owned by Fluidigm, Inc., such as U.S. Application No. 12/170,414, entitled "Method and Apparatus for Determining Copy Number Variation Using Digital PCR." One illustrative embodiment has 12 input ports corresponding to 12 separate sample inputs to the device. The device can have 12 panels, and each of the 12 panels can contain 765 6 nL reaction chambers with a total volume of 4.59 μL per panel. Microfluidic channels can connect the various reaction chambers on the panels to fluid sources. Pressure can be applied to an accumulator in order to open and close valves connecting the reaction chambers to fluid sources. In illustrative embodiments, 12 inlets can be provided for loading of the sample reagent mixture. 48 inlets can be used to provide a source for reagents, which are supplied to the biochip when pressure is applied to accumulator. Additionally, two or more inlets can be provided to provide hydration to the biochip. Hydration inlets are in fluid communication with the device to facilitate the control of humidity associated with the reaction chambers. As will be understood to one of skill in the art, some elastomeric materials that can utilized in the fabrication of the device are gas permeable, allowing evaporated gases or vapor from the reaction chambers to pass through the elastomeric material into the surrounding atmosphere. In a particular embodiment, fluid lines located at peripheral portions of the device provide a shield of hydration liquid, for example, a buffer or master mix, at peripheral portions of the biochip surrounding the panels of reaction chambers, thus reducing or preventing evaporation of liquids present in the reaction chambers. Thus, humidity at peripheral portions of the device can be increased by adding a volatile liquid, for example water, to hydration inlets. In a specific embodiment, a first inlet is in fluid communication with the hydration fluid lines surrounding the panels on a first side of the biochip and the second inlet is in fluid communication with the hydration fluid lines surrounding the panels on the other side of the biochip.

IDT Exhibit 1020.2669