

WHAT IS CLAIMED IS:

1. An artificial transposon, comprising:
 - a first transposase recognition site,
 - a second transposase recognition site, and
 - 5 a barcode disposed therebetween, wherein the barcode comprises a double-stranded nucleic acid sequence comprising
 - a first strand barcode and
 - a second strand barcode.
- 10 2. The transposon of claim 1, wherein the transposase is a hyperactive Tn5 transposase.
3. The transposon of claim 1, wherein the transposase is a Mu transposase.
- 15 4. The transposon of claim 1, wherein the transposase is an IS5 or an IS911 transposase.
5. The transposon of claim 1, wherein the transposon is associated with an affinity tag.
- 20 6. The transposon of claim 5, wherein the affinity tag is associated with a protein that targets specific nucleic acid sequences.
7. The transposon of claim 1, wherein the first strand barcode and the second strand barcode comprise complementary sequences.
- 25 8. The transposon of claim 1, wherein the first strand barcode and the second strand barcode comprise non-complementary sequences.
- 30 9. The transposon of claim 1, further comprising at least one universal primer site.

10. An isolated template nucleic acid, comprising at least a portion of a target nucleic acid or copy thereof and at least first and second transposons of claim 1, wherein the barcode of the first transposon is different from the barcode of the second transposon.
- 5
11. A method for preparing a library of template nucleic acids comprising:
- (a) contacting a target nucleic acid with a plurality of transposon of claim 1 under conditions such that a portion of said of transposons are inserted into the target nucleic acid; and
- 10 (b) fragmenting said target nucleic acid.
12. The method of claim 11, wherein the target nucleic acid comprises cDNAs from a single cell.
- 15 13. The method of claim 11, wherein the target nucleic acid comprises nucleic acids from multiple species.
14. The method of claim 11, wherein the target nucleic acid comprises nucleic acids from multiple haplotypes.
- 20 15. The method of claim 11, wherein the fragmentation comprises contacting the target nucleic acid with a polymerase.
16. The method of claim 15, wherein the polymerase has 3' to 5' exonuclease activity.
- 25 17. A method for preparing a library of template nucleic acids, comprising:
- (a) contacting a target nucleic acid with a plurality of transposons of claim 1, further comprising a first primer site, under conditions such that a portion of said transposons are inserted into the target nucleic acid; and
- 30 (b) hybridizing a primer to the first primer site; and
- (c) amplifying a portion of the transposon sequence.
18. A method of analyzing a target nucleic acid comprising:

- (a) performing the method of claim 17;
- (b) obtaining sequence data from said templates; and
- (c) assembling a representation of a portion of said target nucleic acid from said sequence data.

5

19. The method of claim 18, wherein assembling step (c) comprises identifying the presence of a first barcode in one sequencing read and the presence of a corresponding barcode in another sequencing read, thereby indicating proximity between the two sequencing reads in the target nucleic acid.

10

20. The method of claim 13, wherein step (b) is terminated after the barcode is identified.

1/36

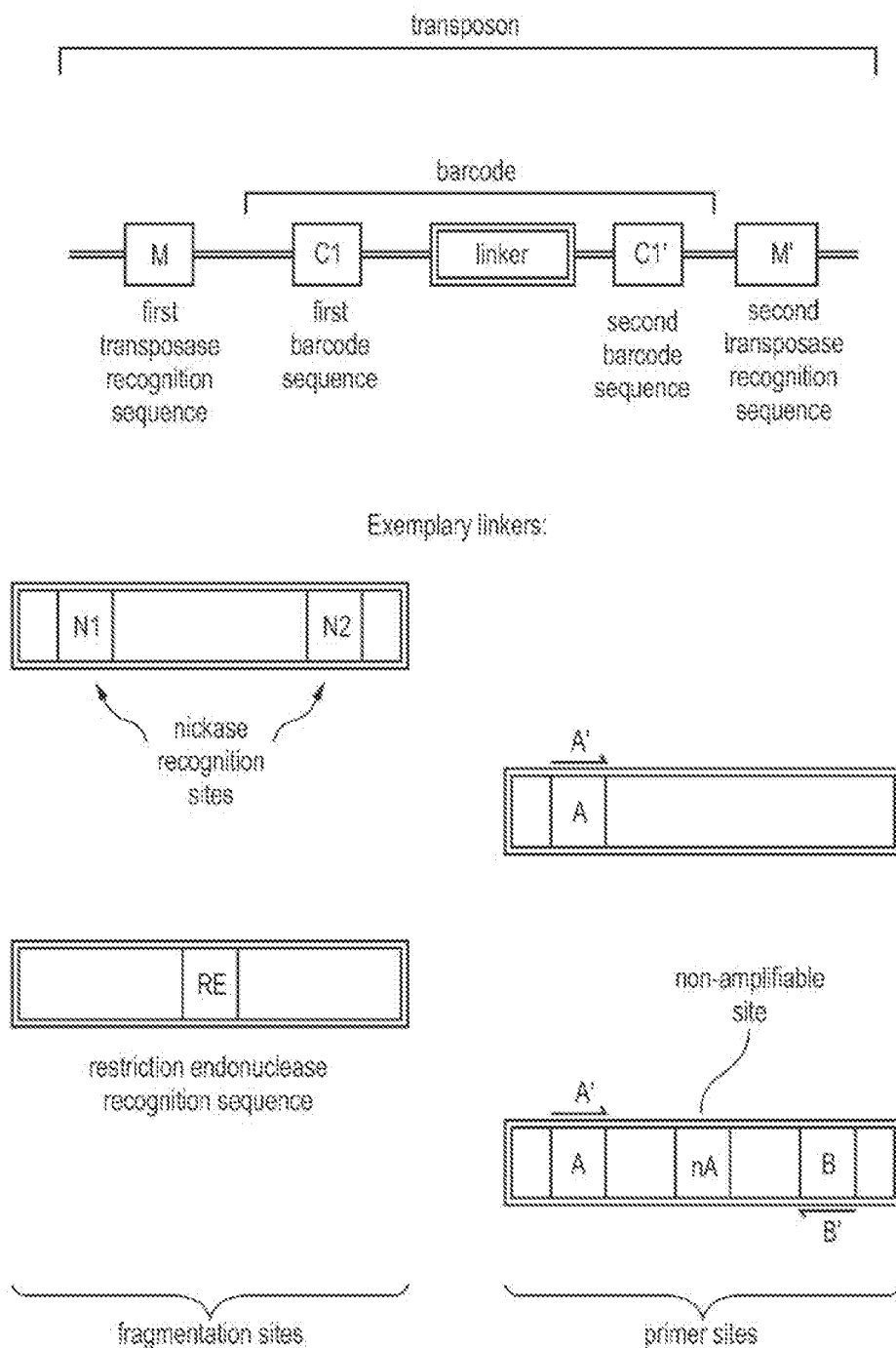


Fig. 1

SUBSTITUTE SHEET (RULE 26)

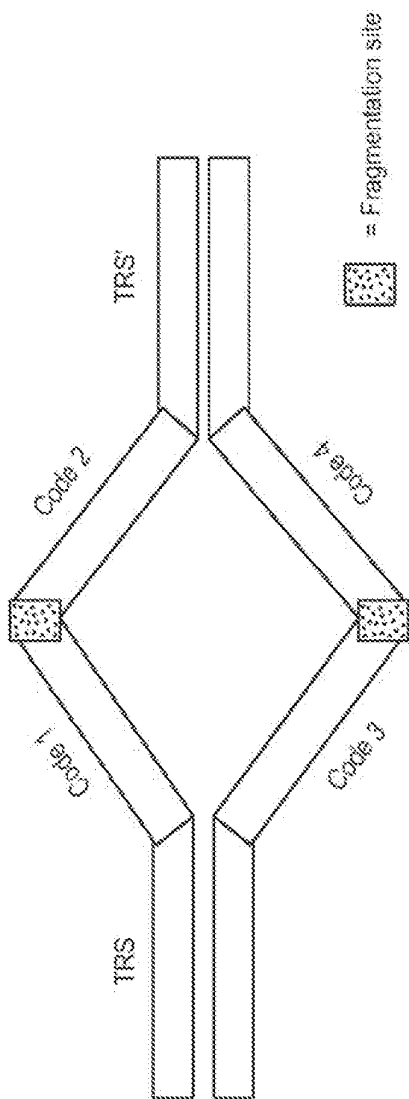


Fig. 2

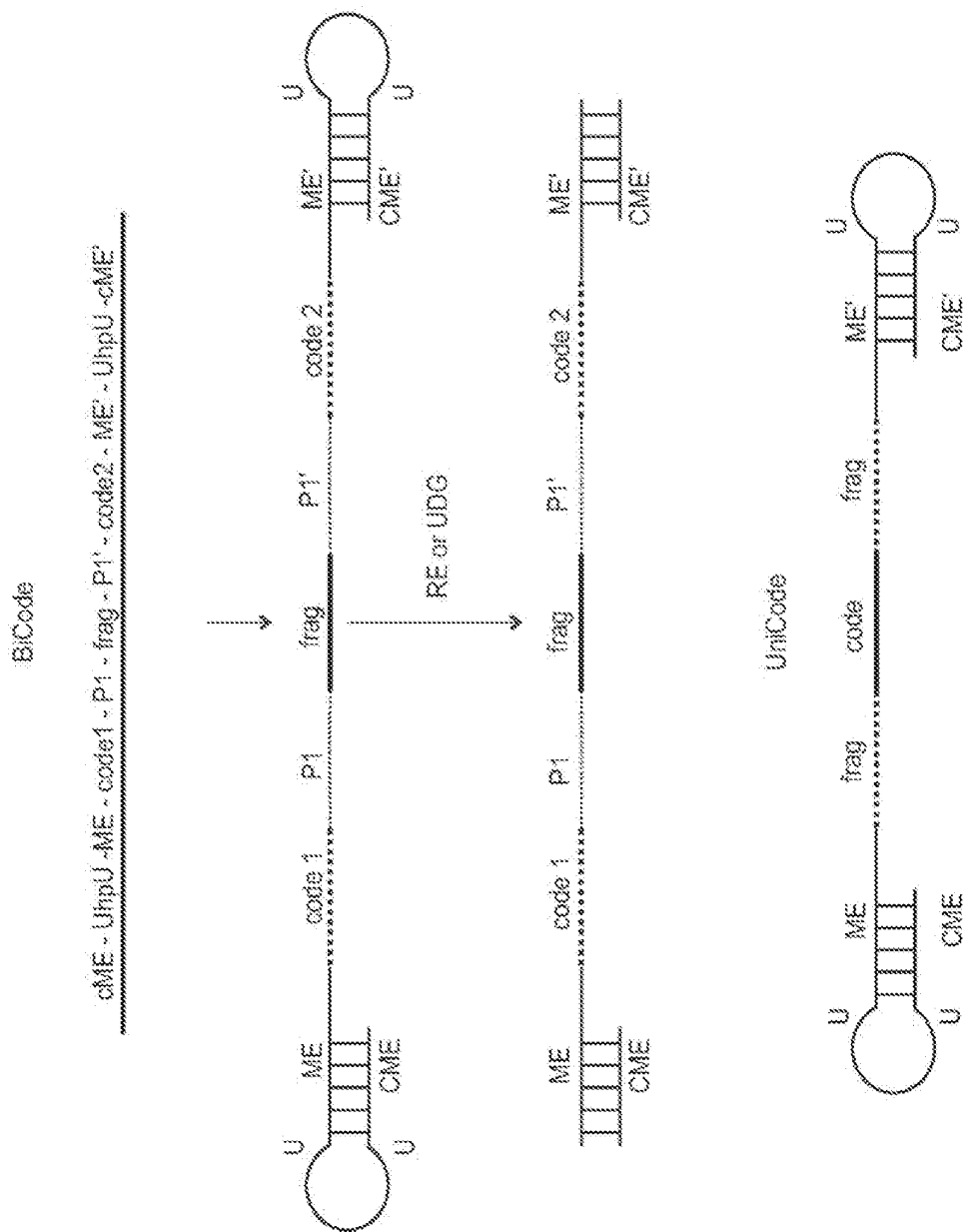


Fig. 3

4/36

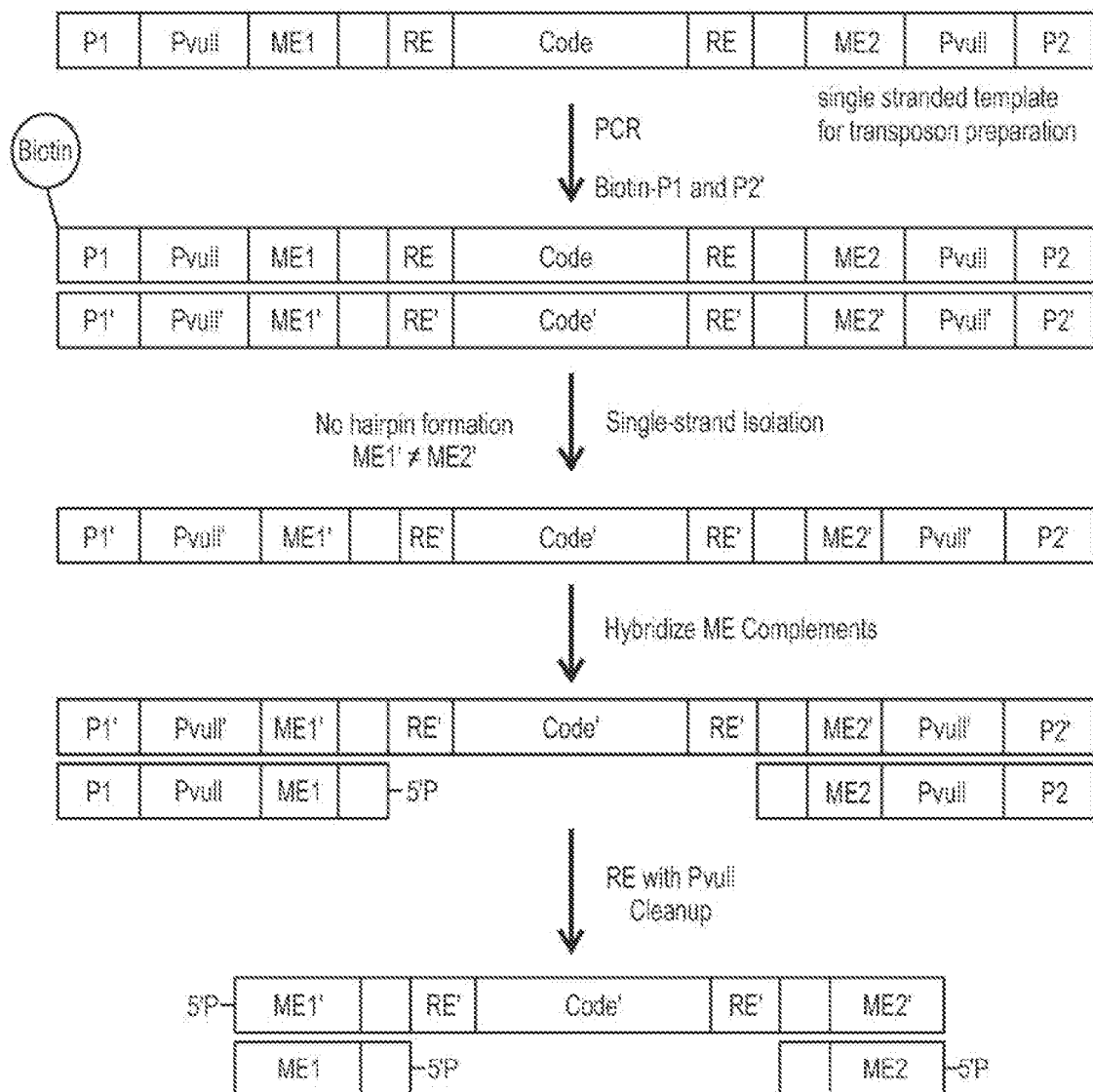


Fig. 4

SUBSTITUTE SHEET (RULE 26)

5/36

Transposon-Transposase Complex Formation

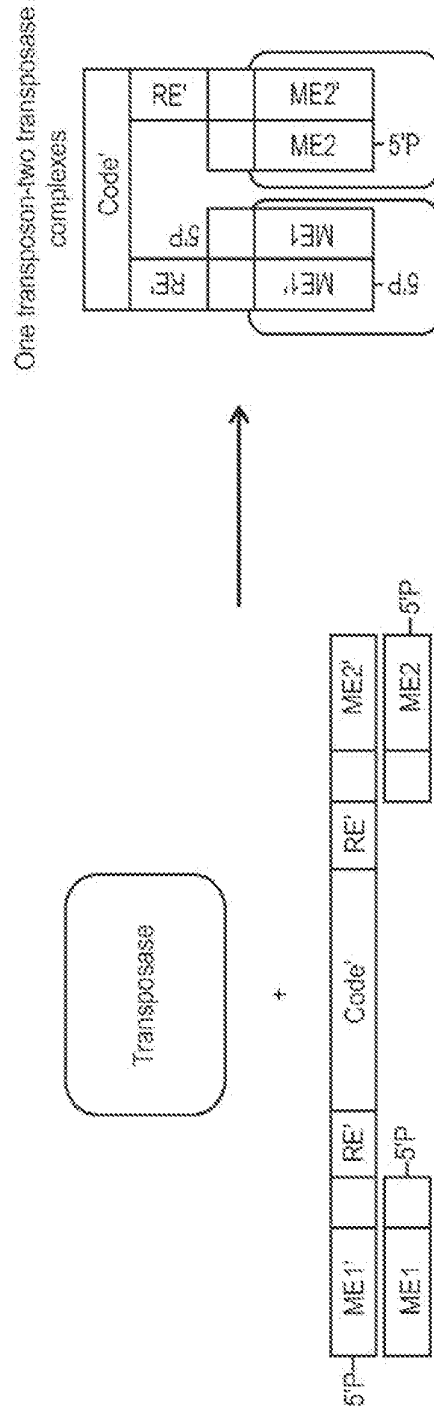


Fig. 5

SUBSTITUTE SHEET (RULE 26)

Transposition of "single-stranded" Transposon

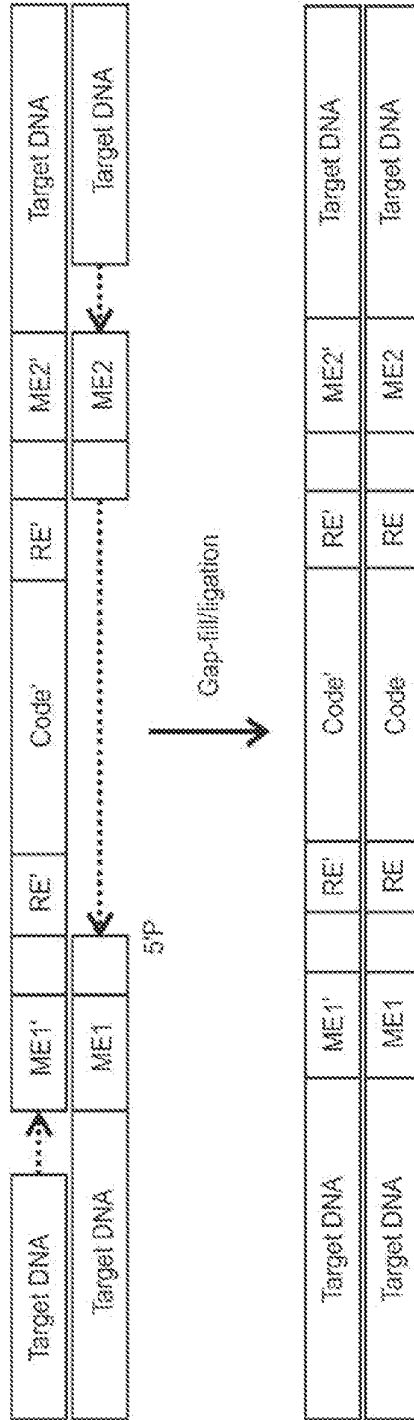


Fig. 6

7/36

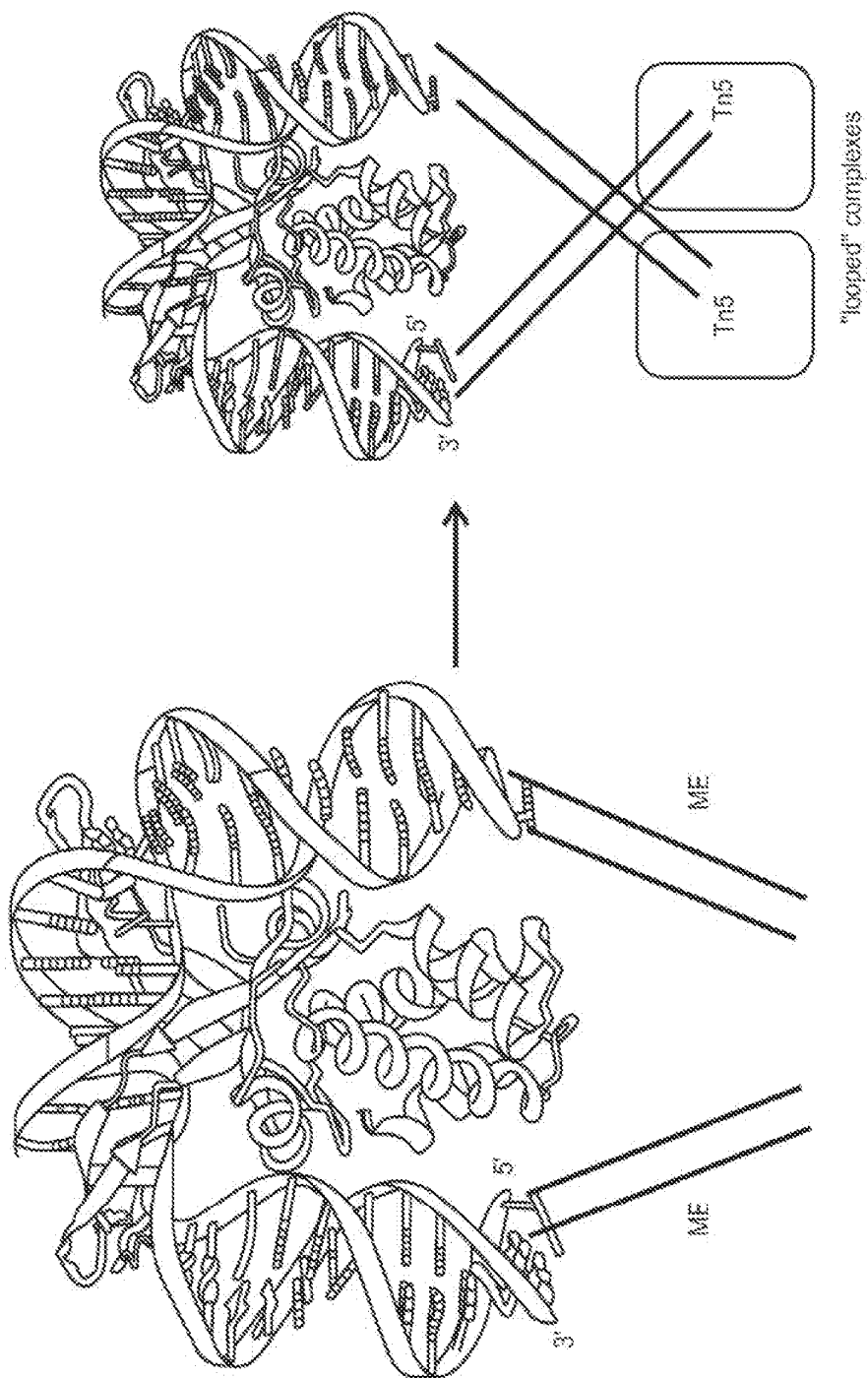


Fig. 7

SUBSTITUTE SHEET (RULE 26)

8/36

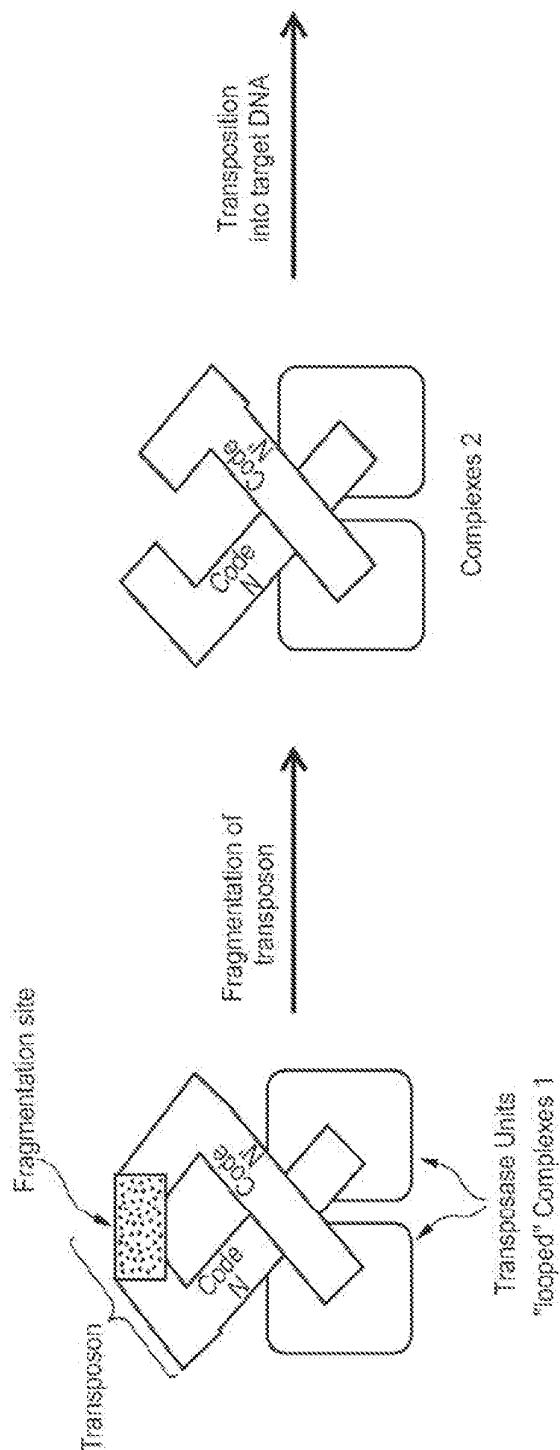


Fig. 8

SUBSTITUTE SHEET (RULE 26)

9/36

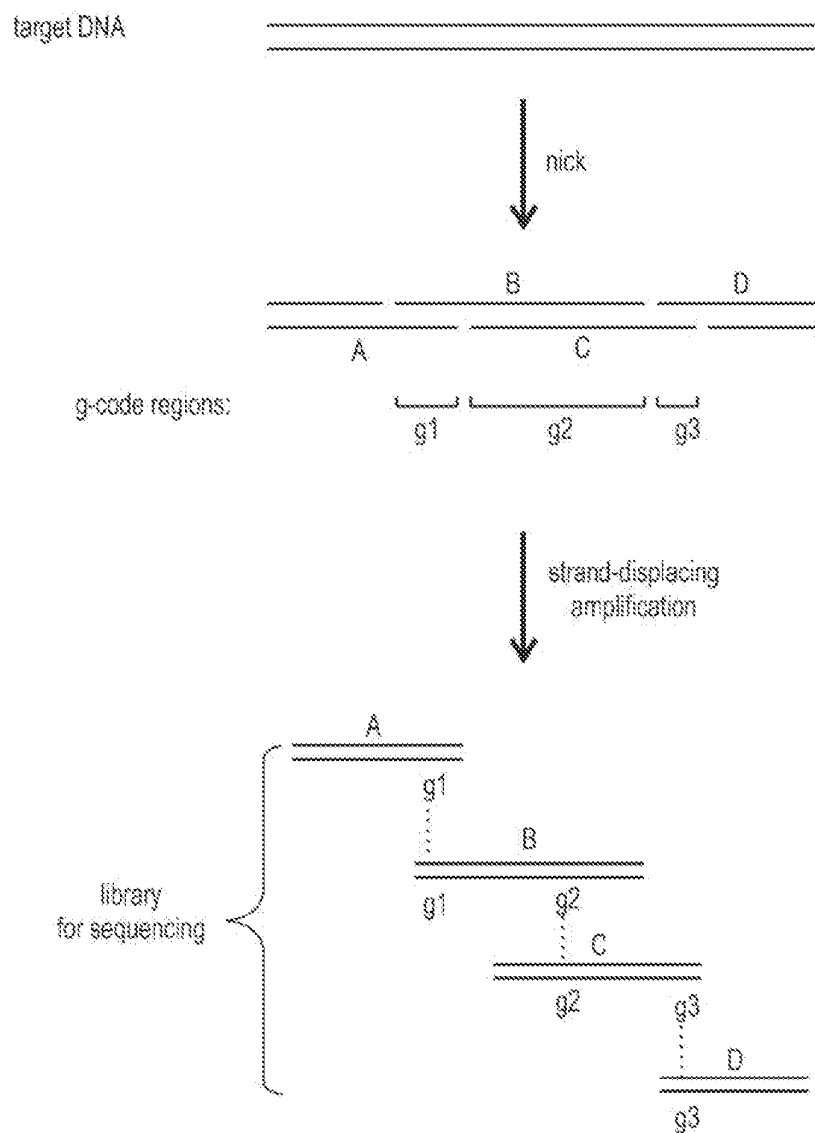


Fig. 9

SUBSTITUTE SHEET (RULE 26)

9/36

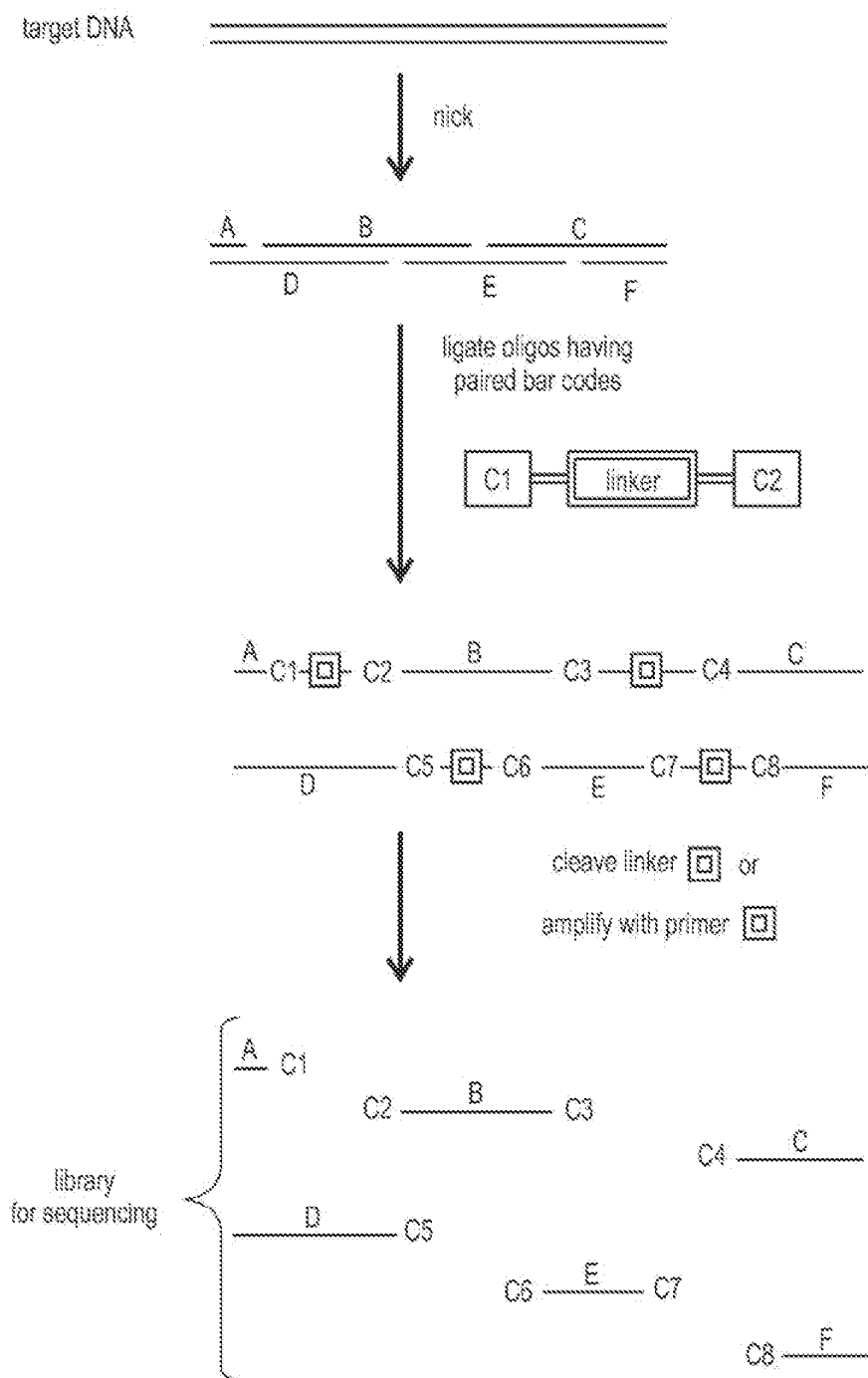


Fig. 10

SUBSTITUTE SHEET (RULE 26)

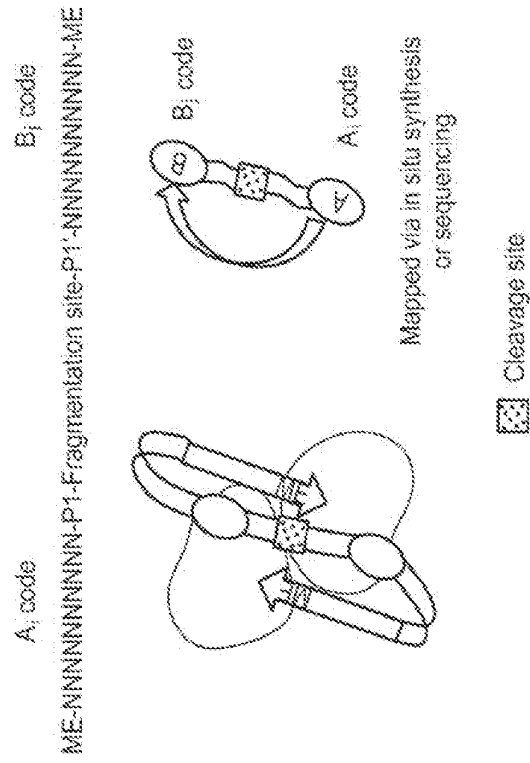
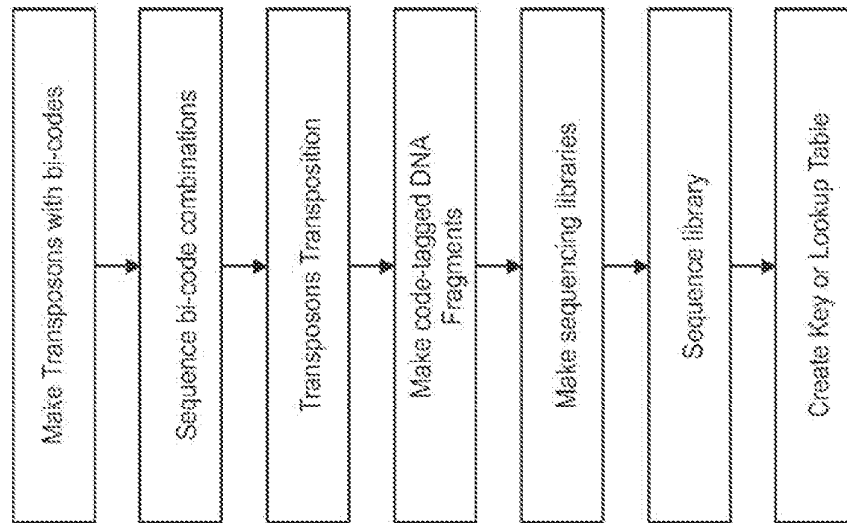


Fig. 11

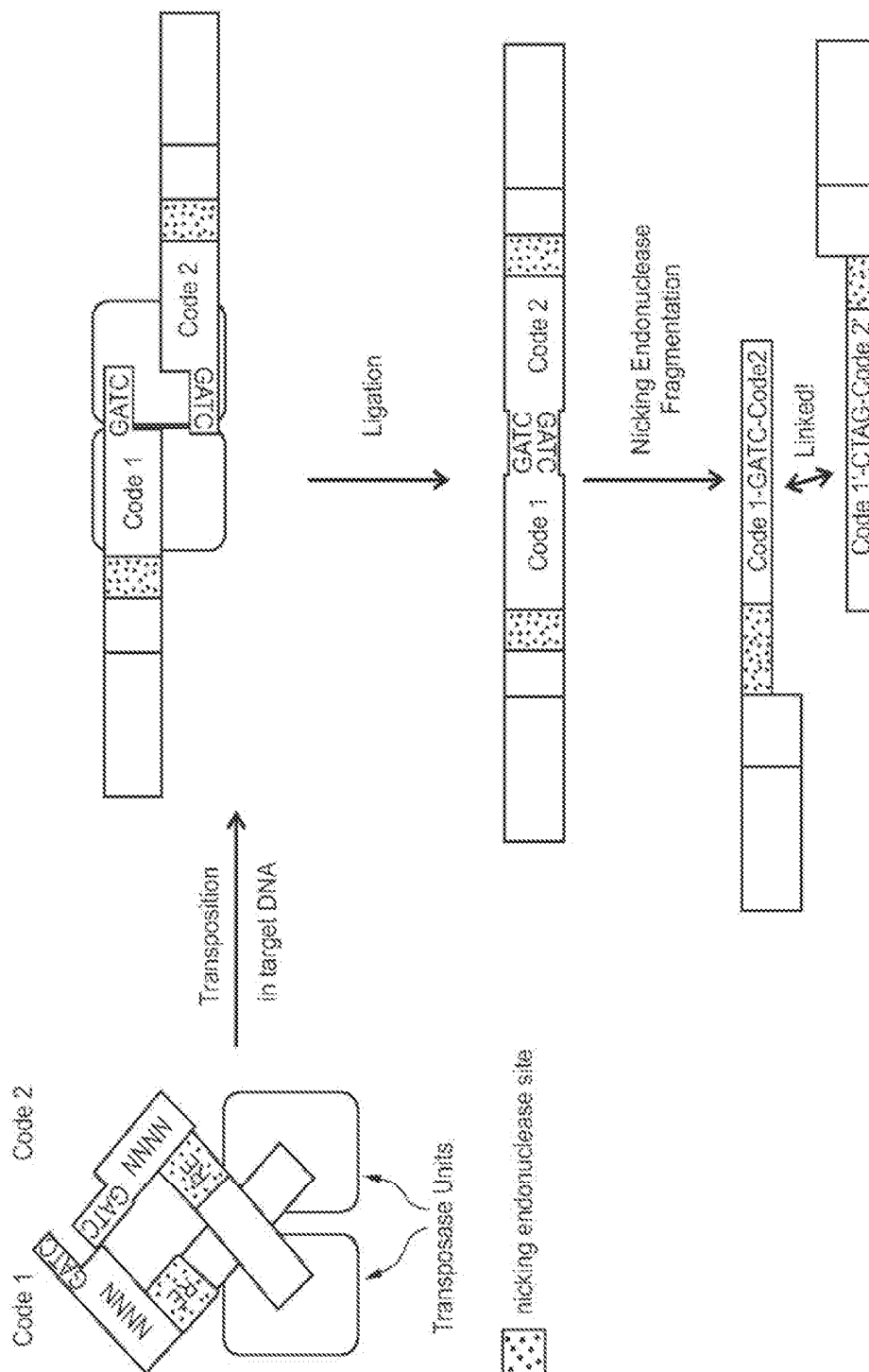


Fig. 12

13/36

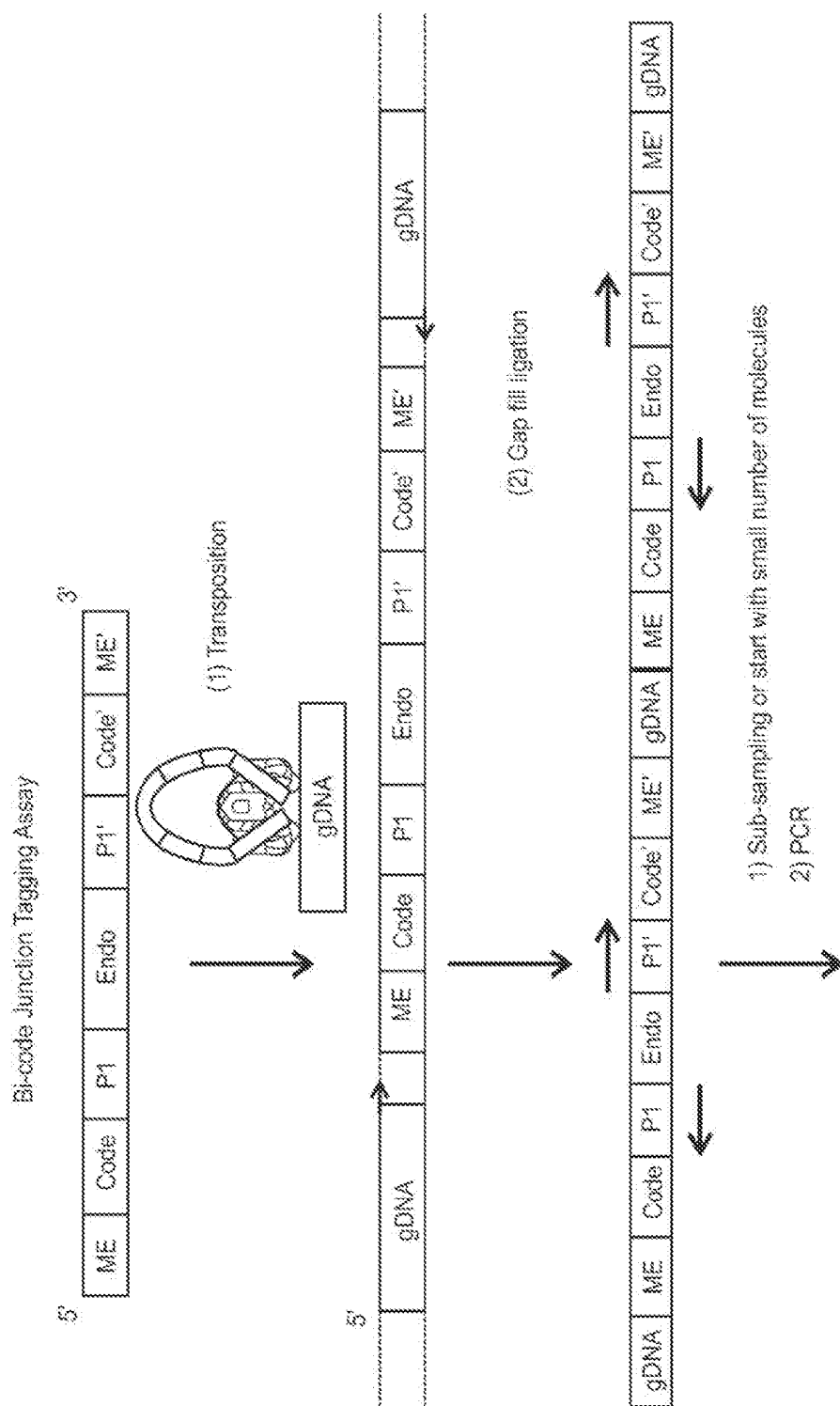


Fig. 13

SUBSTITUTE SHEET (RULE 26)

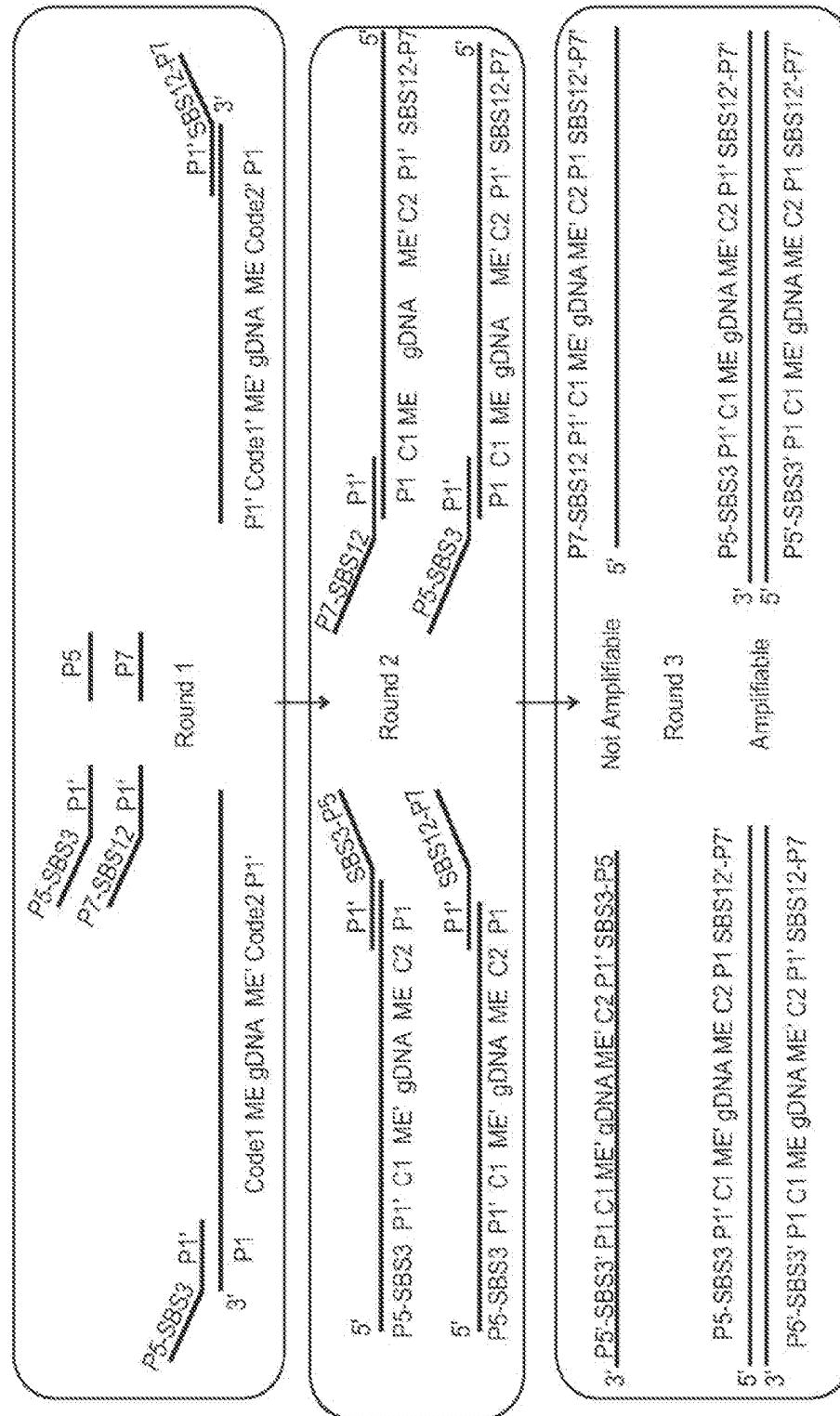


Fig. 14

15/36

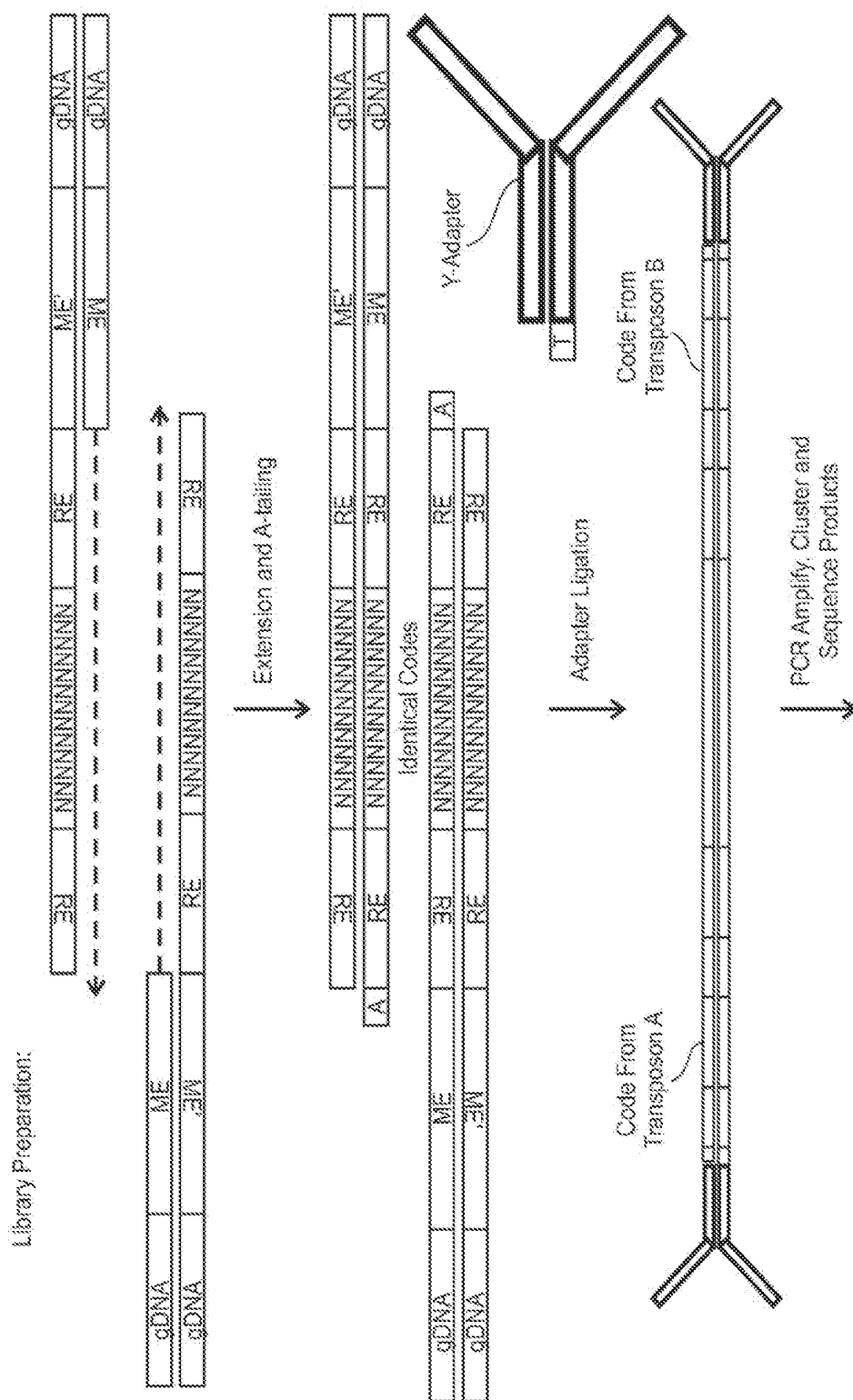
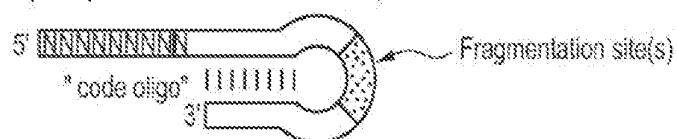


Fig. 15

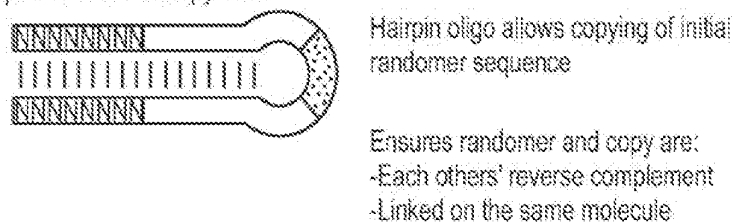
SUBSTITUTE SHEET (RULE 26)

16/36

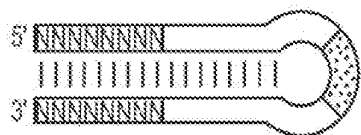
Step 1: synthesize randomer with hairpin



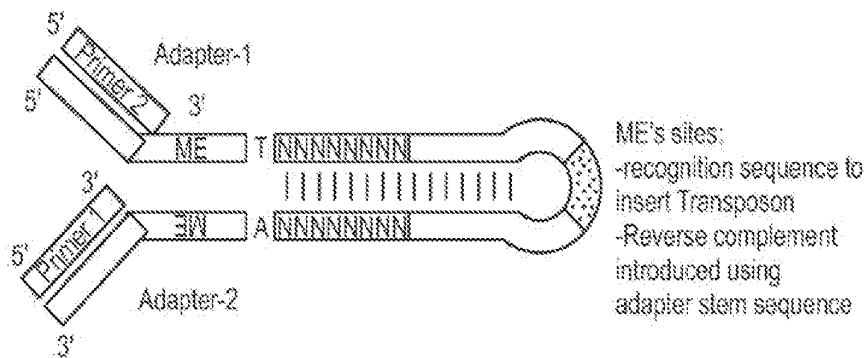
Step 2: extend and copy code



Step 3: A-tailing



Step 4: Adapter Ligation



Step 5: PCR amplification

P2	ME	Code				Code'	ME	P1'
P2'	ME'	Code'				Code	ME'	P1

Fig. 16

SUBSTITUTE SHEET (RULE 26)

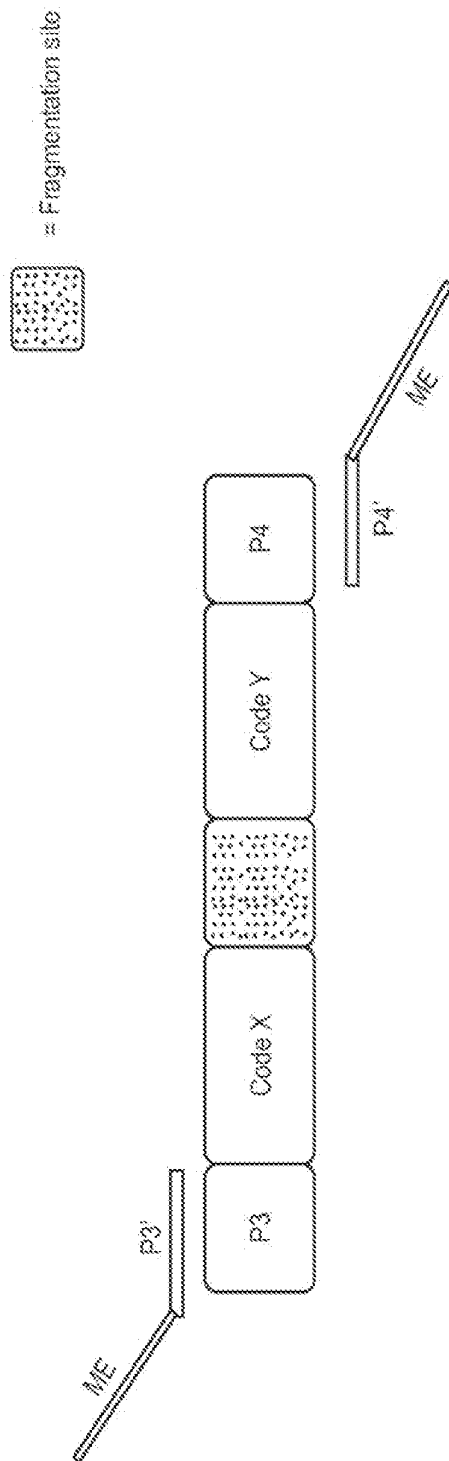


Fig. 17

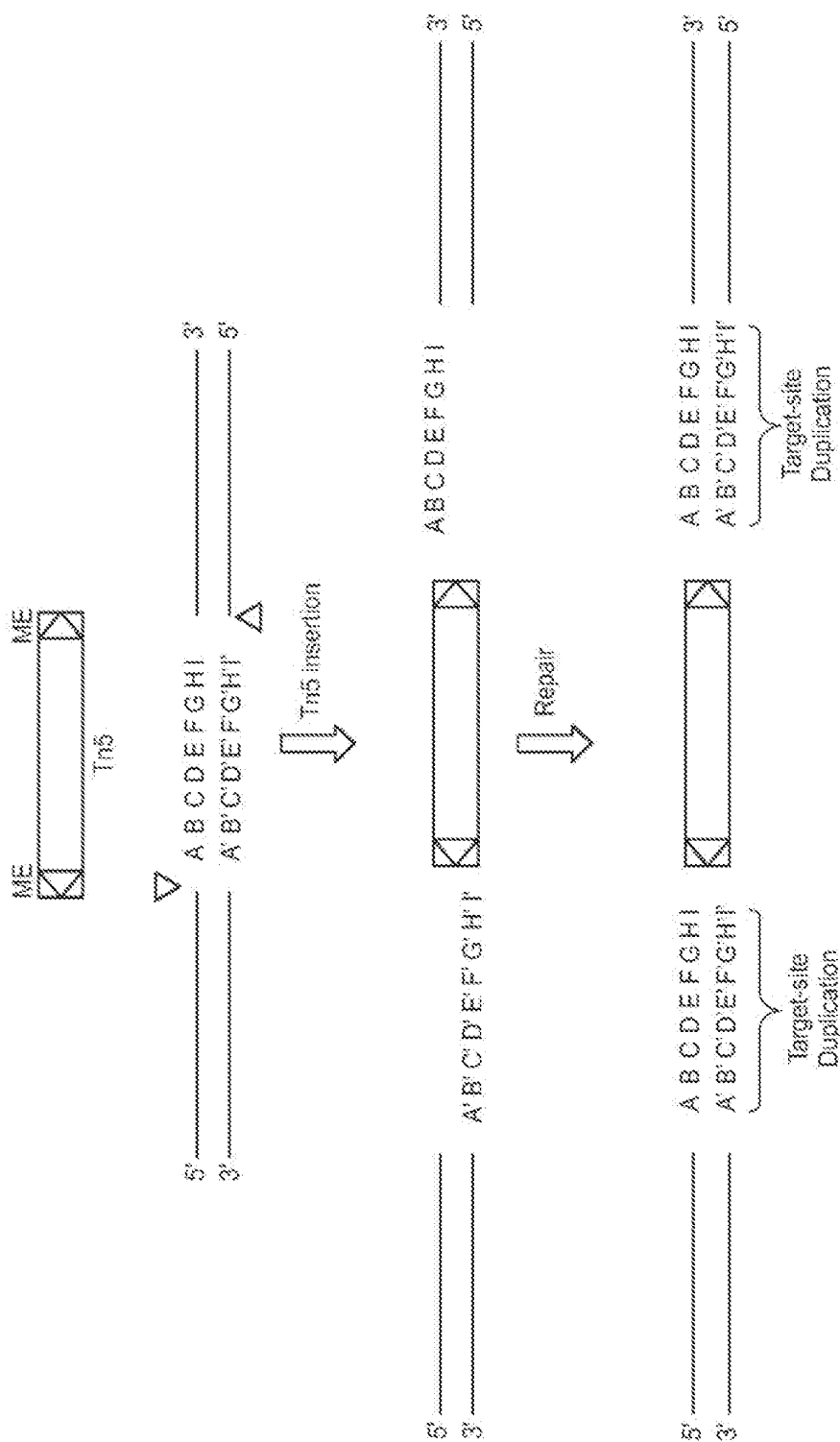


Fig. 18

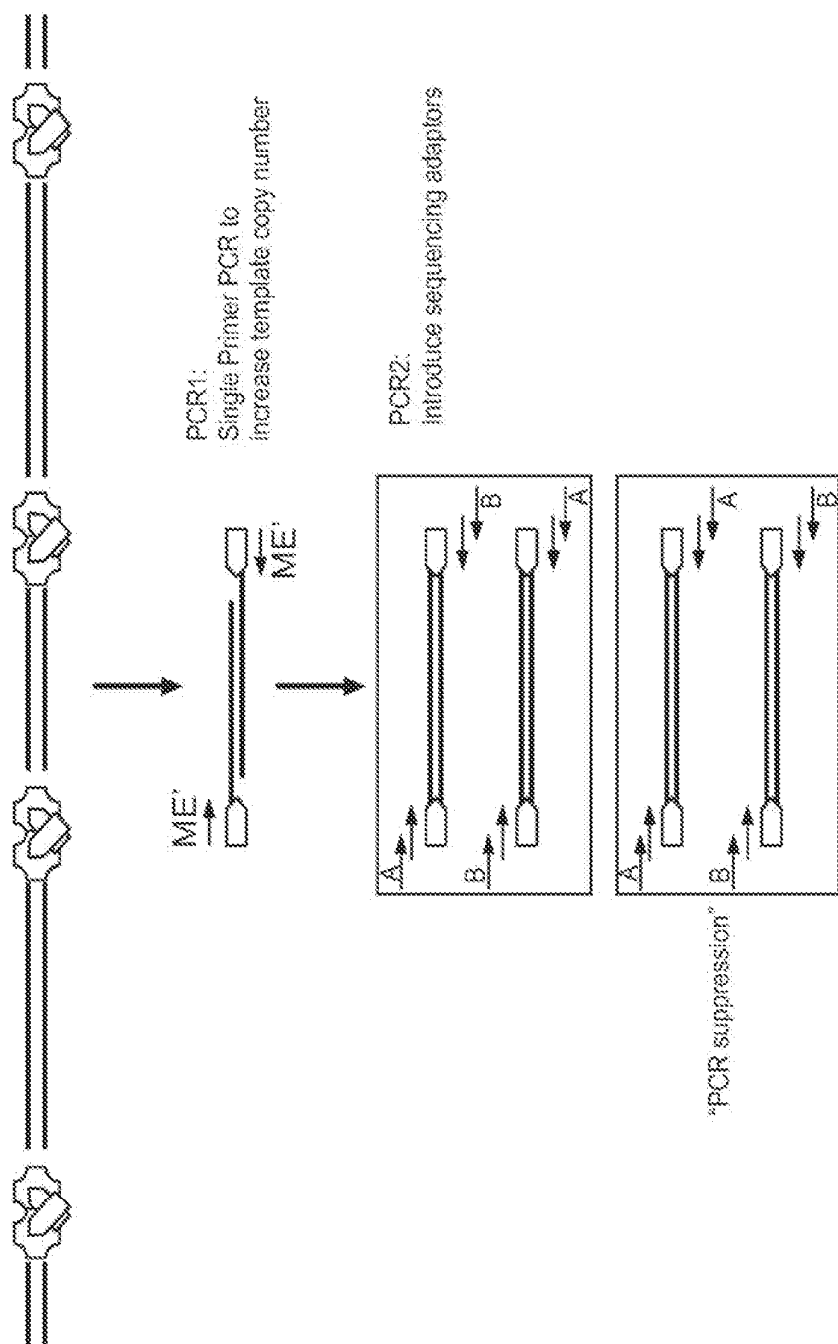


Fig. 19

20/36

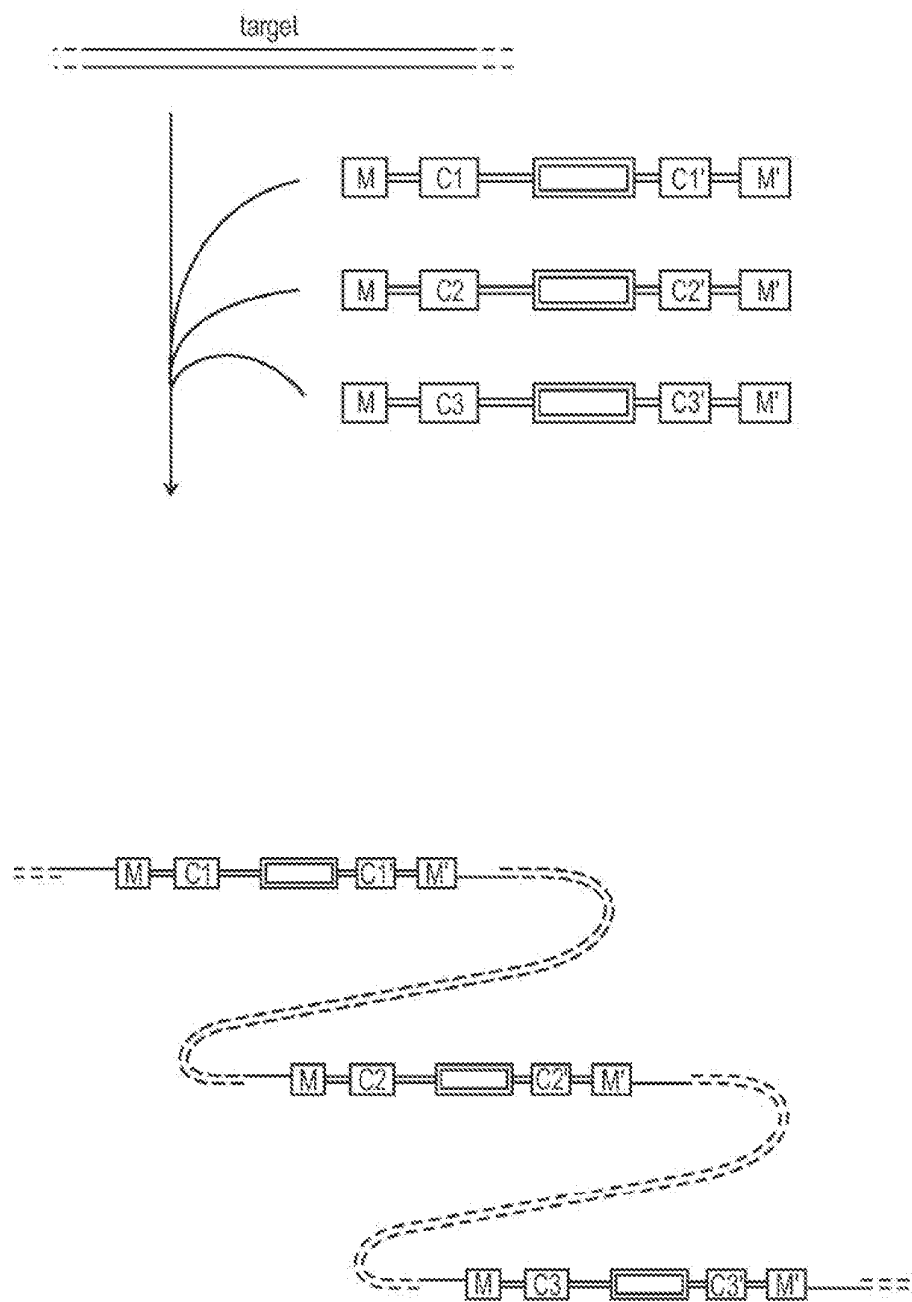


Fig. 20

SUBSTITUTE SHEET (RULE 26)

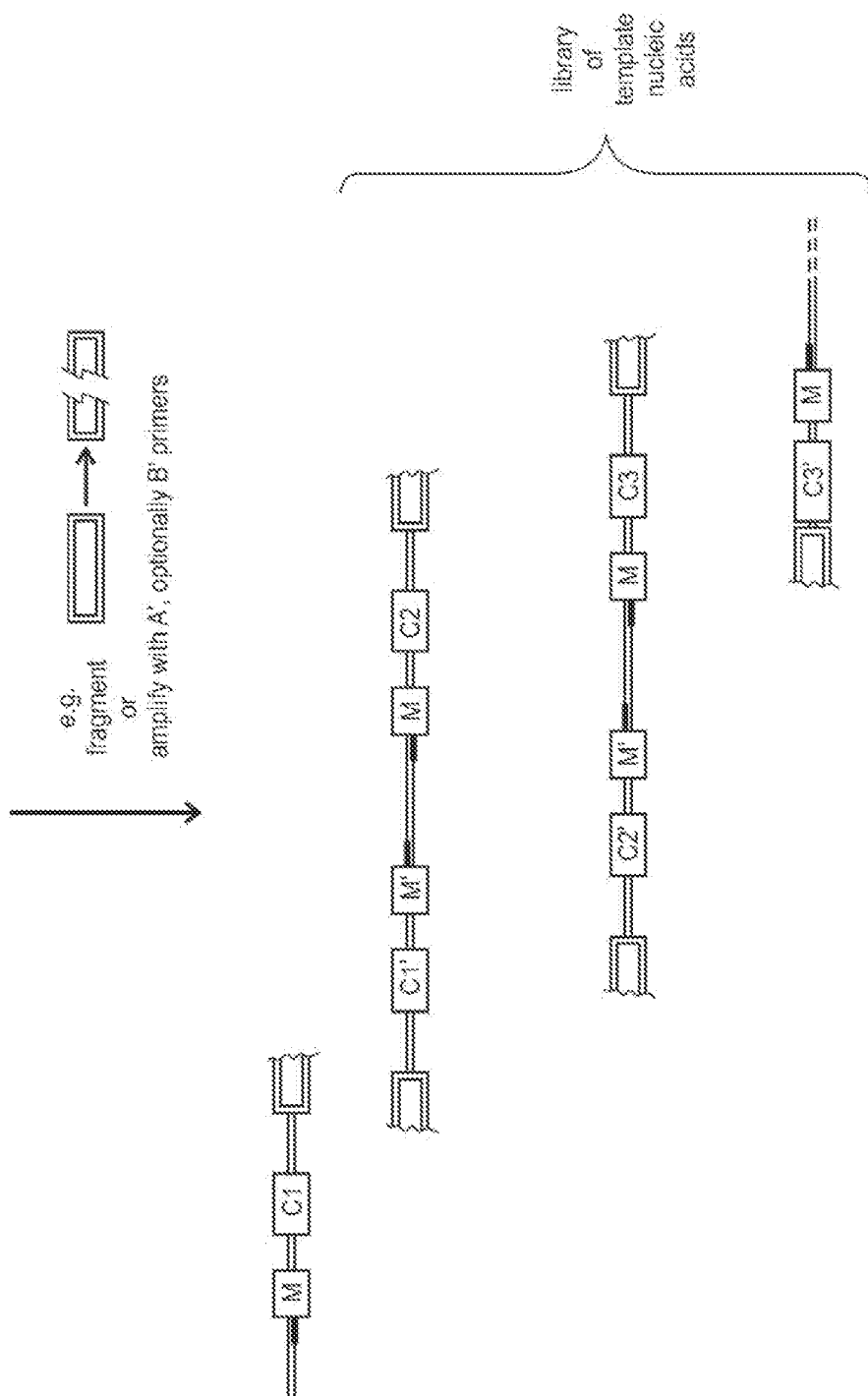


Fig. 21

22/36

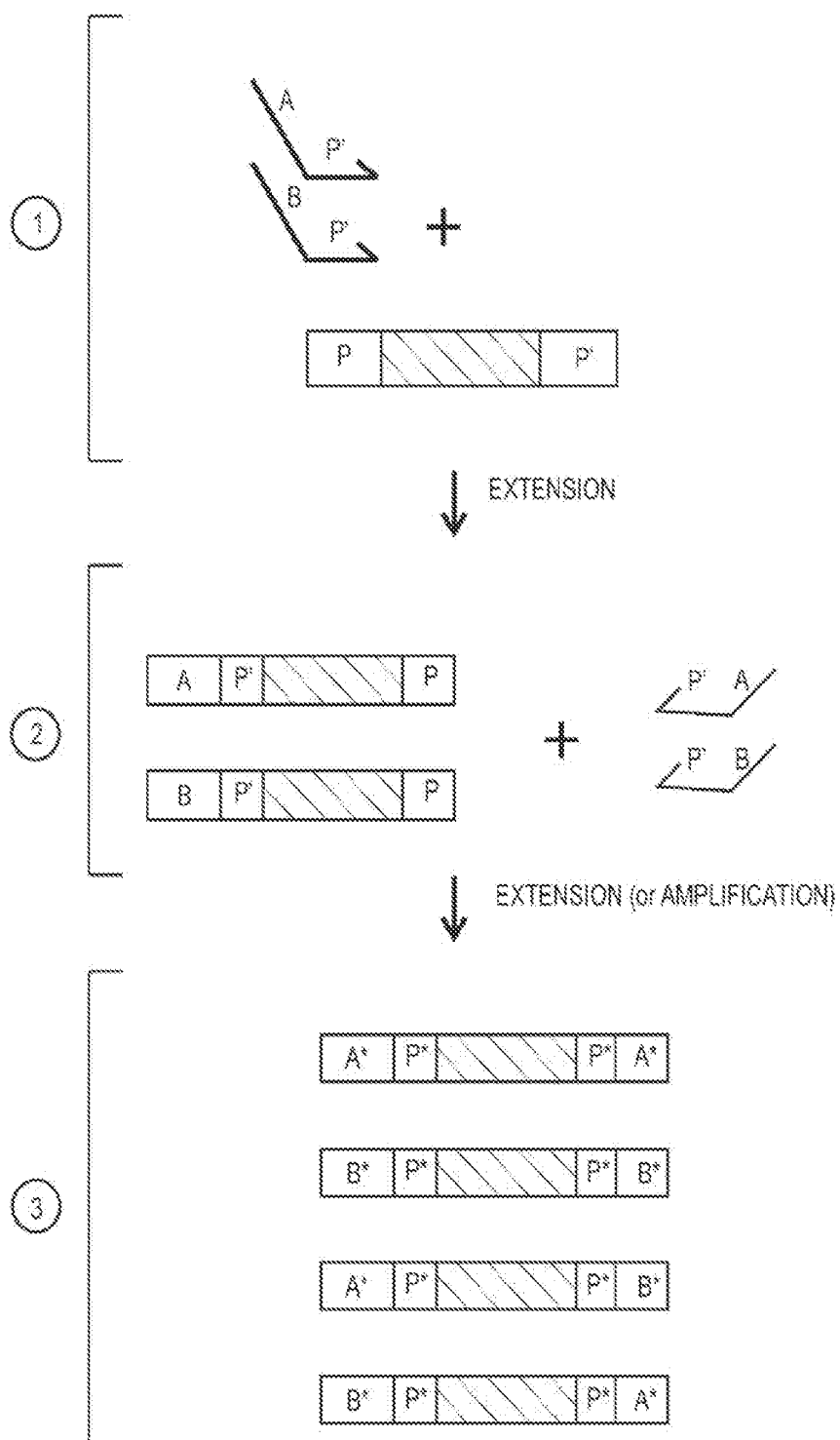


Fig. 22

SUBSTITUTE SHEET (RULE 26)

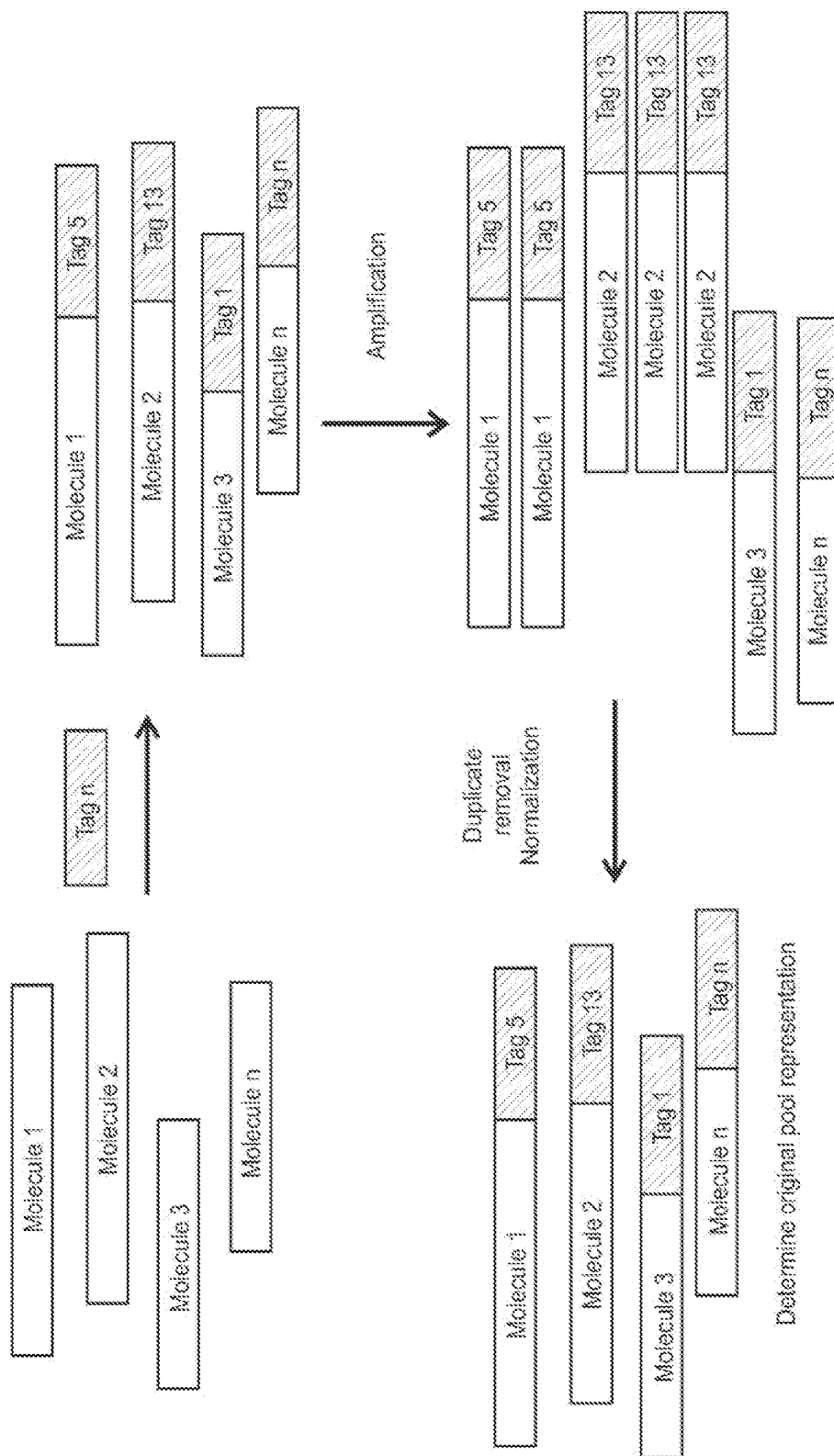


Fig. 23

24/36

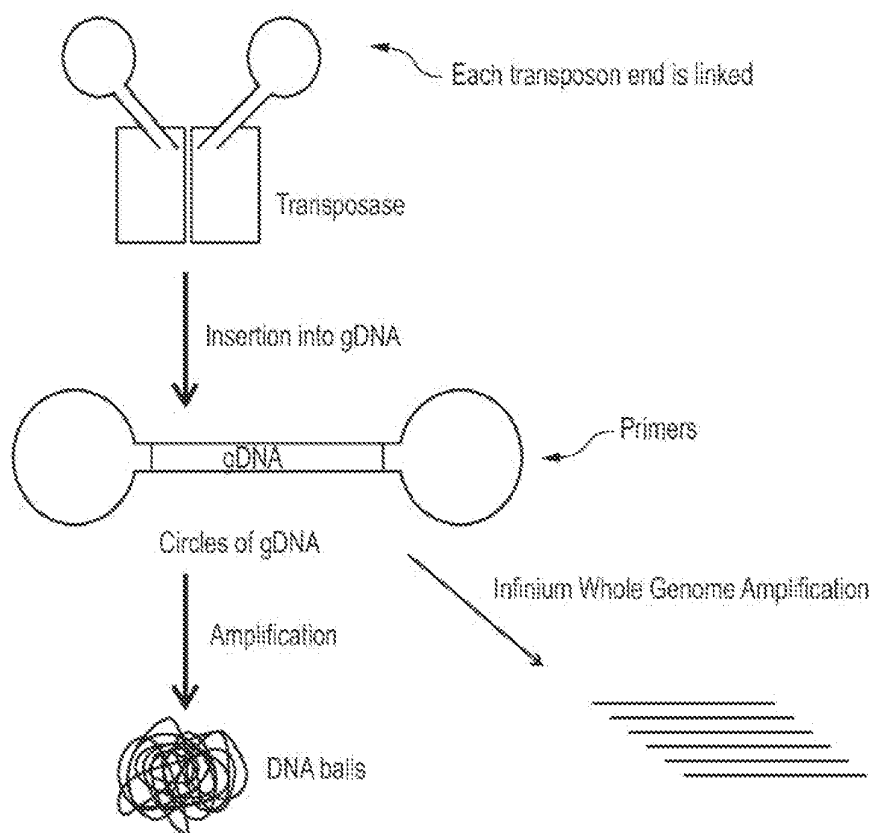


Fig. 24

SUBSTITUTE SHEET (RULE 26)

25/36

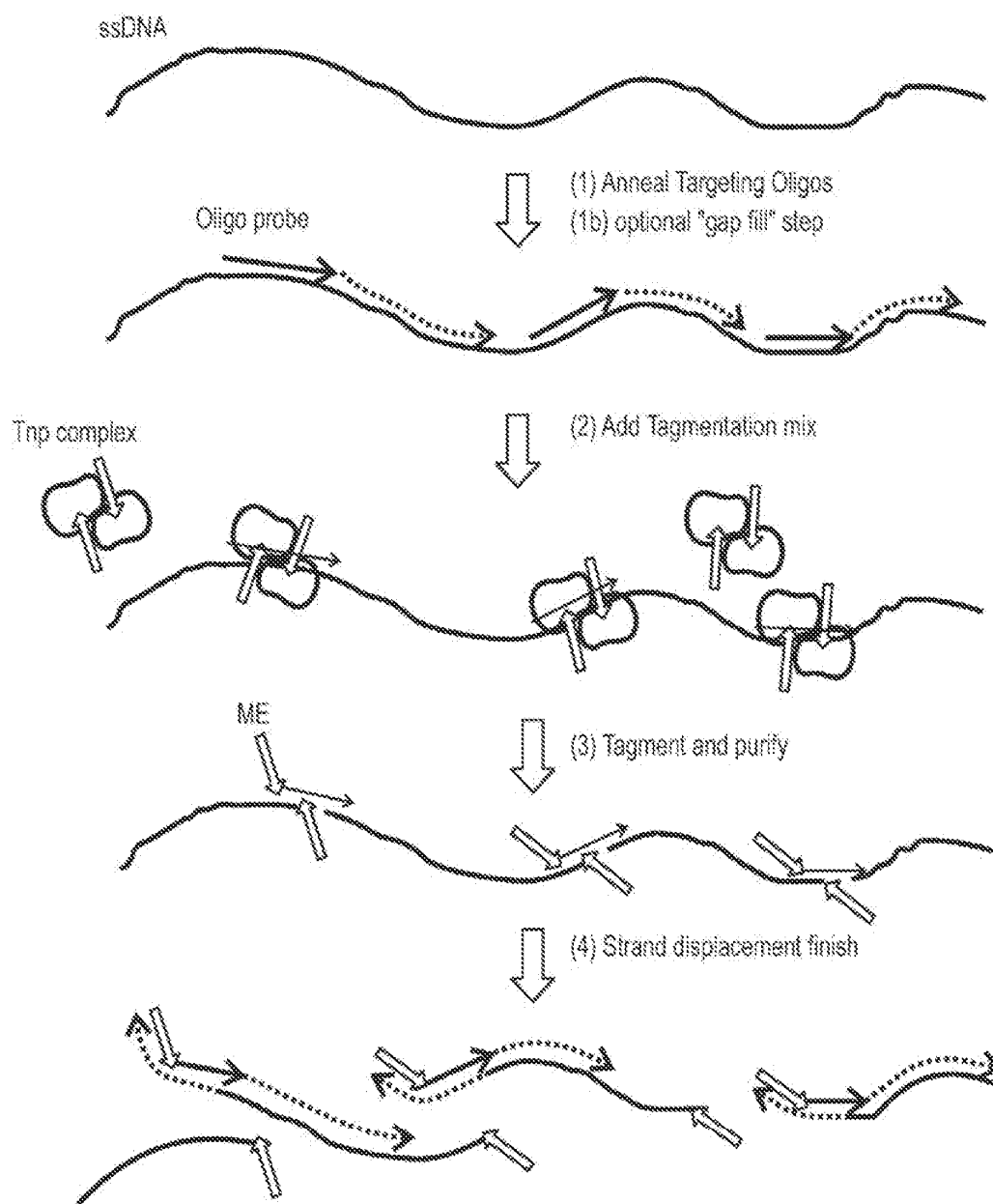


Fig. 25

SUBSTITUTE SHEET (RULE 26)

26/36

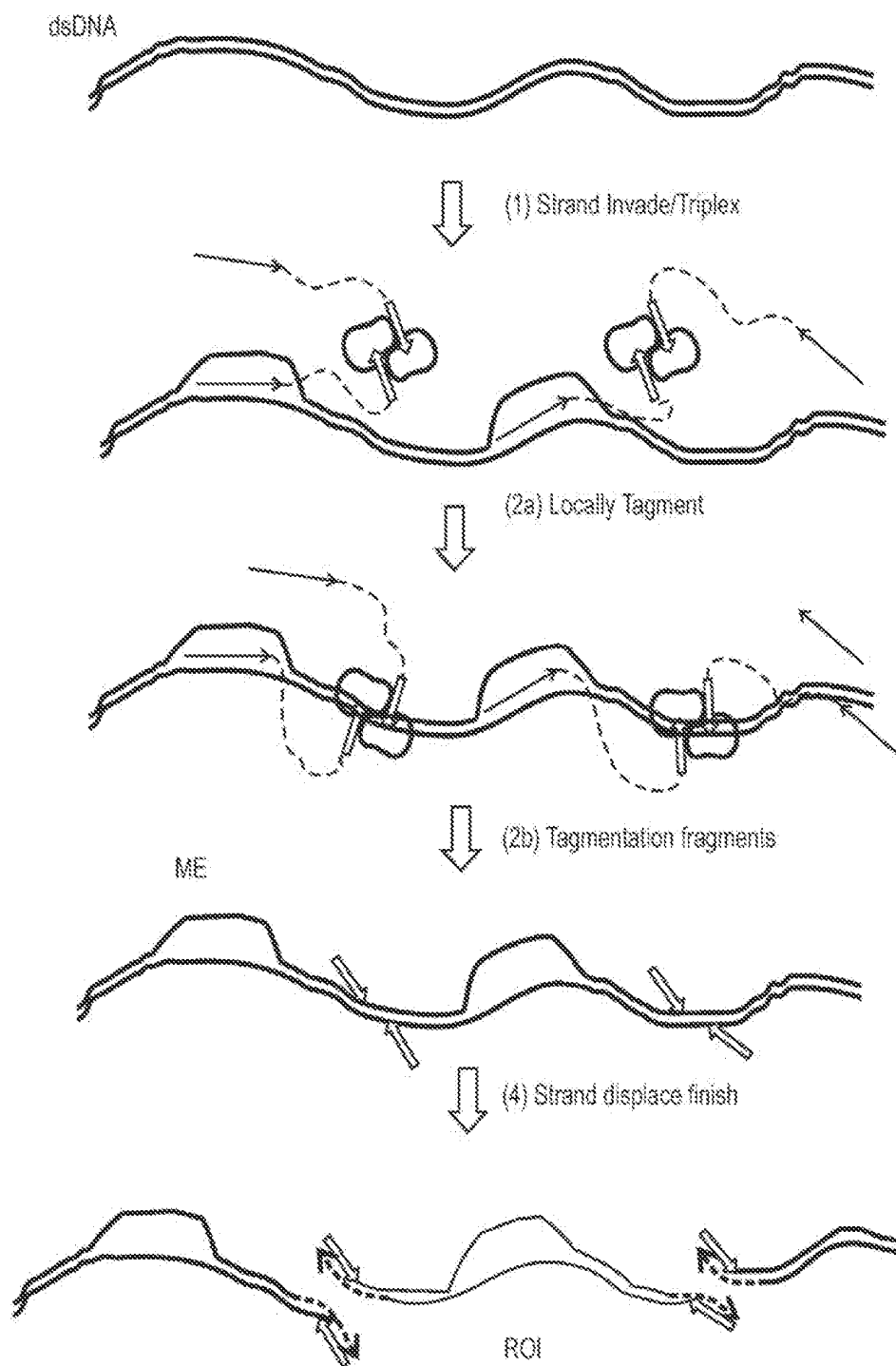


Fig. 26

SUBSTITUTE SHEET (RULE 26)

27/36

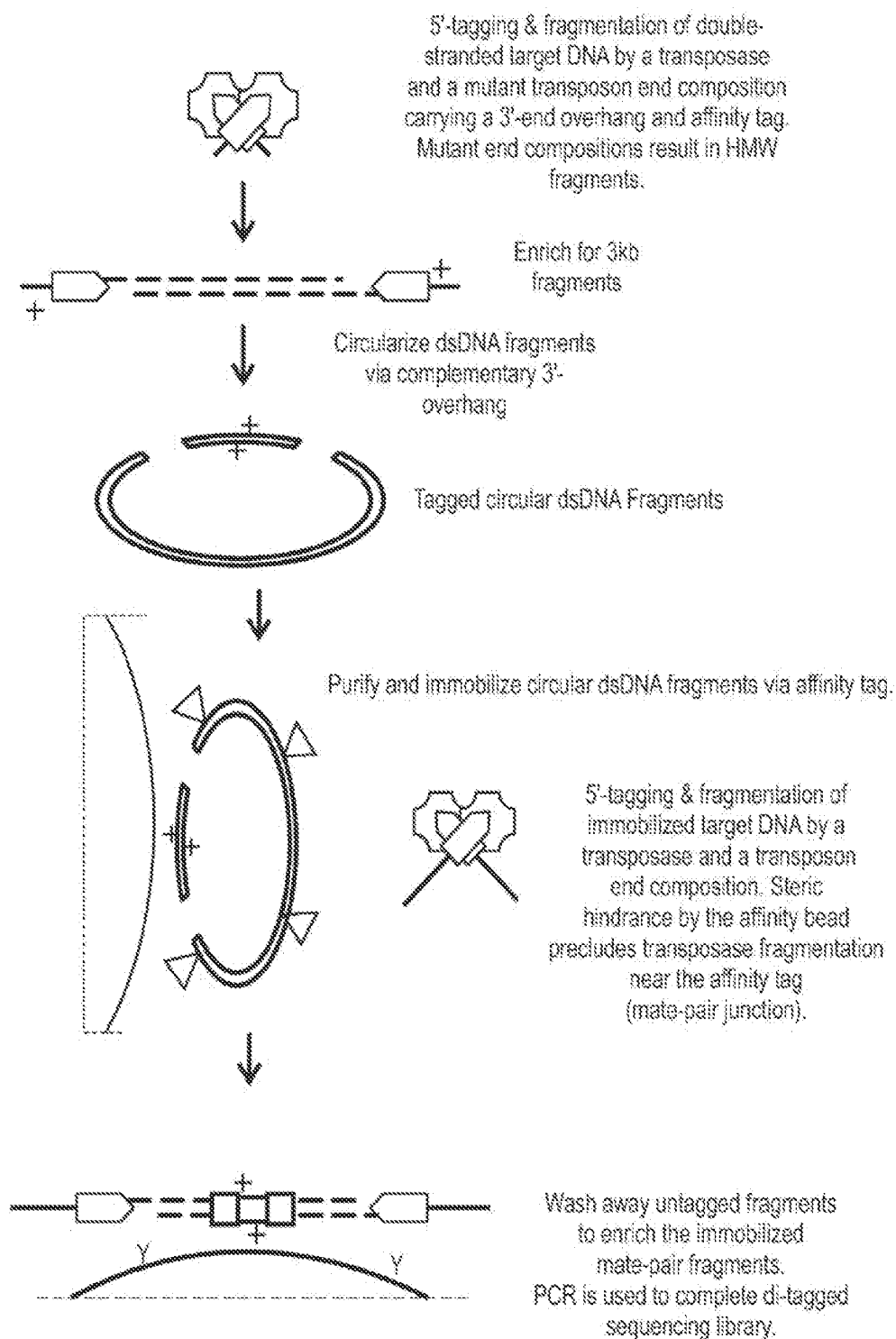


Fig. 27

SUBSTITUTE SHEET (RULE 26)

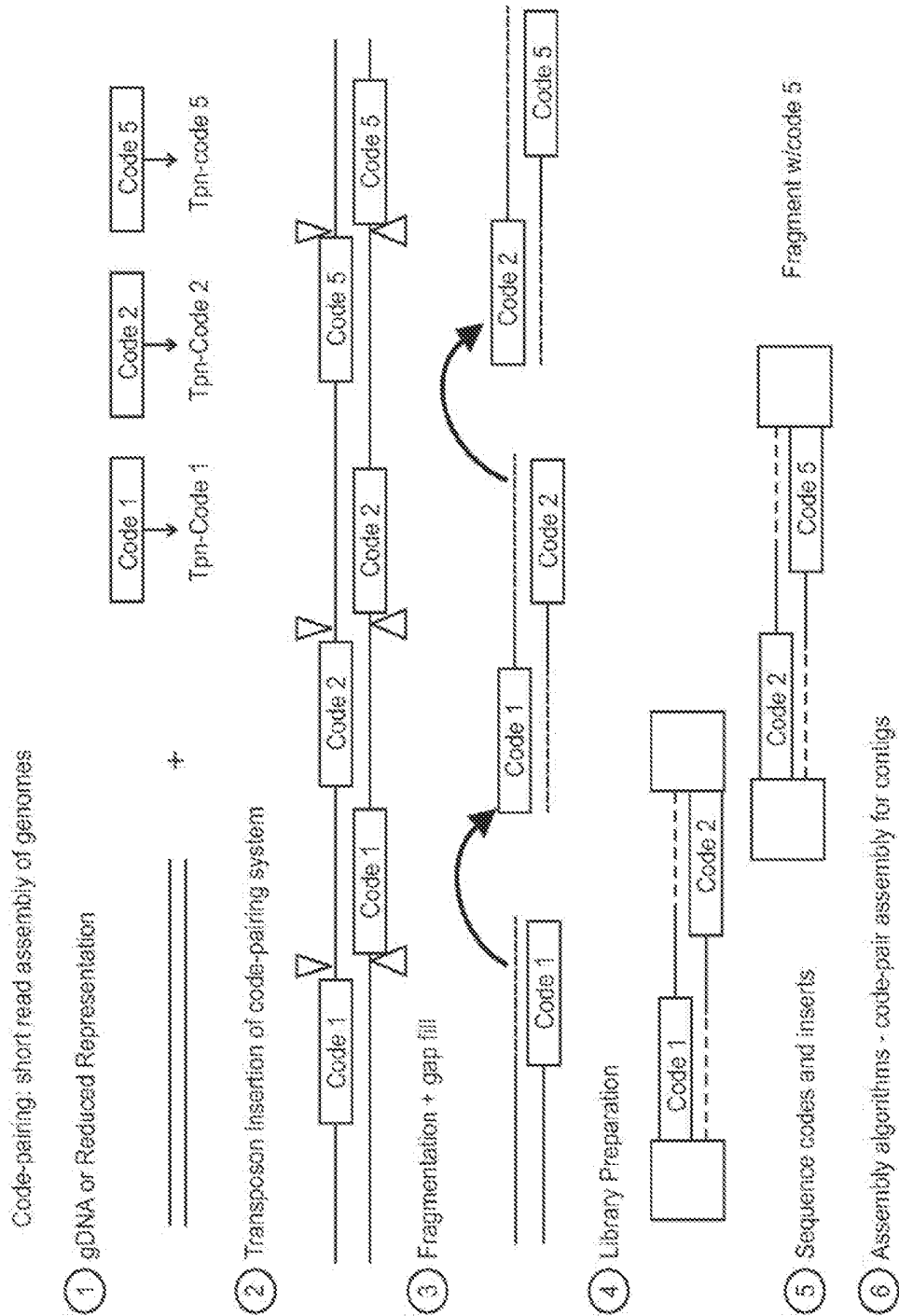


Fig. 28

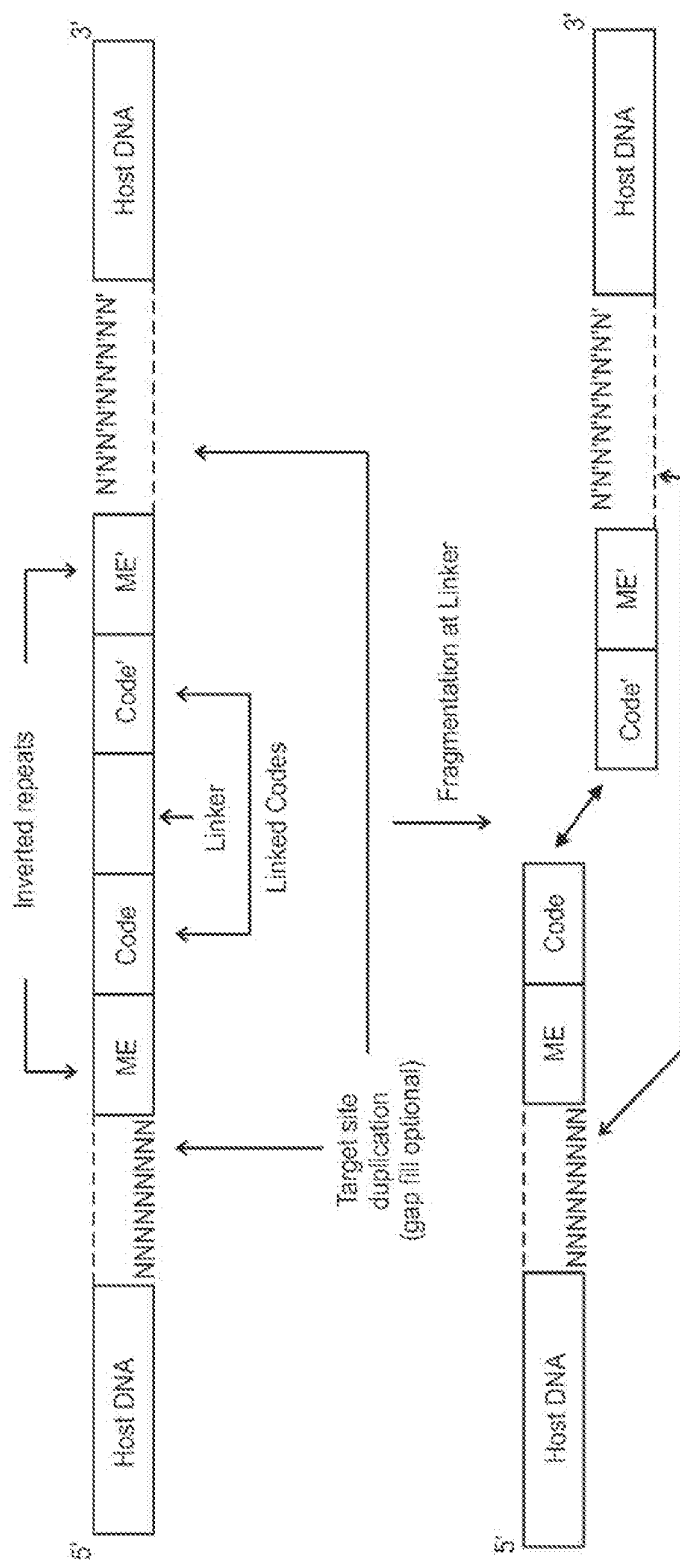


Fig. 29

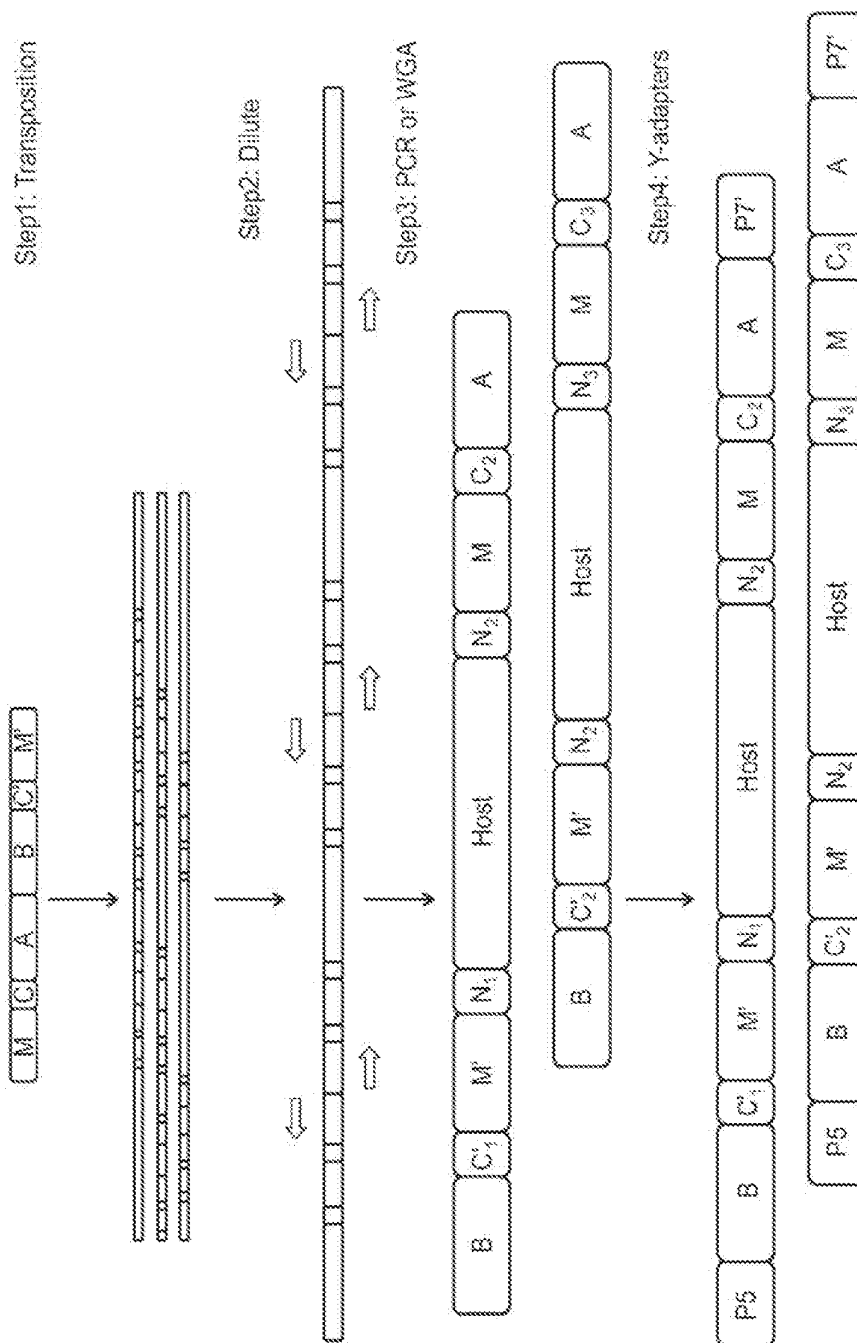


Fig. 30

31/36

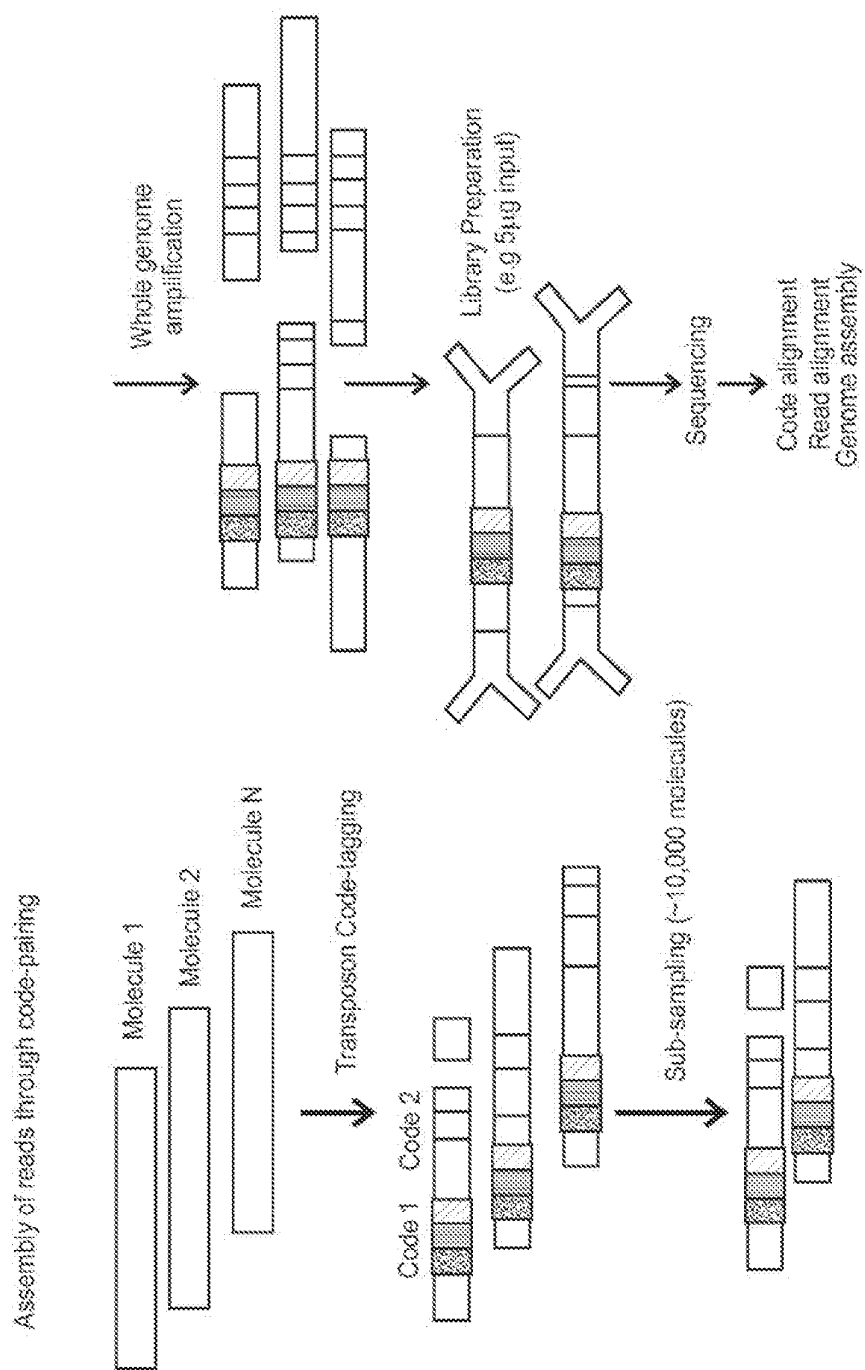


Fig. 31

SUBSTITUTE SHEET (RULE 26)

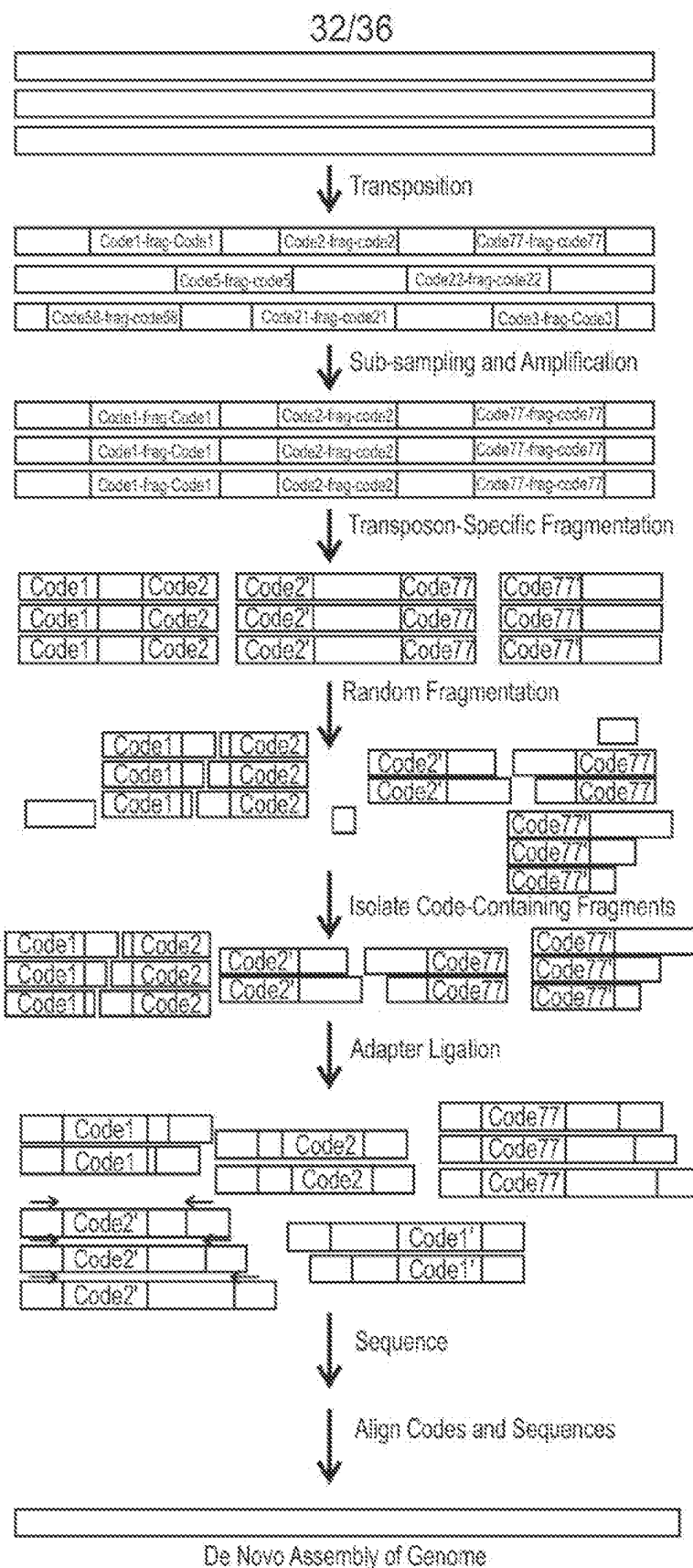


Fig. 32

SUBSTITUTE SHEET (RULE 26)

32/36

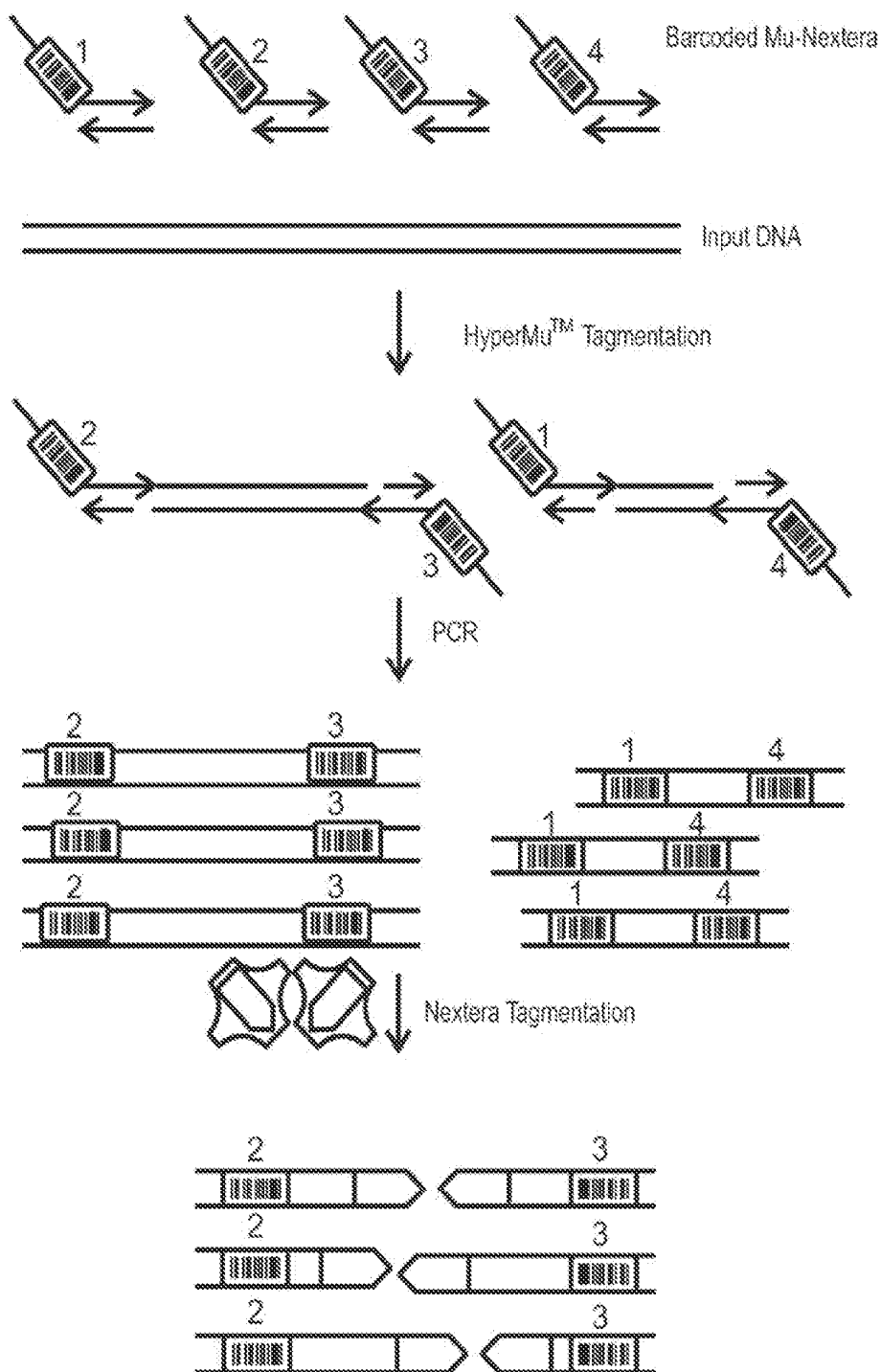


Fig. 33

SUBSTITUTE SHEET (RULE 26)

32/36

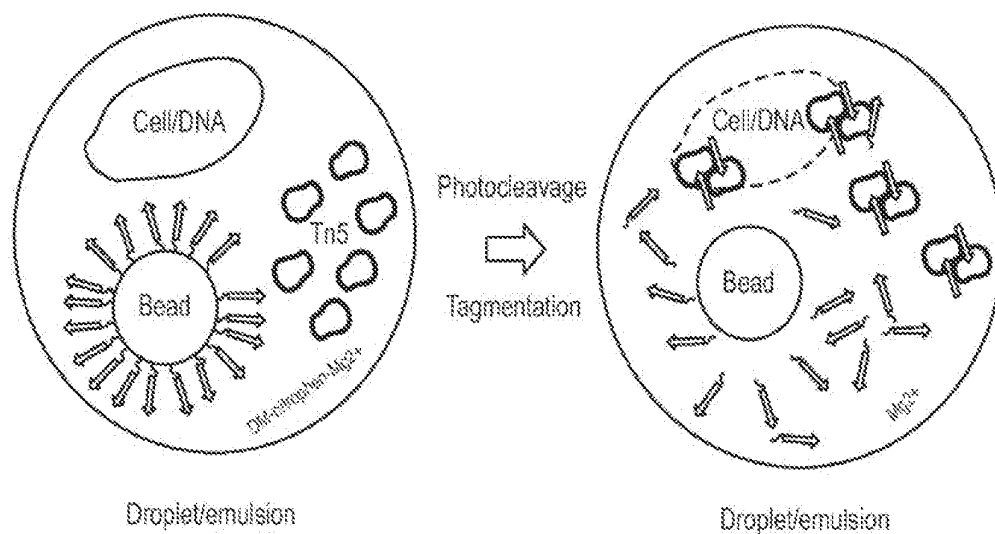


Fig. 34

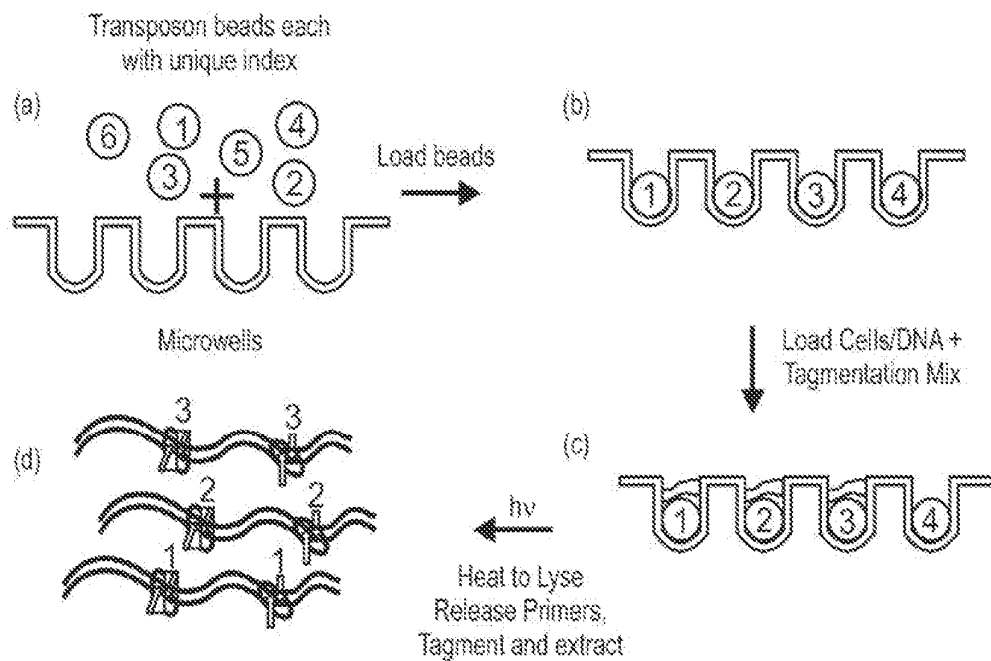


Fig. 35

SUBSTITUTE SHEET (RULE 26)

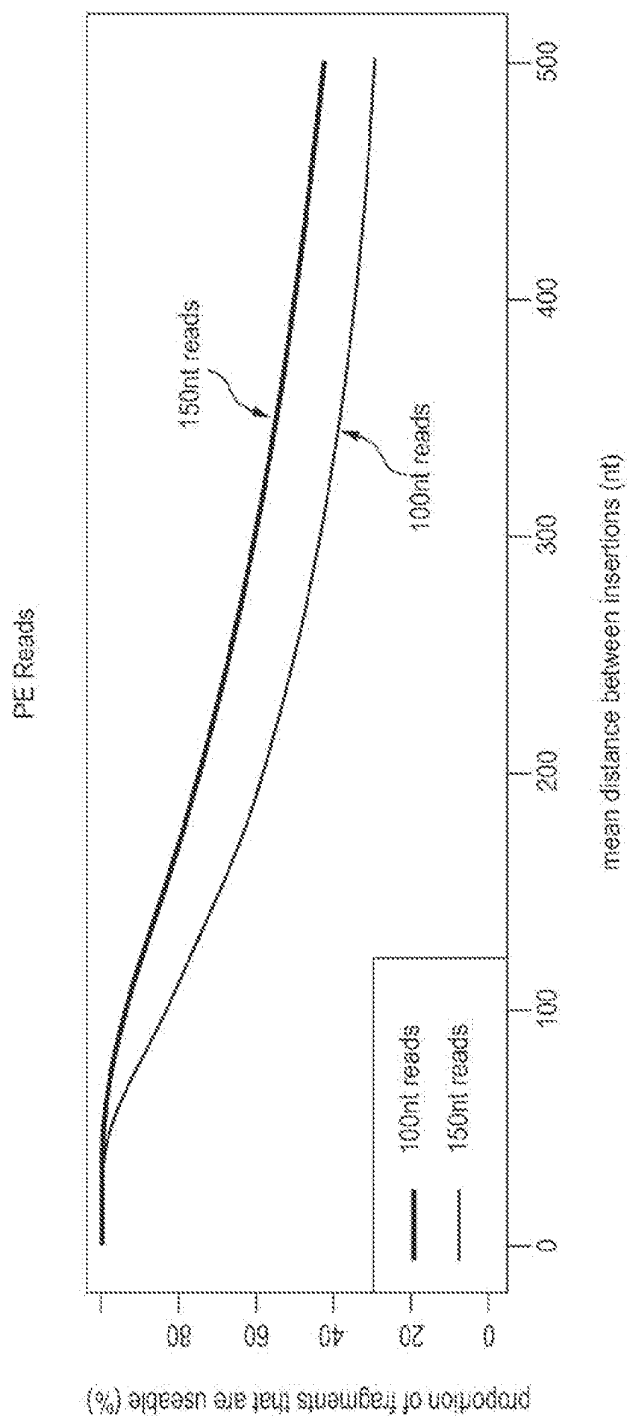


Fig. 36

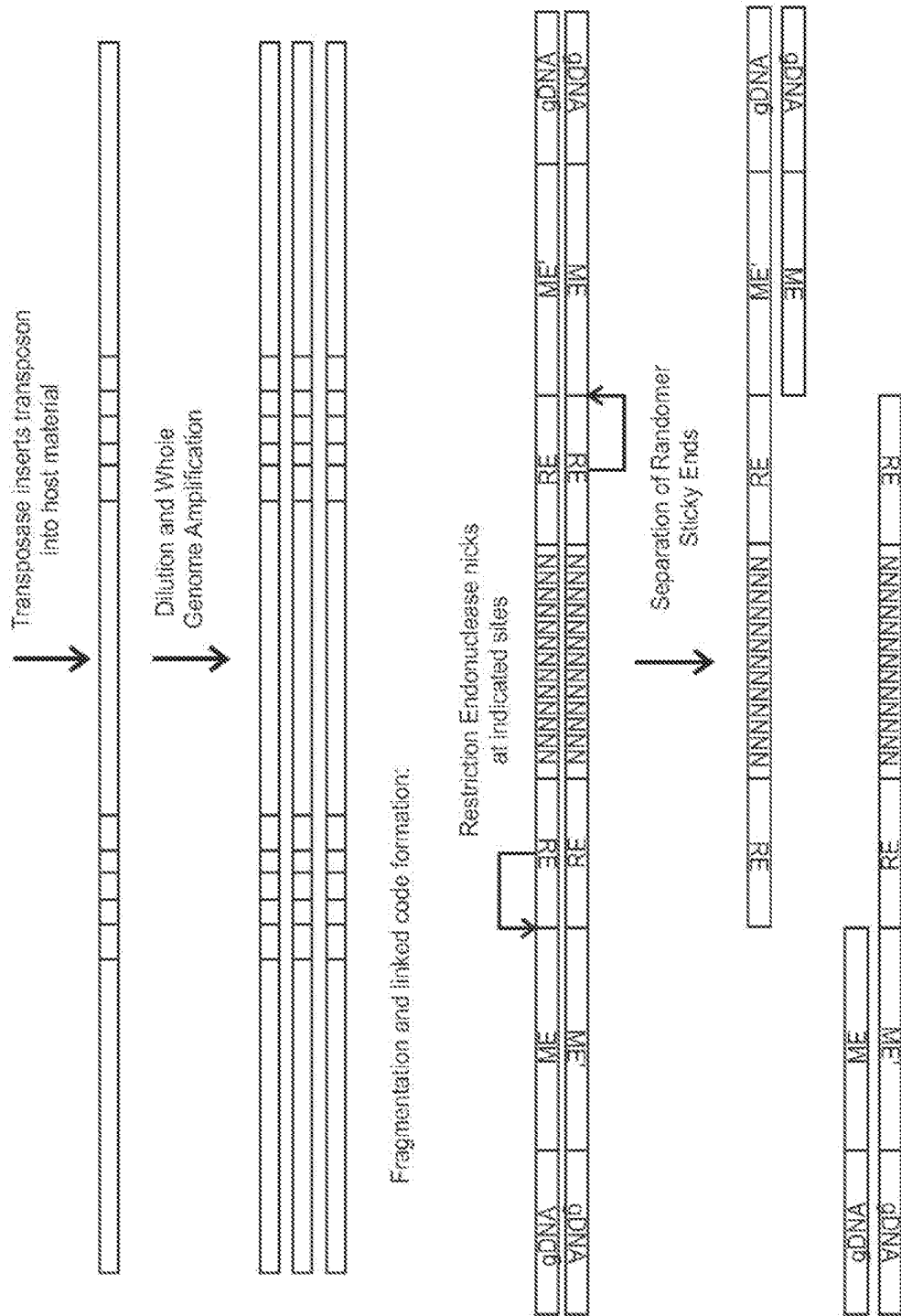


Fig. 37

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2011/059642

A. CLASSIFICATION OF SUBJECT MATTER INV. C12N15/10 C12N15/90 C40B40/08 C40B50/06 ADD.		
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) C12N C40B		
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 practical, search terms used) EP0-Internal, BIOSIS, CAB Data, Sequence Search, WPI Data		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2010/048605 A1 (EPICT TECHNOLOGIES CORP [US]; JENDRISAK JEROME [US]; DAHL GARY [US]; G) 29 April 2010 (2010-04-29) the whole document	1-20
A	SYED FRAZ ET AL: "Optimized library preparation method for next-generation sequencing", NATURE METHODS, vol. 6, no. 10, October 2009 (2009-10), pages 1-11, XP002672202, ISSN: 1548-7091 the whole document ----- -/--	1-20
☒ Further documents are listed in the continuation of Box C. ☒ See patent family annex.		
* 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 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 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 "Z" document member of the same patent family		
Date of the actual completion of the international search		Date of mailing of the international search report
23 March 2012		10/04/2012
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3015		Authorized officer Hornig, Horst

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2011/059642

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	BIMBER BENJAMIN N ET AL: "Whole-genome characterization of human and simian immunodeficiency virus intrahost diversity by ultradeep pyrosequencing.", JOURNAL OF VIROLOGY NOV 2010 LNKD-PUBMED:20844037, vol. 84, no. 22, 15 September 2010 (2010-09-15), pages 12087-12092, XP002672203, ISSN: 1098-5514 the whole document -----	1-20
A	WO 2005/100585 A2 (EPICT [US]; FIANDT MICHAEL J [US]; DAHL GARY A [US]) 27 October 2005 (2005-10-27) the whole document -----	1-20
A	US 7 138 267 B1 (JENDRISAK JEROME J [US] ET AL) 21 November 2006 (2006-11-21) the whole document -----	1-20
X,P	WANG HAOYI ET AL: "Calling Cards enable multiplexed identification of the genomic targets of DNA-binding proteins.", GENOME RESEARCH MAY 2011 LNKD-PUBMED:21471402, vol. 21, no. 5, May 2011 (2011-05), pages 748-755, XP002672204, ISSN: 1549-5469 the whole document -----	1-20
X,P	MARINE RACHEL ET AL: "Evaluation of a transposase protocol for rapid generation of shotgun high-throughput sequencing libraries from nanogram quantities of DNA.", APPLIED AND ENVIRONMENTAL MICROBIOLOGY NOV 2011 LNKD- PUBMED:21948828, vol. 77, no. 22, 23 September 2011 (2011-09-23), pages 8071-8079, XP002672205, ISSN: 1098-5336 the whole document -----	1-20

Form PCT/ISA/210 (continuation of second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2011/059642

Patent document cited in search report	Publication date	Patent family member(s)	Publication date	
WO 2010048605	A1	29-04-2010	CA 2750054 A1	29-04-2010
			CN 102264914 A	30-11-2011
			EP 2376517 A1	19-10-2011
			WO 2010048605 A1	29-04-2010
WO 2005100585	A2	27-10-2005	US 2006014169 A1	19-01-2006
			WO 2005100585 A2	27-10-2005
US 7138267	B1	21-11-2006	NONE	

Form PCT/ISA/210 (patent family annex) (April 2005)

301738-2004
J Bolland.

PATENT COOPERATION TREATY

From the INTERNATIONAL SEARCHING AUTHORITY

PCT

NOTIFICATION OF TRANSMITTAL OF
THE INTERNATIONAL SEARCH REPORT AND
THE WRITTEN OPINION OF THE INTERNATIONAL
SEARCHING AUTHORITY, OR THE DECLARATION

(PCT Rule 44.1)

To: MICHAEL S. TUSCAN COOLEY LLP 1299 PENNSYLVANIA AVENUE, N.W. SUITE 700 WASHINGTON, DC 20004-2400	
Date of mailing (day/month/year)	
Applicant's or agent's file reference NUGN-00101WO	FOR FURTHER ACTION See paragraphs 1 and 4 below
International application No. PCT/US2014/065530	International filing date (day/month/year) 13 November 2014
Applicant NUGEN TECHNOLOGIES, INC.	

1. ☒ The applicant is hereby notified that the international search report and the written opinion of the International Searching Authority have been established and are transmitted herewith.

Filing of amendments and statement under Article 19:
The applicant is entitled, if he so wishes, to amend the claims of the international application (see Rule 46).
When? The time limit for filing such amendments is normally two months from the date of transmittal of the international search report.
How? Directly to the International Bureau of WIPO preferably through ePCT or on paper to, 34 chemin des Colombettes
1211 Geneva 20, Switzerland, Facsimile No.: +41 22 338 82 70
For more detailed instructions, see PCT Applicant's Guide, International Phase, paragraphs 9.004 - 9.011.

2. ☐ The applicant is hereby notified that no international search report will be established and that the declaration under Article 17(2)(a) to that effect and the written opinion of the International Searching Authority are transmitted herewith.

3. ☐ With regard to any protest against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that:
☐ the protest together with the decision thereon has been transmitted to the International Bureau together with any request to forward the texts of both the protest and the decision thereon to the designated Offices.
☐ no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.

4. Reminders

The applicant may submit comments on an informal basis on the written opinion of the International Searching Authority to the International Bureau. These comments will be made available to the public after international publication. The International Bureau will send a copy of such comments to all designated Offices unless an international preliminary examination report has been or is to be established.

Shortly after the expiration of 18 months from the priority date, the international application will be published by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau before the completion of the technical preparations for international publication (Rules 90bis.1 and 90bis.3).

Within 19 months from the priority date, but only in respect of some designated Offices, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase until 30 months from the priority date (in some Offices even later); otherwise, the applicant must, within 20 months from the priority date, perform the prescribed acts for entry into the national phase before those designated Offices. In respect of other designated Offices, the time limit of 30 months (or later) will apply even if no demand is filed within 19 months. For details about the applicable time limits, Office by Office, see www.wipo.int/pct/en/texts/time_limits.html and the PCT Applicant's Guide, National Chapters.

Within 19 months from the priority date, the applicant may request that a supplementary international search be carried out by a different International Searching Authority that offers this service (Rule 45bis.1). The procedure for requesting supplementary international search is described in the PCT Applicant's Guide, International Phase, paragraphs 8.006-8.032.

Name and mailing address of the ISA/ Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3291	Authorized officer Blaine R. Copenhaver PCT Helpdesk: 571-272-4000 Telephone No. 571-272-7774
---	--

Form PCT/ISA/220 (July 2014)

Abstract 4-5-2015
P419 Demand 5-5-2015
Written Opinion 9-13-2015

PATENT COOPERATION TREATY

From the INTERNATIONAL SEARCHING AUTHORITY

To: MICHAEL S. TUSCAN COOLEY LLP 1289 PENNSYLVANIA AVENUE, N.W. SUITE 700 WASHINGTON, DC 20004-2400
--

PCT

NOTIFICATION OF TRANSMITTAL OF
 THE INTERNATIONAL SEARCH REPORT AND
 THE WRITTEN OPINION OF THE INTERNATIONAL
 SEARCHING AUTHORITY, OR THE DECLARATION

(PCT Rule 44.1)

Applicant's or agent's file reference NUGN-00101WO	Date of mailing (day/month/year) 05 MAR 2015
International application No. PCT/US2014/065530	International filing date (day/month/year) 13 November 2014
Applicant NUGEN TECHNOLOGIES, INC.	

<p>1. ☒ The applicant is hereby notified that the international search report and the written opinion of the International Searching Authority have been established and are transmitted herewith.</p> <p>Filing of amendments and statement under Article 19: The applicant is entitled, if he so wishes, to amend the claims of the international application (see Rule 46). When? The time limit for filing such amendments is normally two months from the date of transmittal of the international search report. How? Directly to the International Bureau of WIPO preferably through ePCT or on paper to, 34 chemin des Colombettes 1211 Geneva 20, Switzerland, Facsimile No.: +41 22 538 82 70 For more detailed instructions, see <i>PCT Applicant's Guide</i>, International Phase, paragraphs 9.004 - 9.011.</p> <p>2. ☐ The applicant is hereby notified that no international search report will be established and that the declaration under Article 17(2)(a) to that effect and the written opinion of the International Searching Authority are transmitted herewith.</p> <p>3. ☐ With regard to any protest against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that: ☐ the protest together with the decision thereon has been transmitted to the International Bureau together with any request to forward the texts of both the protest and the decision thereon to the designated Offices. ☐ no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.</p> <p>4. Reminders The applicant may submit comments on an informal basis on the written opinion of the International Searching Authority to the International Bureau. These comments will be made available to the public after international publication. The International Bureau will send a copy of such comments to all designated Offices unless an international preliminary examination report has been or is to be established. Shortly after the expiration of 18 months from the priority date, the international application will be published by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau before the completion of the technical preparations for international publication (Rules 90bis.1 and 90bis.3). Within 19 months from the priority date, but only in respect of some designated Offices, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase until 30 months from the priority date (in some Offices even later); otherwise, the applicant must, within 20 months from the priority date, perform the prescribed acts for entry into the national phase before those designated Offices. In respect of other designated Offices, the time limit of 30 months (or later) will apply even if no demand is filed within 19 months. For details about the applicable time limits, Office by Office, see www.wipo.int/pct/en/texts/time_limits.html and the <i>PCT Applicant's Guide</i>, National Chapters. Within 19 months from the priority date, the applicant may request that a supplementary international search be carried out by a different International Searching Authority that offers this service (Rule 45bis.1). The procedure for requesting supplementary international search is described in the <i>PCT Applicant's Guide</i>, International Phase, paragraphs 8.006-8.032.</p>

Name and mailing address of the ISA/ Mail Stop PCT, After ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer Elaine R. Copenhaver PCT Helpdesk: 571-272-4320 Telephone No. 571-272-7774
---	--

Form PCT/ISA/220 (July 2014)

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference NLGN-001D1WO	FOR FURTHER ACTION see Form PCT/ISA/220 as well as, where applicable, item 5 below.	
International application No. PCT/US2014/065530	International filing date (day/month/year) 13 November 2014	(Earliest) Priority Date (day/month/year) 13 November 2013
Applicant NUSEN TECHNOLOGIES, INC.		

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of 4 sheets.

☐ It is also accompanied by a copy of each prior art document cited in this report.

1. Basis of the report

a. With regard to the language, the international search was carried out on the basis of:

- ☒ the international application in the language in which it was filed.
☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

b. ☐ This international search report has been established taking into account the rectification of an obvious mistake authorized by or notified to this Authority under Rule 91 (Rule 43.6bis(a)).

c. ☒ With regard to any nucleotide and/or amino acid sequence disclosed in the international application, see Box No. I.

2. ☐ Certain claims were found unsearchable (see Box No. II).

3. ☐ Unity of invention is lacking (see Box No. III).

4. With regard to the title,

- ☒ the text is approved as submitted by the applicant.
☐ the text has been established by this Authority to read as follows:

5. With regard to the abstract,

- ☐ the text is approved as submitted by the applicant.
☒ the text has been established, according to Rule 38.2, by this Authority as it appears in Box No. IV. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. With regard to the drawings,

- a. the figure of the drawings to be published with the abstract is Figure No. 3
☒ as suggested by the applicant.
☐ as selected by this Authority, because the applicant failed to suggest a figure.
☐ as selected by this Authority, because this figure better characterizes the invention.
- b. ☐ none of the figures is to be published with the abstract.

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

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2014/065530

Box No. 1 Nucleotide and/or amino acid sequence(s) (Continuation of item 1.a of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing filed or furnished:

a. (means)

☐

on paper

☒

in electronic form

b. (time)

☒

in the international application as filed

☐

together with the international application in electronic form

☐

subsequently to this Authority for the purposes of search

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

Form PCT/ISA/210 (continuation of first sheet (1)) (July 2009)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2014/065530

Box No. IV Text of the abstract (Continuation of item 5 of the first sheet)

The present invention provides methods, compositions and kits for detecting duplicate sequencing reads. In some embodiments, the duplicate sequencing reads are removed. The present invention is based, in part, on compositions and methods for discerning duplicate sequencing reads from a population of sequencing reads. The detection and/or removal of duplicate sequencing reads presented herein is a novel approach to increasing the efficacy of evaluating data generated from high throughput sequence reactions, including complex multiplex sequence reactions.

Form PCT/ISA/210 (continuation of first sheet (3)) (July 2009)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2014/069530

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12N 15/11 (2015.01)

CPC - C12Q 1/6869 (2015.01)

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) - C12N 15/09, 15/10, 15/11; C12P 19/34, C12Q 1/68; C40B 20/00, 20/04 (2015.01)

USPC - 435/6.1, 6.11; 506/2, 23; 536/23.1, 24.2, 25.3

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

CPC - C12Q 1/6869, 1/6874, 2525/18, 2537/143 (2015.01) (keyword delimited)

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

PatBase, Google Patents, PubMed, Google

Search terms used: nucleic acid barcode tag* duplicate* identifier* adapter* sequence*

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2008/093088 A2 (SOLEXA LIMITED) 07 August 2008 (07.08.2008) entire document	21
Y		7, 8, 10-12, 20
Y	US 2012/0208705 A1 (STEEMERS et al) 16 August 2012 (16.08.2012) entire document	1-20
Y	MEYERS, R. 'Myers lab microRNA-seq Protocol, Hudson Alpha Institute for Biotechnology,' 02 May 2013 (02.05.2013), Pgs. 1-15. Retrieved from the Internet: < https://www.encodeproject.org/documents/9f88a08c-d145-447a-a806-e7f3dc44b545/@download/attachment/HA_Myers_miRNA-seq_Protocol_2013-05-02.pdf > on 26 February 2015 (28.02.2015). entire document	1-20
Y	WO 2012/061832 A1 (ILLUMINA, INC.) 10 May 2012 (10.05.2012) entire document	2
P, Y	WO 2014/144022 A1 (NUGEN TECHNOLOGIES, INC.) 18 September 2014 (18.09.2014)	1-21

☐ 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

"I" 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

"G" document member of the same patent family

Date of the actual completion of the international search

18 February 2015

Date of mailing of the international search report

05 MAR 2015

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. Copenhaver

PCT Helpdesk: 571-272-4950

PCT 656P: 571-272-7114

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

PATENT COOPERATION TREATY

From the
INTERNATIONAL SEARCHING AUTHORITY

To: MICHAEL S. TUSCAN
COOLEY LLP
1289 PENNSYLVANIA AVENUE, N.W.
SUITE 700
WASHINGTON, DC 20004-2400

PCT

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43bis.1)

Date of mailing
(day/month/year) **05 MAR 2015**

Applicant's or agent's file reference
NUGN-00101WO

FOR FURTHER ACTION

See paragraph 2 below

International application No.
PCT/US2014/066530

International filing date (day/month/year)
13 November 2014

Priority date (day/month/year)
13 November 2013

International Patent Classification (IPC) or both national classification and IPC
IPC(8) - C12N 15/11 (2015.01)
CPC - C12Q 1/68869 (2015.01)

Applicant NUGEN TECHNOLOGIES, INC.

1. This opinion contains indications relating to the following items:

- ☒ Box No. I Basis of the opinion
- ☐ Box No. II Priority
- ☐ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- ☐ Box No. IV Lack of unity of invention
- ☒ Box No. V Reasoned statement under Rule 43bis.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
- ☐ Box No. VI Certain documents cited
- ☐ Box No. VII Certain defects in the international application
- ☐ Box No. VIII Certain observations on the international application

2. FURTHER ACTION

If a demand for international preliminary examination is made, this opinion will be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notified the International Bureau under Rule 66.1 bis(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

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

Date of completion of this opinion
18 February 2015

Authorized officer:

Brian R. Copenheaver

PCT Helpdesk: 877-272-4505
PCT OSP: 571-272-7754

Form PCT/ISA/237 (cover sheet) (July 2011)

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.

PCT/US2014/005930

Box No. I Basis of this opinion

1. With regard to the language, this opinion has been established on the basis of:
☒ the international application in the language in which it was filed.
☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 25.1(b)).
2. ☐ This opinion has been established taking into account the rectification of an obvious mistake authorized by or notified to this Authority under Rule 91 (Rule 43bis J(a)).
3. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this opinion has been established on the basis of a sequence listing filed or furnished:
 - a. (means)
☐ on paper
☒ in electronic form
 - b. (time)
☒ in the international application as filed
☐ together with the international application in electronic form
☐ subsequently to this Authority for the purposes of search
4. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
5. Additional comments:

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.

PCT/US2014/065530

Box No. V	Reasoned statement under Rule 43bis.1(s)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement			
1.	Statement			
Novelty (N)	Claims	1-20		YES
	Claims	21		NO
Inventive step (IS)	Claims	None		YES
	Claims	1-21		NO
Industrial applicability (IA)	Claims	1-21		YES
	Claims	None		NO
2.	Citations and explanations:			
Claim 21 lacks novelty under PCT Article 33(2) as being anticipated by SOLEXA LIMITED (hereafter SOLEXA LIMITED).				
Regarding claim 21, SOLEXA LIMITED discloses a kit (A kit, Pg. 10, Ln. 26) comprising a plurality of adaptors (A kit comprising a double stranded nucleic acid adaptor, Pg. 10, Lns. 28-29; using universal adaptors, Pg. 11, Lns. 5-8), wherein each adaptor comprises: (i) an indexing primer binding site; (ii) an indexing site; (iii) an identifier site; and (iv) a target sequencing primer binding site (The 'universal' forward and reverse primers include sequences capable of annealing to the 'universal' primer binding sequences in the template construct, Pg. 4, Lns. 5-6; provided that each template is modified by addition of the same universal primer-binding sequences to its 5' and 3' ends; target sequence can therefore be any DNA fragment of interest Pg. 4, Lns. 13-17; each DNA sample, Pg. 4, Ln. 25; the tag and target are read in a single read, Pg. 13, Lns. 23-30; wherein according to the instant application, 'An indexing primer binding site is a nucleotide sequence for binding a primer for an indexing site', Para. [0029]).				
Claims 1, 3, 4-6, 9, and 13-19 lack an inventive step under PCT Article 33(3) as being obvious over Sleemans et al. (hereafter Sleemans) in view of Meyers (hereafter Meyers).				
Regarding claim 1, Sleemans discloses a method for detecting a duplicate sequencing read from a population of sample sequencing reads (in some embodiments, the assembling step further comprises identifying more than one sequencing read comprising the same host tags, Para. [0048]; a sequencing read can be carried out on a single target nucleic acid molecule or simultaneously on a population of target nucleic acid molecules having the same sequence, Para. [0052]; methods described herein can be advantageously carried out in multiplex formats, Para. [0149]) comprising: a) ligating an adaptor to an end of each nucleic acid fragment of a plurality of nucleic acid fragments from one or more samples (The sub-sampled target DNA; adaptors are ligated to...ends of each nucleic acid, Para. [0167]; some embodiments include populations of genomic DNA fragments, Para. [0058]), wherein the adaptor comprises: (ii) an identifier site (in some embodiments, the assembling step further comprises identifying more than one sequencing read comprising the same host tags, Para. [0048]; in some embodiments, the host tag comprises nine nucleotides, Para. [0054]; wherein according to the instant application, 'An identifier site is a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads', Para. [0030]); and (iv) a target sequence primer binding site (Sequencing adapter comprising a first primer site, Para. [0039]; wherein the polynucleotide strand hybridizes to one of the primer sites, Para. [0106]); b) amplifying the adaptor-nucleic acid fragment ligated products; c) generating a population of sequencing reads from amplified adaptor-nucleic acid fragment ligated products (The sub-sampled target DNA; adaptors are ligated to...ends of each nucleic acid, Para. [0167]; steps include...adaptor ligation; and PCR amplify...and sequence products, Para. [0066]; some embodiments include populations of genomic DNA fragments, Para. [0058]; and d) detecting the population of sequencing reads comprising a sequencing read with a duplicate identifier site and target sequence (in some embodiments, the assembling step further comprises identifying more than one sequencing read comprising the same host tags, Para. [0048]; in some embodiments, the host tag comprises nine nucleotides, Para. [0054]; wherein according to the instant application, 'An identifier site is a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads', Para. [0030]; a sequencing read can be carried out on a single target nucleic acid molecule or simultaneously on a population of target nucleic acid molecules having the same sequence, Para. [0052]; methods described herein can be advantageously carried out in multiplex formats, Para. [0149]), but Sleemans fails to explicitly disclose a) ligating an adaptor to a 5' end of each nucleic acid, wherein the adaptor comprises: (i) an indexing primer binding site; (ii) an indexing site. However, Meyers is in the field of adaptor oligo, Pg. 1, second full paragraph and teaches a) ligating an adaptor to a 5' end of each nucleic acid, wherein the adaptor comprises: (i) an indexing primer binding site; (ii) an indexing site (An adaptor oligo...referred to as the 5' adaptor) is ligated to the 5' end. PCR primers have long tails (~30 nt) that extend the length of the product; the tails contain the barcoding sequences (with an index read primer binding site), Pg. 1, second full paragraph; wherein according to the instant application, 'An indexing primer binding site is a nucleotide sequence for binding a primer for an indexing site', Para. [0029]). It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the method of Sleemans to include a) ligating an adaptor to a 5' end of each nucleic acid, wherein the adaptor comprises: (i) an indexing primer binding site; (ii) an indexing site as taught by Meyers as adaptors ligated on the 5' end and PCR primers containing barcoding sequences (with an index read primer binding site) may efficiently be used to generate sequence reads (Myers, Pg. 1, second full paragraph).				

Form PCT/ISA/237 (Box No. V) (July 2011)

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.

PCT/US2014/085530

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

Regarding claim 3, modified Steemers discloses the method of claim 1, modified Steemers further discloses wherein the identifier site is sequenced (in some embodiments, the assembling step further comprises identifying more than one sequencing read comprising the same host tags, Para. [0046]; in some embodiments, the host tag comprises nine nucleotides, Para. [0064]; wherein according to the instant application, 'An identifier site is a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads', Para. [0030]; a sequencing read can be carried out on a single target nucleic acid molecule or simultaneously on a population of target nucleic acid molecules having the same sequence, Para. [0092]), but modified Steemers fails to explicitly disclose wherein the indexing site is sequenced.

However, Myers is in the field of adaptor oligo, Pg. 1, second full paragraph and teaches a indexing site is sequenced (The preparation of multiplexed (barcoded) libraries... suitable for sequencing on the Illumina HiSeq, Pg. 1, first full paragraph; an adaptor oligo... (referred to as the 5' adaptor) is ligated to the 5' end; PCR primers have long tails (~30 nt) that extend the length of the product; the tails contain the barcoding sequences (with an index read primer binding site), the Illumina sequencing primer site, Pg. 1, second full paragraph; wherein according to the instant application, 'An indexing primer binding site is a nucleotide sequence for binding a primer for an indexing site', Para. [0029]).

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the method of Steemers to include a indexing site is sequenced as taught by Myers as multiplexed (barcoded) libraries can be sequenced with an index read primer binding site to produce efficient sequence reads.

Furthermore, it would have been obvious to one of ordinary skill in the art at the time of the invention to have the identifier site sequenced with the indexing site, since the use of a one piece sequence construction instead of the separate sequence pieces disclosed in the prior art would have required only routine skill in the art. The motivation for doing so would be to ensure that both regions are equally represented in sequence reads.

Regarding claim 4, modified Steemers discloses the method of claim 1, modified Steemers further discloses wherein the identifier site is sequenced (in some embodiments, the assembling step further comprises identifying more than one sequencing read comprising the same host tags, Para. [0046]; in some embodiments, the host tag comprises nine nucleotides, Para. [0064]; wherein according to the instant application, 'An identifier site is a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads', Para. [0030]; a sequencing read can be carried out on a single target nucleic acid molecule or simultaneously on a population of target nucleic acid molecules having the same sequence, Para. [0092]), but modified Steemers fails to explicitly disclose wherein the indexing site is sequenced.

However, Myers is in the field of adaptor oligo, Pg. 1, second full paragraph and teaches a indexing site is sequenced (The preparation of multiplexed (barcoded) libraries... suitable for sequencing on the Illumina HiSeq, Pg. 1, first full paragraph; an adaptor oligo... (referred to as the 5' adaptor) is ligated to the 5' end; PCR primers have long tails (~30 nt) that extend the length of the product; the tails contain the barcoding sequences (with an index read primer binding site), the Illumina sequencing primer site, Pg. 1, second full paragraph; wherein according to the instant application, 'An indexing primer binding site is a nucleotide sequence for binding a primer for an indexing site', Para. [0029]).

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the method of Steemers to include a indexing site is sequenced as taught by Myers as multiplexed (barcoded) libraries can be sequenced with an index read primer binding site to produce efficient sequence reads.

Furthermore, it would have been obvious to one of ordinary skill in the art at the time of the invention to have the identifier site sequenced separately from the indexing site, since separating sequences that were once integral involves only routine skill in the art. The motivation for doing so would be to ensure that both regions are separated during sequence reading.

Regarding claim 5, modified Steemers discloses the method of claim 1, modified Steemers further discloses wherein the identifier site is sequenced with the target sequence (in some embodiments, the assembling step further comprises identifying more than one sequencing read comprising the same host tags, Para. [0046]; in some embodiments, the host tag comprises nine nucleotides, Para. [0064]; wherein according to the instant application, 'An identifier site is a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads', Para. [0030]; a sequencing read can be carried out on a single target nucleic acid molecule or simultaneously on a population of target nucleic acid molecules having the same sequence, Para. [0092]; methods described herein can be advantageously carried out in multiplex formats, Para. [0149]).

Regarding claim 6, modified Steemers discloses the method of claim 1, modified Steemers further discloses wherein the identifier site is sequenced from the target sequence (in some embodiments, the assembling step further comprises identifying more than one sequencing read comprising the same host tags, Para. [0046]; in some embodiments, the host tag comprises nine nucleotides, Para. [0064]; wherein according to the instant application, 'An identifier site is a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads', Para. [0030]; a sequencing read can be carried out on a single target nucleic acid molecule or simultaneously on a population of target nucleic acid molecules having the same sequence, Para. [0092]; methods described herein can be advantageously carried out in multiplex formats, Para. [0149]), but modified Steemers fails to explicitly disclose wherein the identifier site is sequenced separately from the target sequence.

It would have been obvious to one of ordinary skill in the art at the time of the invention to have the identifier site sequenced separately from the target sequence, since separating parts that were once integral involves only routine skill in the art. The motivation for doing so would be to ensure that each sequence region alone is most efficiently amplified in optimized sequence reads.

Regarding claim 9, modified Steemers discloses the method of claim 1, modified Steemers further discloses wherein the parity of nucleic acid fragments is generated from more than one sample (The sub-sampled target DNA; adaptors are ligated to ends of each nucleic acid, Para. [0167]; steps include... adaptor ligation; and PCR amplify... and sequence products, Para. [0080]; some embodiments include populations of genomic DNA fragments, Para. [0058]).

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US2014/065530

Supplemental Box

In case the space in any of the preceding boxes is not sufficient,
Continuation of:

Regarding claim 13, modified Steemers discloses the method of claim 1, modified Steemers further discloses wherein the nucleic acid fragments are DNA fragments, RNA fragments, or DNA/RNA fragments (in some embodiments include populations of genomic DNA fragments, Para. [0058]).

Regarding claim 14, modified Steemers discloses the method of claim 13, modified Steemers further discloses wherein the nucleic acid fragments are genomic DNA fragments or cDNA fragments (in some embodiments include populations of genomic DNA fragments, Para. [0058]).

Regarding claim 15, modified Steemers discloses the method of claim 1, but modified Steemers fails to explicitly disclose wherein the indexing site is between 2 and 8 nucleotides in length. However, Myers is in the field of adaptor oligo, Fig. 1, second full paragraph and teaches an indexing site is 30 nucleotides in length (An adaptor oligo... (referred to as the 5' adaptor) is ligated to the 5' end; PCR primers have long tails (~30 nt) that extend the length of the product; the tails contain the barcoding sequences (with an index read primer binding site), Fig. 1, second full paragraph; wherein according to the instant application, 'An indexing primer binding site is a nucleotide sequence for binding a primer for an indexing site', Para. [0029]).

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the method of Steemers to include an indexing site is 30 nucleotides in length as taught by Myers as adaptors ligated on the 5' end and PCR primers containing barcoding sequences (with an index read primer binding site) may efficiently be used to generate sequence reads (Myers, Fig. 1, second full paragraph).

Furthermore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to have an indexing site between 2 and 8 nucleotides in length, since where the general conditions of the claim are disclosed in the prior art, discovering the optimum or workable ranges involves only routine skill in the art. The motivation for doing so would be to ensure that an indexing site between 2 and 8 nucleotides could be identified in the sequence reads.

Regarding claim 16, modified Steemers discloses the method of claim 1, but modified Steemers fails to explicitly disclose wherein the indexing site is about 5 nucleotides in length.

However, Myers is in the field of adaptor oligo, Fig. 1, second full paragraph and teaches an indexing site is 30 nucleotides in length (An adaptor oligo... (referred to as the 5' adaptor) is ligated to the 5' end; PCR primers have long tails (~30 nt) that extend the length of the product; the tails contain the barcoding sequences (with an index read primer binding site), Fig. 1, second full paragraph; wherein according to the instant application, 'An indexing primer binding site is a nucleotide sequence for binding a primer for an indexing site', Para. [0029]).

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the method of Steemers to include an indexing site is 30 nucleotides in length as taught by Myers as adaptors ligated on the 5' end and PCR primers containing barcoding sequences (with an index read primer binding site) may efficiently be used to generate sequence reads (Myers, Fig. 1, second full paragraph).

Furthermore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to have an indexing site about 5 nucleotides in length, since where the general conditions of the claim are disclosed in the prior art, discovering the optimum or workable ranges involves only routine skill in the art. The motivation for doing so would be to ensure that an indexing site about 5 nucleotides in length could be identified in the sequence reads.

Regarding claim 17, modified Steemers discloses the method of claim 1, modified Steemers further discloses wherein the identifier site is 6 nucleotides in length (in some embodiments, the assembling step further comprises identifying more than one sequencing read comprising the same host tags, Para. [0046]; in some embodiments, the host tag comprises nine nucleotides, Para. [0064]; wherein according to the instant application, 'An identifier site is a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads', Para. [0030]), but modified Steemers fails to explicitly disclose wherein the identifier site is between 1 and 8 nucleotides in length. It would have been obvious to one of ordinary skill in the art at the time the invention was made to have an identifier site between 1 and 8 nucleotides in length, since where the general conditions of the claim are disclosed in the prior art, discovering the optimum or workable ranges involves only routine skill in the art. The motivation for doing so would be to ensure that at least 1 to 8 nucleotides of the identifier site could be identified in the sequence reads.

Regarding claim 18, modified Steemers discloses the method of claim 1, modified Steemers further discloses wherein the identifier site is about 9 nucleotides in length (in some embodiments, the assembling step further comprises identifying more than one sequencing read comprising the same host tags, Para. [0046]; in some embodiments, the host tag comprises nine nucleotides, Para. [0064]; wherein according to the instant application, 'An identifier site is a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads', Para. [0030]), but modified Steemers fails to explicitly disclose wherein the identifier site is about 8 nucleotides in length.

It would have been obvious to one of ordinary skill in the art at the time the invention was made to have an identifier site about 8 nucleotides in length, since where the general conditions of the claim are disclosed in the prior art, discovering the optimum or workable ranges involves only routine skill in the art. The motivation for doing so would be to ensure that at least 8 nucleotides of the identifier site could be identified in the sequence reads.

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US2014/066630

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

Regarding claim 19, modified Steemers discloses the method of claim 1, but modified Steemers fails to explicitly disclose wherein the indexing primer binding site is a universal indexing primer binding site. However, Myers is in the field of adaptor oligo... (referred to as the 5' adaptor) is ligated to the 5' end; PCR primers have long tails (~30 nt) that extend the length of the product; the tails contain the barcoding sequences (with an index read primer binding site), Pg. 1, second full paragraph; wherein according to the instant application, 'An indexing primer binding site is a nucleotide sequence for binding a primer for an indexing site', Para. [0029]; an indexing primer binding site is a binding site for a universal indexing primer, i.e., the indexing primer binding site is a universal indexing primer binding site, Para. [0035].

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the method of Steemers to include an indexing primer binding site is a universal indexing primer binding site as taught by Myers as adaptors ligated on the 5' end and PCR primers containing barcoding sequences with an index read primer binding site which can be a universal site, may efficiently be used to generate sequence reads (Myers, Pg. 1, second full paragraph).

Claim 2 lacks an inventive step under PCT Article 33(3) as being obvious over Steemers et al. (hereafter Steemers) in view of Meyers (hereafter Meyers) and ILLUMINA, INC. (hereafter ILLUMINA, INC.).

Regarding claim 2, modified Steemers discloses the method of claim 1, modified Steemers further discloses wherein the method further comprises the population of sequence reads the sequencing read with a duplicate identifier site and target sequence (in some embodiments, the assembling step further comprises identifying more than one sequencing read comprising the same host tags, Para. [0048]; in some embodiments, the host tag comprises nine nucleotides, Para. [0064]; wherein according to the instant application, 'An identifier site is a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads', Para. [0030]; a sequencing read can be carried out on a single target nucleic acid molecule or simultaneously on a population of target nucleic acid molecules having the same sequence, Para. [0032]; methods described herein can be advantageously carried out in multiplex formats, Para. [0149], but Steemers fails to explicitly disclose wherein removing from the population of sequence reads the sequencing read with a duplicate identifier site and target sequence.

However, ILLUMINA, INC. is in the field of removing duplicate sequences, Pg. 42, Lns. 1-3 and teaches removing from the population of sequence reads with a duplicate identifier site and target sequence (in some embodiments, the assembling step further comprises identifying more than one sequencing read comprising the same host tags, Pg. 7, Lns. 1-3; a sequencing read can be carried out on a single target nucleic acid molecule or simultaneously on a population of target nucleic acid molecules having the same sequence, Pg. 18, Lns. 8-10; removing duplicates and counting the original representation of molecules, each having a unique tag; thus, the cDNAs can be amplified and sequenced, Pg. 42, Lns. 1-3).

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the method of Steemers to include removing from the population of sequence reads with a duplicate identifier site and target sequence as taught by ILLUMINA, INC. as removing duplicates and counting the original representation of molecules, each having a unique tag enables nucleic acids can be amplified and sequenced from various populations of sequence reads (ILLUMINA, INC., Pg. 7, Lns. 1-3, Pg. 18, Lns. 8-10, and Pg. 42, Lns. 1-3).

Claims 7, 8, 10-12, and 20 lack an inventive step under PCT Article 33(3) as being obvious over Steemers et al. (hereafter Steemers) in view of Meyers (hereafter Meyers) and SOLEXA LIMITED (hereafter SOLEXA LIMITED).

Regarding claim 7, modified Steemers discloses the method of claim 1, modified Steemers further discloses wherein the adaptor comprises (ii) the identifier site; and (iv) the target sequence primer binding site (in some embodiments, the assembling step further comprises identifying more than one sequencing read comprising the same host tags, Para. [0048]; in some embodiments, the host tag comprises nine nucleotides, Para. [0064]; wherein according to the instant application, 'An identifier site is a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads', Para. [0030]; a sequencing read can be carried out on a single target nucleic acid molecule or simultaneously on a population of target nucleic acid molecules having the same sequence, Para. [0032]; methods described herein can be advantageously carried out in multiplex formats, Para. [0149], but modified Steemers fails to explicitly disclose wherein the adaptor comprises from 5' to 3': (i) the indexing primer binding site; (ii) the indexing site; (iii) the identifier site; and (iv) the target sequence primer binding site.

However, SOLEXA LIMITED is in the field of indexing samples (Title) and teaches an adaptor comprises from 5' to 3' (i) the indexing primer binding site; (ii) the indexing site (The 'universal' forward and reverse primers include sequences capable of annealing to the 'universal' primer binding sequences in the template construct, Pg. 4, Lns. 5-8; provided that each template is modified by addition of the same universal primer-binding sequences to its 5' and 3' ends; target sequence can therefore be any DNA fragment of interest Pg. 4, Lns. 13-17; each DNA sample, Pg. 4, Ln. 28; wherein according to the instant application, 'An indexing primer binding site is a nucleotide sequence for binding a primer for an indexing site', Para. [0029]; the tags are on the adaptors, pg. 20, Ln. 10; ligating identical adaptor polynucleotide molecules... to the 5'... ends of the target polynucleotide, Pg. 20, Lns. 19-20 and 23-24).

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the method of Steemers to include an adaptor comprising from 5' to 3' (i) the indexing primer binding site; (ii) the indexing site as taught by SOLEXA LIMITED as adaptors containing (i) the indexing primer binding site; (ii) the indexing site; (iii) the identifier site; and (iv) the target sequence primer binding site in a 5' to 3' orientation can be most efficient in detecting duplicate sequences.

Furthermore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to have the adaptor comprising from 5' to 3': (i) the indexing primer binding site; (ii) the indexing site; (iii) the identifier site; and (iv) the target sequence primer binding site, since rearranging adaptor parts of an invention involves only routine skill in the art. The motivation for doing so would be to ensure that the 5' to 3' orientation of sequences are most efficient in the determining duplicate reads.

Supplemental Box

In case the space in any of the preceding boxes is not sufficient,
Continuation of:

Regarding claim 8, modified Steemers discloses the method of claim 1, modified Steemers further discloses wherein the adaptor comprises (iii) the target sequence primer binding site; and (iv) the identifier site (in some embodiments, the assembling step further comprises identifying more than one sequencing read comprising the same host tags, Para. [0046]; in some embodiments, the host tag comprises nine nucleotides, Para. [0054]; wherein according to the instant application, 'An identifier site is a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads', Para. [0030]; a sequencing read can be carried out on a single target nucleic acid molecule or simultaneously on a population of target nucleic acid molecules having the same sequence, Para. [0052]; methods described herein can be advantageously carried out in multiplex formats, Para. [0149]), but modified Steemers fails to explicitly disclose wherein the adaptor comprises from 5' to 3' (i) the indexing primer binding site; (ii) the indexing site; (iii) the target sequence primer binding site; and (iv) the identifier site.

However, SOLEXA LIMITED is in the field of indexing samples (Title) and teaches an adaptor comprises from 5' to 3' (i) the indexing primer binding site; (ii) the indexing site (The 'universal' forward and reverse primers include sequences capable of annealing to the 'universal' primer binding sequences in the template construct, Pg. 4, Lns. 5-8; provided that each template is modified by addition of the same universal primer-binding sequences to its 5' and 3' ends; target sequence can therefore be any DNA fragment of interest Pg. 4, Lns. 13-17; each DNA sample, Pg. 4, Ln. 25; wherein according to the instant application, 'An indexing primer binding site is a nucleotide sequence for binding a primer for an indexing site', Para. [0029]; the tags are on the adaptors, Pg. 20, Ln. 10; ligating identical adaptor polynucleotide molecules... to the 5'...ends of the target polynucleotide, Pg. 23, Lns. 19-20 and 23-24).

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the method of Steemers to include an adaptor comprising from 5' to 3' (i) the indexing primer binding site; (ii) the indexing site as taught by SOLEXA LIMITED as adaptors containing (i) the indexing primer binding site; (ii) the indexing site; (iii) the target sequence primer binding site; and (iv) the identifier site in a 5 to 3 orientation can be most efficient in detecting duplicate sequences.

Furthermore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to have the adaptor comprising from 5' to 3' (i) the indexing primer binding site; (ii) the indexing site; (iii) the target sequence primer binding site; and (iv) the identifier site, since rearranging adaptor parts of an invention involves only routine skill in the art. The motivation for doing so would be to ensure that the 5' to 3' orientation of sequences are most efficient in the determining duplicate reads.

Regarding claim 10, modified Steemers discloses the method of claim 5, but modified Steemers fails to explicitly disclose wherein the nucleic acid fragments from each sample has the same indexing site.

However, SOLEXA LIMITED is in the field of indexing samples (Title) and teaches a nucleic acid fragments from each sample has the same indexing site (The universal forward and reverse primers include sequences capable of annealing to the 'universal' primer binding sequences in the template construct, Pg. 4, Lns. 5-8; provided that each template is modified by addition of the same universal primer-binding sequences to its 5' and 3' ends; target sequence can therefore be any DNA fragment of interest Pg. 4, Lns. 13-17; each DNA sample, Pg. 4, Ln. 25; wherein according to the instant application, 'An indexing primer binding site is a nucleotide sequence for binding a primer for an indexing site', Para. [0029]).

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the method of Steemers to include nucleic acid fragments from each sample having the same indexing site as taught by SOLEXA LIMITED as samples having the same indexing site can be sequenced together, eliminating the need for multiple sequencing runs.

Regarding claim 11, modified Steemers discloses the method of claim 10, but modified Steemers fails to explicitly disclose wherein the sequencing reads are separated based on the indexing site.

However, SOLEXA LIMITED is in the field of indexing samples (Title) and teaches sequencing reads are separated based on the indexing site (The universal forward and reverse primers include sequences capable of annealing to the 'universal' primer binding sequences in the template construct, Pg. 4, Lns. 5-8; provided that each template is modified by addition of the same universal primer binding sequences to its 5' and 3' ends; target sequence can therefore be any DNA fragment of interest Pg. 4, Lns. 13-17; each DNA sample, Pg. 4, Ln. 25; the tag and target are read in a single read, Pg. 13, Lns. 29-30; wherein according to the instant application, 'An indexing primer binding site is a nucleotide sequence for binding a primer for an indexing site', Para. [0029]; wherein said sequencing is performed in two separate reads, Pg. 97, Claim 4).

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the method of Steemers to include sequencing reads are separated based on the indexing site as taught by SOLEXA LIMITED as duplicate sequence reads in multiple samples can be determined when sequence reads are separated based on tags and/or identifiers.

Regarding claim 12, modified Steemers discloses the method of claim 11, modified Steemers further discloses wherein the separation of sequencing reads is performed (in some embodiments, the assembling step further comprises identifying more than one sequencing read comprising the same host tags, Para. [0046]), but modified Steemers discloses fails to explicitly disclose wherein the separation of sequencing reads is performed prior to step d).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to have the separation of sequencing reads is performed prior to step d), since selection of any order of process steps which did not result in a new or unexpected result would involve only routine skill in the art. The motivation for doing so would be to ensure the duplicate sequence reads are identified accurately.

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.

PCT/US2014/065530

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

Regarding claim 20, modified *Stemmers* discloses the method of claim 1, but modified *Stemmers* fails to explicitly disclose wherein the target sequence primer binding site is a universal target sequence primer binding site.
However, SOLEXA LIMITED is in the field of *indexing samples* (Title) and teaches a target sequence primer binding site is a universal target sequence primer binding site (The universal forward and reverse primers include sequences capable of annealing to the 'universal' primer binding sequences in the template construct, Pg. 4, Lns. 5-8; provided that each template is modified by addition of the same universal primer-binding sequences to its 5' and 3' ends, target sequence can therefore be any DNA fragment of interest Pg. 4, Lns. 13-17; each DNA sample, Pg. 4, Ln. 25; the tag and target are read in a single read, Pg. 13, Lns. 29-30; wherein according to the instant application, 'An indexing primer binding site is a nucleotide sequence for binding a primer for an indexing site', Para. [0026]).
It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the method of *Stemmers* to include a target sequence primer binding site is a universal target sequence primer binding site as taught by SOLEXA LIMITED as samples having the same universal target sequence primer binding site can be sequenced together, eliminating the need for multiple sequencing runs.

Claims 1-21 meet the criteria set out in PCT Article 36(4), and thus have industrial applicability because the subject matter claimed can be made or used in industry.

Search History:

Limited Classification Search

The Patent Analyst performed a limited classification search within the following US, IPC, CPC, ECLA, or F-Term classification areas:

U.S. Class/Subclass(es): 435/6.1, 6.11; 506/2, 23; 536/23.1, 24.2, 25.3

IPC(8) Class/Subclass(es): C12N 15/09, 15/10, 15/11; C12P 19/34, C12Q 1/68; C40B 20/00, 20/04 (2015.01)

CPC Class/Subclass(es): C12Q 1/6869, 1/6874, 2525/16, 2537/143

See Global Search Results.

Global Patent Literature Text Search

The Patent Analyst performed the following global text search, which was not limited by classification but may or may not have been limited by other criteria:

PATBASE

#	SEARCH	Results
1	PA=("NUGEN_TECHNOLOGIES" OR "NUGEN_TECHNOLOGIES_INC" OR "NUGEN_TECHNOLOGIES_INC201_INDUSTRIAL_RD_SAN_CARLOS" OR "NUGEN_TECHNOLOGIES_INC821_INDUSTRIAL_ROADSAN_CARLOS")	28
2	INV=("AMORESE_DOUGLAS" OR "AMORESE_DOUGLAS_A" OR "AMORESE_DOUGLAS_A_HOCKESSIN_DELAWARE_19707" OR "AMORESE_DOUGLAS_A_HOCKESSIN_DELAWARE_19707_US" OR "AMORESE_DOUGLAS_ALAN")	38
3	INV=("AMORESE_DOUGLAS" OR "AMORESE_DOUGLAS_A" OR "AMORESE_DOUGLAS_A_HOCKESSIN_DELAWARE_19707" OR "AMORESE_DOUGLAS_A_HOCKESSIN_DELAWARE_19707_US" OR "AMORESE_DOUGLAS_ALAN")	38
4	INV=("SCOLNICK_JONATHAN")	4
5	INV=("SCOLNICK_JONATHAN")	4
6	INV=("SCHROEDER_BEN")	1
7	INV=("SCHROEDER_BENJAMIN" OR "SCHROEDER_BENJAMIN_G" OR "SCHROEDER_BENJAMIN_R")	18

	OR "SCHROEDER_BENJAMIN_V")	
8	PN=US20090298075	1
9	PN=WO2011156529	1
10	PN=US20100330574	1
11	PN=WO2008106987	1
12	(C12N15/09 OR C12N15/10 OR C12N15/11 OR C12P19/34 OR C12Q1/68 OR C40B20/00 OR C40B20/04)	1
13	PN=US20120208705	1
14	PN=US20140256571	1
15	PN=US8847799	1
16	PN=WO2008093098	1
17	FT=((nucleic acid) and (barcode) AND (tag*) AND (duplicate*) AND (identifier*) AND (adaptor*))	142
18	FT=((nucleic acid) and (barcode) AND (tag*) AND (duplicate*) AND (identifier*) AND (adaptor*)) and CPC=((C12Q1/6869 OR C12Q1/6874 OR C12Q2537/143 OR C12Q2525/16)) and UC=((435/6.1 OR 435/6.11 OR 506/2 OR 506/23 OR 536/23.1 OR 536/24.2 OR 536/25.3))	25
19	FT=((nucleic acid) and (barcode) AND (tag*) AND (duplicate*) AND (identifier*) AND (adaptor*)) and IC=((C12N15/09 OR C12N15/10 OR C12N15/11 OR C12P19/34 OR C12Q1/68 OR C40B20/00 OR C40B20/04)) and CPC=((C12Q1/6869 OR C12Q1/6874 OR C12Q2537/143 OR C12Q2525/16)) and UC=((435/6.1 OR 435/6.11 OR 506/2 OR 506/23 OR 536/23.1 OR 536/24.2 OR 536/25.3))	25
20	FT=((nucleic acid) and (sequence*) AND (tag*) AND (duplicate*) AND (identifier*) AND (adaptor*)) and IC=((C12N15/09 OR C12N15/10 OR C12N15/11 OR C12P19/34 OR C12Q1/68 OR C40B20/00 OR C40B20/04)) and CPC=((C12Q1/6869 OR C12Q1/6874 OR C12Q2537/143 OR C12Q2525/16)) and UC=((435/6.1 OR 435/6.11 OR 506/2 OR 506/23 OR 536/23.1 OR 536/24.2 OR 536/25.3))	40

GOOGLE PATENT: www.google.com/advanced_patent_search

No. of Hits	Text String
6,240	Single Primer Extension (SPET) Reaction
609	indexing primer binding site
420	method of obtaining duplicate sequence reads, ligated adaptor on 5' end,

	index site
363	method of detecting duplicated sequence, ligated adaptor on 5' end, index site
5,430	Sequencing duplicate nucleic acid fragments
662	Sequencing duplicate nucleic acid fragments and ligated adaptor at 5' end
3,92	identifying a duplicate reads
1,720	identify duplicate sequence reads
517	kit comprising adaptor, index primer binding site, index site
1,110,000	a method for multiplexing duplicate sequencing read
1,020	Barcoding nucleic acids
383,000	universal indexing primer binding site

Computer Accessed Text Databases Searched

The Patent Analyst searched the following computer accessed text databases:

GOOGLE: www.google.com

Search	Query	Items found
#1	A method for detecting a duplicate sequencing read	51,200,000
#2	detecting duplicate sequence reads using multiplex and ligated adaptors	78,800
#3	indexing primer binding site and adaptors	738,000

PUBMED: www.ncbi.nlm.nih.gov/pubmed/advanced

Search	Query	Items found
#1	SCHROEDER, Ben	1
#2	SCOLNICK Jonathan[Author]	6
#3	AMORESE Douglas[Author]	0

Date search was completed: 18 February 2015

DF/JJ

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
7 August 2008 (07.08.2008)

PCT

(10) International Publication Number
WO 2008/093098 A2

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

(74) Agents: WHITE, Nina, Louise et al.; BOULT WADDE
TENNANT, Verulam Gardens, 79 Gray's Inn Road, Lon-
don WC1X 8BT (GB).

(21) International Application Number:
PCT/GB2008/000137

(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, 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, LI, LU, LY, MA, MD, ME, MG, MK, MN,
MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH,
PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV,
SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN,
ZA, ZM, ZW.

(22) International Filing Date: 1 February 2008 (01.02.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/899,221 2 February 2007 (02.02.2007) US

(71) Applicant (for all designated States except US): SOLEXA
LIMITED [GB/GB]; Chesterford Research Park, Little
Chesterford, Saffron Walden Essex CB10 1XL (GB).

(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, MT, NL,
NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG,
CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(72) Inventors; and

(75) Inventors/Applicants (for US only): BIGNALL, Helen
[GB/GB]; SOLEXA LIMITED, Chesterford Research
Park, Little Chesterford, Saffron Walden Essex CB10 1XL
(GB). FRASER, Louise [GB/GB]; SOLEXA LIMITED,
Chesterford Research Park, Little Chesterford, Saffron
Walden Essex CB10 1XL (GB). GORMLEY, Niall
[IE/GB]; SOLEXA LIMITED, Chesterford Research Park,
Little Chesterford, Saffron Walden Essex CB10 1XL (GB).

Published:

— without international search report and to be republished
upon receipt of that report
— with sequence listing part of description published sepa-
rately in electronic form and available upon request from
the International Bureau

WO 2008/093098 A2

(54) Title: METHODS FOR INDEXING SAMPLES AND SEQUENCING MULTIPLE NUCLEOTIDE TEMPLATES

(57) Abstract: The invention relates to methods for indexing samples during the sequencing of polynucleotide templates, resulting in the attachment of tags specific to the source of each nucleic acid sample such that after a sequencing run, both the source and sequence of each polynucleotide can be determined. Thus, the present invention pertains to analysis of complex genomes (e.g., human genomes), as well as multiplexing less complex genomes, such as those of bacteria, viruses, mitochondria, and the like.

- 1 -

Methods for indexing samples and sequencing multiple
polynucleotide templates

Field of the invention

5 The invention relates to methods for indexing samples during the sequencing of polynucleotide templates, resulting in the attachment of tags specific to the source of each nucleic acid sample such that after a sequencing run, both the source and sequence of each polynucleotide can be
10 determined. Thus, the present invention pertains to analysis of complex genomes (e.g., human genomes), as well as multiplexing less complex genomes, such as those of bacteria, viruses, mitochondria, and the like.

15 Background to the invention

 Several publications and patent documents are referenced in this application in order to more fully describe the state of the art to which this invention pertains. The disclosure of each of these publications and
20 documents is incorporated by reference herein.

 Advances in the study of biological molecules have been led, in part, by improvement in technologies used to characterise the molecules or their biological reactions. In particular, the study of the nucleic acids DNA and RNA
25 has benefited from developing technologies used for sequence analysis.

 The study of complex genomes, in particular, the search for the genetic basis of disease in humans requires genotyping on a massive scale. Screens for numerous genetic
30 markers performed on populations large enough to yield statistically significant data are needed before associations can be made between a given genotype and a

- 2 -

particular disease. However large-scale genotyping is demanding in terms of the cost of both materials and labour involved, and the time taken to perform the study, especially if the methodology employed involves separate serial analysis of individual DNA samples. One shortcut is to pool DNA from many individuals and to determine parameters such as the ratio of changes at certain positions in the genome. Such measurements of 'allele frequency' in the pool of samples can be used to correlate the relationship between the changes in the genome sequence and the occurrence of a disease. Hence, an association study involving 1000 patients would in theory only necessitate a 'one-pot' reaction for each genetic change. Pooling therefore represents an effective technique for analysing large quantities of samples in a facile manner.

One disadvantage of pooling samples prior to analysis is that information pertaining to individual DNA samples is lost; only global information such as allele frequencies is gathered, as there is no easy method for discerning which individuals gave rise to a particular genotype. An ability to genotype large populations in a small number of reactions, while retaining the information relating to the source of the individual samples, would yield the information content of a full non-pooled population screen in the time and at the cost of a pooled reaction.

Several of the new methods employed for high throughput DNA sequencing (*Nature*. 437, 376-380 (2005); *Science*. 309, 5741, 1728-1732 (2005)) rely on a universal amplification reaction, whereby a DNA sample is randomly fragmented, then treated such that the ends of the different fragments all contain the same DNA sequence. Fragments with universal ends can be amplified in a single reaction with a single pair of

- 3 -

amplification primers. Separation of the library of fragments to the single molecule level prior to amplification ensures that the amplified molecules form discrete populations that can then be further analysed. Such
5 separations can be performed either in emulsions (*Nature*, 437, 376-380 (2005); *Science*, 309, 5741, 1728-1732 (2005)), or on a surface (*Nucleic Acids Research* 27, e34 (1999); *Nucleic Acids Research* 15, e87 (2000)).

WO 98/44151 and WO 00/18957 both describe methods of
10 forming polynucleotide arrays based on 'solid-phase' nucleic acid amplification, which is a bridging amplification reaction wherein the amplification products are immobilised on a solid support in order to form arrays comprised of nucleic acid clusters or 'colonies'. Each cluster or colony
15 on such an array is formed from a plurality of identical immobilised polynucleotide strands and a plurality of identical immobilised complementary polynucleotide strands. The arrays so-formed are generally referred to herein as 'clustered arrays' and their general features will be
20 further understood by reference to WO 98/44151 or WO 00/18957, the contents of both documents being incorporated herein in their entirety by reference.

In common with all amplification techniques, solid-phase bridging amplification requires the use of forward and
25 reverse amplification primers which include 'template-specific' nucleotide sequences which are capable of annealing to sequences in the template to be amplified, or the complement thereof, under the conditions of the annealing steps of the amplification reaction. The sequences
30 in the template to which the primers anneal under conditions of the amplification reaction may be referred to herein as 'primer-binding' sequences.

- 4 -

Certain embodiments of the methods described in WO 98/44151 and WO 00/18957 make use of 'universal' primers to amplify templates comprising a variable template portion that it is desired to amplify flanked 5' and 3' by common or
5 'universal' primer binding sequences. The 'universal' forward and reverse primers include sequences capable of annealing to the 'universal' primer binding sequences in the template construct. The variable template portion, or 'target' may itself be of known, unknown or partially known
10 sequence. This approach has the advantage that it is not necessary to design a specific pair of primers for each target sequence to be amplified; the same primers can be used for amplification of different templates provided that each template is modified by addition of the same universal
15 primer-binding sequences to its 5' and 3' ends. The variable target sequence can therefore be any DNA fragment of interest. An analogous approach can be used to amplify a mixture of templates (targets with known ends), such as a plurality or library of target nucleic acid molecules (e.g.
20 genomic DNA fragments), using a single pair of universal forward and reverse primers, provided that each template molecule in the mixture is modified by the addition of the same universal primer-binding sequences.

DNA from more than one source can be sequenced on an
25 array if each DNA sample is first tagged to enable its identification after it has been sequenced. Many low scale DNA-tag methodologies already exist, for example fluorescent labelling (Haughland, Handbook of Fluorescent Probes and Research Products, Invitrogen/Molecular Probes), but these
30 are limited in scope to less than 10 or so reactions in parallel. DNA tags can be added to the ends of DNA fragments by cloning, of example as described in US 5604097. The tags

- 5 -

consist of eight four base 'words', where each word uses only three bases (A,T and C) in various combinations resulting in a total of 16,777,216 different tags that all have the same base pair composition and melting points. Such tags are used to label target molecules in a sample so that after an amplification reaction, each original molecule in the sample has a unique tag. The tags can then be used to 'sort' the sample onto beads containing sequences complementary to the tags such that each bead contains multiple copies of a single amplified target sequence (Brenner et al., (2000) Nature Biotechnology, 18, 630). In this application the tags are not sequenced, so the method does not provide a method of analysing targets from multiple samples, but rather a method of sorting a mixture of amplified templates from a single sample. The problem with enabling the method for individual samples rather than individual molecules is that the tags are synthesised in a combinatorial manner, meaning that all 16,777,216 different sequences are obtained in a single mixture in the same tube. Whilst this is ideal for treating one sample such that each individual molecule in the sample carries a different tag, it does not permit attachment of the same tag to every molecule in the sample.

DNA samples from multiple sources can, however, be tagged with different nucleic acid tags such that the source of the sample can be identified. Previous application WO05068656 describes the generic concept of indexing samples. In order to utilise this invention on arrays of amplified single molecule templates, for example as described in WO9844151, WO06099579 or WO04069849, it is advantageous to prepare the nucleic acids using the novel method described herein. The optimised DNA sample

- 6 -

preparation techniques described herein are applicable to any method where the samples are amplified prior to sequencing. The DNA sample preparation techniques presented herein describe in detail the optimal placements of the sequencing primers and indexed tags within the DNA constructs to be sequenced.

Summary of the invention

The present inventors have developed methods for indexing samples, wherein the samples are amplified from isolated single template molecules. Using the techniques of the invention, it is possible to prepare a nucleic acid sample for a sequencing reaction wherein both the target sequence and the tag sequence can be determined.

In a first embodiment of the invention, a method is presented for sequencing nucleic acid sequences on an array and identifying subsets of nucleic acid sequences on an array, wherein each subset of nucleic acid sequences is isolated from a different source, wherein the method comprises the steps of:

- (a) providing at least two samples of randomly fragmented double stranded nucleic acid targets, wherein each of said randomly fragmented double stranded nucleic acid targets is isolated from a different source;
- (b) ligating a universal adaptor to the ends of each target fragment of each sample to generate adaptor-target-adaptors of each sample, wherein each of said adaptor-target-adaptors comprises a target fragment flanked by universal adaptor sequences and said universal adaptor comprises a region of double stranded nucleic acid and at least one region of single stranded nucleic acid;

- 7 -

- (c) amplifying adaptor-target-adaptors of each sample with two or more sample specific amplification primers to generate amplified nucleic acids, wherein one of said amplification primers comprises a sample specific tag sequence, and wherein amplified nucleic acids of each sample comprise said sample specific tag sequence and said sample specific tag sequence differentiates amplified nucleic acids originating from different samples;
- (d) pooling the amplified nucleic acids of different samples;
- (e) immobilising the amplified nucleic acids of different samples on an array to generate an array of immobilised fragments;
- (f) sequencing the immobilised fragments on the array to determine a sequence read of each immobilised target fragment and identify the sample specific tag sequence of each immobilised fragment, thereby determining both a nucleic acid sequence of said immobilized fragment and identifying the immobilized fragment as a member of a subset of nucleic acids on the array.

In a second embodiment of the invention, a method is presented for sequencing nucleic acid sequences on an array and identifying subsets of nucleic acid sequences on an array, wherein each subset of nucleic acid sequences is isolated from a different source, wherein the method comprises the steps of:

- (a) providing at least two samples of randomly fragmented double stranded nucleic acid targets, wherein each of said randomly fragmented double stranded nucleic acid targets is isolated from a different source;
- (b) ligating a sample specific tagged adaptor to the ends of each target fragment of each sample to generate adaptor-

- 8 -

target-adaptors of each sample, wherein each of said adaptor-target-adaptors comprises a target fragment flanked by tagged adaptor sequences and said tagged adaptor comprises a region of double stranded nucleic acid and at least one region of single stranded nucleic acid and a sample specific tag that differentiates adaptor-target-adaptors originating from different samples;

(c) pooling the adaptor-target-adaptors of different samples to generate a pooled sample of adaptor-target-adaptors;

(d) amplifying the pooled sample of adaptor-target-adaptors with a pair of universal amplification primers to generate an amplified pooled sample;

(e) immobilising the amplified pooled sample on an array to generate an array of immobilised fragments;

(f) sequencing the array to determine a sequence read of each immobilised target fragment and identify the sample specific tag sequence of each immobilised fragment, thereby determining both a nucleic acid sequence of said immobilized fragment and identifying the immobilized fragment as a member of a subset of nucleic acids on the array.

In a third embodiment of the invention, a method is presented for sequencing nucleic acid sequences on an array and identifying subsets of nucleic acid sequences on an array, wherein each subset of nucleic acid sequences is isolated from a different source, wherein the method comprises the steps of:

(a) providing at least two samples of randomly fragmented double stranded nucleic acid targets, wherein each of said randomly fragmented double stranded nucleic acid targets is isolated from a different source;

- 9 -

- (b) ligating a sample specific tagged adaptor to the ends of each target fragment of each sample to generate adaptor-target-adaptors of each sample, wherein each of said adaptor-target-adaptors comprises a target fragment flanked by tagged adaptor sequences and said tagged adaptor comprises a region of double stranded nucleic acid and at least one region of single stranded nucleic acid and a sample specific tag that differentiates adaptor-target-adaptors originating from different samples, and wherein the sample specific tag sequence is attached directly to the target fragment in each adaptor-target-adaptor with no intervening nucleotides between the sample specific tag sequence and target fragment sequences;
- (c) amplifying adaptor-target-adaptors of each sample with a pair of amplification primers complementary to the sample specific tag sequence to generate amplified nucleic acids from each sample;
- (d) pooling the amplified nucleic acids from each sample to generate a pool of amplified nucleic acid samples;
- (e) immobilising the pool of amplified nucleic acid samples on an array to generate an array of immobilised fragments;
- (f) sequencing the array to determine a sequence read of each immobilised target fragment and identify the sample specific tag sequence of each immobilised fragment, thereby determining both a nucleic acid sequence of said immobilized fragment and identifying the immobilized fragment as a member of a subset of nucleic acids on the array.

In accordance with the first or second embodiment of the invention, sequencing can be performed in two separate reads. In accordance with the third embodiment of the invention, sequencing of the target fragment and the sample

- 10 -

specific tag sequence can be performed in a single sequencing read.

With respect to the first, second, or third embodiments, the method encompasses a step wherein the
5 pooled samples are amplified on a solid surface prior to sequencing. In a particular aspect, the solid surface on which amplification is performed is a collection of beads. In a more particular aspect, the collection of beads may be isolated as single beads in an emulsion.

10 With respect to the first, second, or third embodiments, the method encompasses a step wherein the array of immobilized fragments is amplified prior to sequencing to generate an amplified array. In a particular aspect, an amplified array so produced is a clustered array of
15 amplified single molecules. Such clustered arrays may be formed, for example, by solid-phase nucleic acid amplification or isothermal solid-phase nucleic acid amplification.

The present method encompasses sequencing that involves
20 cycles of ligation with labelled oligonucleotides or cycles of addition of nucleotides. In a particular aspect of the invention, the nucleotides are labelled. In a more particular aspect the nucleotides are labelled with fluorophores.

25 In an aspect of any of the first, second, or third embodiments of the invention, the sequencing step may be carried out on an array of immobilised beads.

A further aspect of the invention comprises a kit comprising a double stranded nucleic acid adapter and at
30 least two different primer polynucleotides that carry a different tag sequence that does not hybridise to the adapter.

- 11 -

Brief description of the figures

5 **Figure 1** shows a schematic of an indexed sample construct using the first method of the invention, using universal adaptors. Common adaptors are used for all tags, tags are added using one unique amplification primer and sequencing is carried out in two reads.

10 **Figure 2** shows a schematic of an indexed sample construct using the second method of the invention, using tagged adaptors. Tagged adaptors are used for each sample and amplified with one pair of universal primers. Sequencing is carried out in two reads.

15 **Figure 3** shows a schematic of an indexed sample construct using the third method of the invention, using adjacent tag-targets. Two adaptors and two amplification primers are used per tag and sequencing carried out in a single read.

20 **Figure 4** shows numerical data on two sequencing reads obtained from a sample prepared according to the first method of the invention, amplified into an array of clusters and sequenced with two sequencing reads, one for the target
25 and one for the tag.

Figure 5 shows the concept of tagging a sample for use in a paired end reading process. Libraries are amplified at the PCR step using a nested PCR reaction with 3 primers (2
30 universal and 1 tag oligo.

- 12 -

Figure 6a shows a nested PCR reaction using three amplification primers, one of which carries the tag.

Figure 6b shows the same nested PCR concept as figure 6, but with the benefit of the sequences of the various adapters and primers.

Figure 7 shows the embodiment of the method where the first sequencing read is performed, followed by removing the first sequencing primer, hybridising a second sequencing primer to determine the tag sequence, inverting the strand on the surface using a strand resynthesis protocol, and performing a second read at the opposite end of the fragment to the first read to generate a pair of reads from each fragment along with the tag read.

Figure 8 shows some exemplary data obtained using the method shown in figure 7.

Figure 9 shows the sequence of the various oligonucleotides used in the process

Figure 10 shows a schematic of MJ Research thermocycler, coupled with an 8-way peristaltic pump Ismatec IPC ISM931 equipped with Ismatec tubing (orange/yellow, 0.51 mm ID).

Detailed description of the invention

The present invention is directed to methods that advance techniques for nucleic acid analysis, such that target nucleic acids from a plurality of sources can be tagged with a unique, identifiable sequence of nucleic acid bases and then sequenced on a single array. As described

- 13 -

herein, the presence of a characteristic nucleic acid sequence tag on an immobilised target molecule permits identification of the source of the target, concurrent with a sequence read from the target. This is a dramatic
5 improvement over pre-existing array technologies which generally require a two-step process involving an initial sequencing step for sequencing the pooled nucleic acid targets, followed by a second analysis step wherein the source of the nucleic acid is determined.

10 The limitations of such pre-existing array technologies are evident in WO 98/44151 and WO 00/18957, for example, which both describe methods of nucleic acid amplification which allow amplification products to be immobilised on a solid support in order to form arrays comprised of clusters
15 or "colonies" formed from a plurality of identical immobilised polynucleotide strands and a plurality of identical immobilised complementary strands. The nucleic acid molecules present in DNA colonies on the clustered arrays prepared according to these methods can provide
20 templates for sequencing reactions, but to date only a single sequencing read can be obtained from each immobilised strand in each colony. The methods described herein allow for two or more reads on a single amplified strand, thereby allowing analysis of the tag sequence independently of the
25 target sequence, and also allowing the possibility of a paired read from the opposite end of each fragment as well as a third read for the tag sequence. The methods also allow the preparation of constructs where the tag sequence is adjacent to the target sequence, and the tag and target are
30 read in a single read, which again can be turned into a paired read methodology if desired.

- 14 -

As will be apparent to the skilled reader, references herein to a particular nucleic acid sequence may, depending on the context, also refer to nucleic acid molecules which comprise the nucleic acid sequence. Sequencing of a target
5 fragment means that a read of the chronological order of bases is established. The bases do not need to be contiguous, although this is preferred, nor does every base on the entire fragment have to be sequenced.

The following passages describe different aspects of
10 the invention in greater detail. Each aspect of the invention may be combined with any other aspect or aspects of the invention unless clearly indicated to the contrary. In particular, any feature indicated as being preferred or advantageous may be combined with any other feature of
15 features indicated as being preferred or advantageous.

The terms 'target nucleic acid sequence', 'target nucleic acid molecule', 'target nucleic acid' and 'target nucleic acid fragment' may be used interchangeably to refer to nucleic acid molecules that it is desired to sequence on
20 an array according to the invention. The target nucleic acid may be essentially any nucleic acid of known or unknown sequence. It may be, for example, a fragment of genomic DNA or cDNA. Sequencing may result in determination of the sequence of the whole, or a part of the target molecule. The
25 targets can be derived from a primary nucleic acid sample that has been randomly fragmented. The targets can be processed into templates suitable for amplification by the placement of universal amplification sequences at the ends of each target fragment. The targets can also be obtained
30 from a primary RNA sample by reverse transcription into cDNA.

- 15 -

As used herein, the term 'polynucleotide' refers to deoxyribonucleic acid (DNA), but where appropriate the skilled artisan will recognise that the method may also be applied to ribonucleic acid (RNA). The terms should be understood to include, as equivalents, analogs of either DNA or RNA made from nucleotide analogs and to be applicable to single stranded (such as sense or antisense) and double stranded polynucleotides. The term as used herein also encompasses cDNA, that is complementary or copy DNA produced from an RNA template, for example by the action of reverse transcriptase.

The primary polynucleotide molecules may originate in double-stranded DNA (dsDNA) form (e.g. genomic DNA fragments, PCR and amplification products and the like) or may have originated in single-stranded form, as DNA or RNA, and been converted to dsDNA form. By way of example, mRNA molecules may be copied into double-stranded cDNAs suitable for use in the method of the invention using standard techniques well known in the art. The precise sequence of the primary polynucleotide molecules is generally not material to the invention, and may be known or unknown.

In a particular embodiment, the primary polynucleotide molecules are RNA molecules. In an aspect of this embodiment, RNA isolated from specific samples is first converted to double-stranded DNA using techniques known in the art. In accordance with the present method, the double-stranded DNA is then tagged or indexed with a sample specific tag. Different preparations of such double-stranded DNA comprising sample specific tags can be generated, in parallel, from RNA isolated from different specific samples. Subsequently, different preparations of double-stranded DNA comprising different sample specific

- 16 -

tags can be mixed, sequenced *en masse*, and the identity of each sequenced fragment determined with respect to the sample from which it was isolated/derived by virtue of the presence of a sample specific tag.

5 In a particular embodiment, the primary polynucleotide molecules are DNA molecules. More particularly, the primary polynucleotide molecules represent the entire genetic complement of an organism, and are genomic DNA molecules which include both intron and exon sequences (coding
10 sequence), as well as non-coding regulatory sequences such as promoter and enhancer sequences. Although it could be envisaged that particular sub-sets of polynucleotide sequences or genomic DNA could also be used, such as, for example, particular chromosomes. Yet more particularly, the
15 sequence of the primary polynucleotide molecules is not known. Still yet more particularly, the primary polynucleotide molecules are human genomic DNA molecules. The DNA target molecules may be treated chemically or enzymatically either prior to, or subsequent to any random
20 fragmentation processes, and prior to or subsequent to the ligation of the adaptor sequences.

 Methods of the invention are especially useful for the parallel sequencing of a large number of relative small samples in the same experiment. In a sequencing platform
25 with a high degree of parallelism, it may be possible to analyse 50-100 million reads in the same experiment, each to a length of 50 bases or more. Thus it may easily be possible to determine more than 5 billion base pairs of sequence from a single experiment. For a genome of 100,000-1 million base
30 pairs, this represents a much larger coverage of each base pair than is actually needed for accurate sequencing. This feature is even more pronounced for viral or mitochondrial

- 17 -

genomes or collections of PCR fragments that may be 10,000 base pairs or less. For a 10 kB sample, sequenced in a run of 5 billion base pairs, each base pair will appear on average 500,000 times. For accurate sequencing, it may be desirable for each base pair to appear only 20 times, and thus it will be possible to sequence 25,000 10kB fragments in the sequencing run. The use of 8 base pair tags gives a possibility of 4^8 (65536) different samples in the same experiment. For the sake of clarity, sequencing of the tags can be performed in a separate read from the sequencing of the target fragments, so the read length or accuracy of the read from the sample is not lowered by the need to sequence the tags.

For genome sizes of 1 mB, at a depth of 20X, it is possible to sequence 250 samples in a 5 billion base pair experiment. The 250 samples can be coded by as few as a four base tag, but extra bases may be added to ensure accurate analysis. Thus, the primary polynucleotide molecules may be a genome of any organism, but the methods of the present invention may be used to particular advantage when analysing smaller genomes, such as those of bacteria, viruses, mitochondria and the like, which have genome sizes in the regions of a few million base pairs or lower.

Random fragmentation refers to the fragmentation of a polynucleotide molecule in a non-ordered fashion by enzymatic, chemical or mechanical means. Such fragmentation methods are known in the art and utilise standard methods (Sambrook and Russell, Molecular Cloning, A Laboratory Manual, third edition). For the sake of clarity, generating smaller fragments of a larger piece of nucleic acid via specific PCR amplification of such smaller fragments is not equivalent to fragmenting the larger piece of nucleic acid

- 18 -

because the larger piece of nucleic acid sequence remains in intact (i.e., is not fragmented by the PCR amplification). Moreover, random fragmentation is designed to produce fragments irrespective of the sequence identity or position
5 of nucleotides comprising and/or surrounding the break. More particularly, the random fragmentation is by mechanical means such as nebulisation or sonication to produce fragments of about 50 base pairs in length to about 1500 base pairs in length, still more particularly 50-700 base
10 pairs in length, yet more particularly 50-400 base pairs in length. Most particularly, the method is used to generate smaller fragments of from 50-150 base pairs in length.

Fragmentation of polynucleotide molecules by mechanical means (nebulization, sonication and Hydroshear for example)
15 results in fragments with a heterogeneous mix of blunt and 3'- and 5'-overhanging ends. It is therefore desirable to repair the fragment ends using methods or kits (such as the Lucigen DNA terminator End Repair Kit) known in the art to generate ends that are optimal for insertion, for example,
20 into blunt sites of cloning vectors. In a particular embodiment, the fragment ends of the population of nucleic acids are blunt ended. More particularly, the fragment ends are blunt ended and phosphorylated. The phosphate moiety can be introduced via enzymatic treatment, for example, using
25 polynucleotide kinase.

In a particular embodiment, the target polynucleotide sequences are prepared with single overhanging nucleotides by, for example, activity of certain types of DNA polymerase such as Taq polymerase or Klenow exo minus polymerase which
30 has a nontemplate-dependent terminal transferase activity that adds a single deoxynucleotide, for example, deoxyadenosine (A) to the 3' ends of, for example, PCR

- 19 -

products. Such enzymes can be utilised to add a single nucleotide 'A' to the blunt ended 3' terminus of each strand of the target polynucleotide duplexes. Thus, an 'A' could be added to the 3' terminus of each end repaired duplex strand of the target polynucleotide duplex by reaction with Taq or Klenow exo minus polymerase, whilst the adaptor polynucleotide construct could be a T-construct with a compatible 'T' overhang present on the 3' terminus of each duplex region of the adaptor construct. This end modification also prevents self-ligation of both vector and target such that there is a bias towards formation of the combined ligated adaptor-target sequences.

The method of the invention utilises nucleic acid sequence tags as markers characteristic of the source of particular target molecules on the array. A nucleic acid sequence tag characteristic of the source is attached to each of the target molecules in each isolated sample before the sample is immobilised for sequencing. The tag is not itself formed by part of the target nucleic acid molecule, but becomes part of the template for amplification. Generally the tag will be a synthetic sequence of nucleotides which is added to the target as part of the template preparation step. Accordingly, a sample specific tag is a nucleic acid sequence tag which is attached to each of the target molecules of a particular sample, the presence of which is indicative of or is used to identify the sample or source from which the target molecules were isolated.

Preferably the nucleic acid sequence tag may be up to 20 nucleotides in length, more preferably 1-10 nucleotides, and most preferably 4-6 nucleotides in length. A four nucleotide tag gives a possibility of multiplexing 256

- 20 -

samples on the same array, a six base tag enables 4096 samples to be processed on the same array.

In the first embodiment of the invention, the tag sequences are introduced in a two step process involving
5 ligation of an adaptor that is common to all samples and amplification with a tagged amplification primer. See Figures 1 and 6. In the second embodiment the tags are on the adaptor, and the amplification can be performed with universal amplification primers. See Figure 2. In the third
10 embodiment, the tags are on the adaptors, but at the terminus of the duplex region such that the tag and target sequences are contiguous. See Figure 3. The third embodiment allows for the sequencing to be performed in a single read, but requires both strands of the adaptor to be
15 modified and therefore requires more oligonucleotides as two unique adaptor strands are needed for each sample.

In each embodiment, the target nucleic acids of each specific fragmented sample are treated by first ligating identical adaptor polynucleotide molecules ('mismatched
20 adaptors', the general features of which are defined below, and further described in copending application US 20070128624, the contents of which are incorporated herein by reference in their entirety) to the 5' and 3' ends of the target polynucleotide duplexes (which may be of known,
25 partially known or unknown sequence) to form adaptor-target constructs and then carrying out an initial primer extension reaction in which extension products complementary to both strands of each individual adaptor-target construct are formed. The resulting primer extension products, and
30 optionally amplified copies thereof, collectively provide a library of template polynucleotides. The term library refers to the collection of target fragments containing known

- 21 -

common sequences at their 3' and 5' ends, and may also be referred to as a 3' and 5' modified library.

Depending on the embodiment of the invention, the adaptors may be universal for all samples, or one or both
5 strands of the duplexes may carry the tag sequence to code or track the identity of the samples.

The adaptor polynucleotides used in the method of the invention are referred to herein as 'mismatched' adaptors because, as will be explained in detail herein, it is
10 essential that the adaptors include a region of sequence mismatch, i.e., they must not be formed by annealing of fully complementary polynucleotide strands.

Mismatched adaptors for use in the invention are formed by annealing of two partially complementary polynucleotide
15 strands so as to provide, when the two strands are annealed, at least one double-stranded region and at least one unmatched single-stranded region.

The 'double-stranded region' of the adaptor is a short double-stranded region, typically comprising 5 or more
20 consecutive base pairs, formed by annealing of the two partially complementary polynucleotide strands. This term refers to a double-stranded region of nucleic acid in which the two strands are annealed and does not imply any particular structural conformation.

25 Generally it is advantageous for the double-stranded region to be as short as possible without loss of function. In this context, 'function' refers to the ability of the double-stranded region to form a stable duplex under standard reaction conditions for an enzyme-catalysed nucleic
30 acid ligation reaction, which will be well known to the skilled reader (e.g. incubation at a temperature in the range of 4°C to 25°C in a ligation buffer appropriate for

- 22 -

the enzyme), such that the two strands forming the adaptor, remain partially annealed during ligation of the adaptor to a target molecule. It is not absolutely necessary for the double-stranded region to be stable under the conditions typically used in the annealing steps of primer extension or PCR reactions.

Since identical adaptors are ligated to both ends of each target molecule, the target sequence in each adaptor-target construct will be flanked by complementary sequences derived from the double-stranded region of the adaptors. The longer the double-stranded region, and hence the complementary sequences derived therefrom in the adaptor-target constructs, the greater the possibility that the adaptor-target construct is able to fold back and base-pair to itself in these regions of internal self-complementarity under the annealing conditions used in primer extension and/or PCR. It is, therefore, generally preferred for the double-stranded region to be 20 or less, 15 or less, or 10 or less base pairs in length in order to reduce this effect. The stability of the double-stranded region may be increased, and hence its length potentially reduced, by the inclusion of non-natural nucleotides which exhibit stronger base-pairing than standard Watson-Crick base pairs.

It is preferred, but not absolutely essential, for the two strands of the adaptor to be 100% complementary in the double-stranded region. It will be appreciated that one or more nucleotide mismatches may be tolerated within the double-stranded region, provided that the two strands are capable of forming a stable duplex under standard ligation conditions.

Adaptors for use in the invention will generally include a double-stranded region forming the 'ligatable' end

- 23 -

of the adaptor, i.e. the end that is joined to a target polynucleotide in the ligation reaction. The ligatable end of the adaptor may be blunt or, in other embodiments, short 5' or 3' overhangs of one or more nucleotides may be present to facilitate/promote ligation. The 5' terminal nucleotide at the ligatable end of the adaptor is phosphorylated to enable phosphodiester linkage to a 3' hydroxyl group on the target polynucleotide.

The term 'unmatched region' refers to a region of the adaptor wherein the sequences of the two polynucleotide strands forming the adaptor exhibit a degree of non-complementarity such that the two strands are not capable of fully annealing to each other under standard annealing conditions for a primer extension or PCR reaction. The unmatched region(s) may exhibit some degree of annealing under standard reaction conditions for an enzyme-catalysed ligation reaction, provided that the two strands revert to single stranded form under annealing conditions in an amplification reaction.

It is to be understood that the 'unmatched region' is provided by different portions of the same two polynucleotide strands which form the double-stranded region(s). Mismatches in the adaptor construct can take the form of one strand being longer than the other, such that there is a single stranded region on one of the strands, or a sequence selected such that the two strands do not hybridise, and thus form a single stranded region on both strands. The mismatches may also take the form of 'bubbles', wherein both ends of the adaptor construct(s) are capable of hybridising to each other and forming a duplex, but the central region is not. The portion of the strand(s) forming the unmatched region are not annealed under conditions in

- 24 -

which other portions of the same two strands are annealed to form one or more double-stranded regions. For avoidance of doubt it is to be understood that a single-stranded or single base overhang at the 3' end of a polynucleotide duplex that subsequently undergoes ligation to the target sequences does not constitute an 'unmatched region' in the context of this invention.

The lower limit on the length of the unmatched region will typically be determined by function, for example, the need to provide a suitable sequence for binding of a primer for primer extension, PCR and/or sequencing. Theoretically there is no upper limit on the length of the unmatched region, except that in general it is advantageous to minimise the overall length of the adaptor, for example, in order to facilitate separation of unbound adaptors from adaptor-target constructs following the ligation step. Therefore, it is generally preferred that the unmatched region should be less than 50, or less than 40, or less than 30, or less than 25 consecutive nucleotides in length.

The precise nucleotide sequence of the adaptors is generally not material to the invention and may be selected by the user such that the desired sequence elements are ultimately included in the common sequences of the library of templates derived from the adaptors to, for example, provide binding sites for particular sets of universal amplification primers and/or sequencing primers. Additional sequence elements may be included, for example, to provide binding sites for sequencing primers which will ultimately be used in sequencing of template molecules in the library, or products derived from amplification of the template library, for example on a solid support.

- 25 -

Although the precise nucleotide sequence of the adaptor is generally non-limiting to the invention, the sequences of the individual strands in the unmatched region should be such that neither individual strand exhibits any internal self-complementarity which could lead to self-annealing, formation of hairpin structures, etc. under standard annealing conditions. Self-annealing of a strand in the unmatched region is to be avoided as it may prevent or reduce specific binding of an amplification primer to this strand.

The mismatched adaptors are preferably formed from two strands of DNA, but may include mixtures of natural and non-natural nucleotides (e.g. one or more ribonucleotides) linked by a mixture of phosphodiester and non-phosphodiester backbone linkages. Other non-nucleotide modifications may be included such as, for example, biotin moieties, blocking groups and capture moieties for attachment to a solid surface, as discussed in further detail below.

The adaptor constructs may contain exonuclease resistant modifications such as phosphorothioate linkages. Such modifications reduce the number of adaptor-dimers present in the library, since the two adaptors can not undergo ligation without removal of their non complementary overhangs. The adaptors can be treated with an exonuclease enzyme prior to the ligation reaction with the target, to ensure that the overhanging ends of the strands can not be removed during the ligation process. Treatment of the adaptors in this manner reduces the formation of the adaptor-dimers at the ligation step.

Ligation methods are known in the art and utilise standard methods (Sambrook and Russell, Molecular Cloning, A Laboratory Manual, third edition). Such methods utilise

- 26 -

ligase enzymes such as DNA ligase to effect or catalyse joining of the ends of the two polynucleotide strands of, in this case, the adaptor duplex construct and the target polynucleotide duplexes, such that covalent linkages are
5 formed. The adaptor duplex construct may contain a 5'-phosphate moiety in order to facilitate ligation to the target 3'-OH. The target contains a 5'-phosphate moiety, either residual from the shearing process, or added using an enzymatic treatment step, and has been end repaired, and
10 optionally extended by an overhanging base or bases, to give a 3'-OH suitable for ligation. In this context, joining means covalent linkage of polynucleotide strands which were not previously covalently linked. In a particular aspect of the invention, such joining takes place by formation of a
15 phosphodiester linkage between the two polynucleotide strands, but other means of covalent linkage (e.g. non-phosphodiester backbone linkages) may be used.

Optionally the combined ligated polynucleotide sequences and unligated adaptor polynucleotide constructs
20 may be purified from any components of the ligation reaction, such as enzymes, buffers, salts and the like. Suitable purification methods are known in the art and utilise standard methods (Sambrook and Russell, Molecular Cloning, A Laboratory Manual, third edition).

25 In a next step according to the invention an amplification reaction is prepared. The contents of an amplification reaction are known by one skilled in the art and include appropriate substrates (such as dNTPs), enzymes (e.g. a DNA polymerase) and buffer components required for
30 an amplification reaction. Generally amplification reactions require at least two amplification primers, often denoted 'forward' and 'reverse' primers (primer oligonucleotides)

- 27 -

that are capable of annealing specifically to a part of the polynucleotide sequence to be amplified under conditions encountered in the primer annealing step of each cycle of an amplification reaction. In certain embodiments the forward and reverse primers may be identical. Thus the primer oligonucleotides must include an 'adaptor-target specific portion', being a sequence of nucleotides capable of annealing to a part of, that is, a primer-binding sequence, in the polynucleotide molecule to be amplified (or the complement thereof if the template is viewed as a single strand) during the annealing step.

Depending on the embodiment of the invention, the amplification primers may be universal for all samples, or one of the forward or reverse primers may carry the tag sequence that codes for the sample source. The amplification primers may hybridise across the tag region of the ligated adaptor, in which case unique primers will be needed for each sample nucleic acid. The amplification reaction may be performed with more than two amplification primers. In order to prevent the amplification of ligated adapter-adapter dimers, the amplification primers can be modified to contain nucleotides that hybridise across the whole of the ligated adapter and into the ligated template (or the dNTP's attached to the 3' end thereof). This first amplification primer can be modified and treated to help prevent exonucleases digestion of the strands, and thus it may be advantageous to have a first amplification primer that is universal and can amplify all samples rather than modifying and treating each of the tagged primers separately. The tagged primer can be introduced as a sample specific third primer in the amplification reaction, but does not need to be specially modified and treated to lower exonuclease

- 28 -

digestion. The nested PCR approach is shown in figure 6. In the case of this embodiment the third amplification primer that carries the tag needs to contain at sequence the same as at least a portion of the first amplification primer such
5 that it can be used to amplify the duplex resulting from extension of the first amplification primer.

In the context of the present invention, the term 'polynucleotide molecule to be amplified' refers to the original or starting adaptor-target-adaptor sequence added
10 to the amplification reaction. The 'adaptor-target specific portion' in the forward and reverse amplification primers refers to a sequence capable of annealing to the original or initial adaptor-target-adaptor present at the start of the amplification reaction and reference to the length of the
15 'adaptor-target specific portion' relates to the length of the sequence in the primer which anneals to the starting adaptor-target. It will be appreciated that if the primers contain any nucleotide sequence which does not anneal to the starting adaptor-target in the first amplification cycle
20 then this sequence may be copied into the amplification products (assuming the primer does not contain a moiety which prevents read-through of the polymerase). Hence the amplified template strands produced in the first and subsequent cycles of amplification may be longer than the
25 starting adaptor-target strands.

As the mis-matched adapters can be different lengths, the length of adapter sequence added to the 3' and 5' ends of each strand may be different. The amplification primers may also be of different length to each other, and may
30 hybridise to different lengths of the adapter, and therefore the length added to the ends of each strand can be controlled. In the case of nested PCR, the three or more

- 29 -

amplification primers can be designed to be longer than the primer used to amplify the previous amplicon, so the length of the added nucleotides is fully controllable and may be hundreds of base pairs if desired. In the example shown in figure 6b, the first amplification primer adds 13 bases to the ligated adapter, and the third amplification primer adds a further 27 bases such that one end of the amplicon is 40 bases longer than the short arm of adapter-target construct. The short arm of the adapter is 20 bases in length, meaning that the prepared template comprises the genomic region plus 60 added bases at the end. In figure 6b, the second amplification primer is 25 bases longer than the long arm of adapter, which is 32 bases in length plus the additional T that hybridise across the dATP nucleoside added to the genomic sample. Thus the prepared template comprises the genomic fragment, plus the added dATP, plus 57 known bases. Thus in full, one strand of each template duplex comprises from the 5' end: 60 known bases, T, the genomic fragment, A, 57 known bases-3' end. This strand is fully complementary to a sequence: 5'-57 known bases, T, genomic fragment, A, 60 known bases-3' end. The length 57 and 6 are arbitrary, and shown for the purpose of clarification, and should not be taken as limiting. The length of the added sequences may be 20-100 bases or more depending on the desired experimental design.

The forward and reverse primers may be of sufficient length to hybridise to the whole of the adaptor sequence and at least one base of the target sequence (or the nucleotide dNTP added as a 3'-overhang on the target strands). The forward and reverse primers may also contain a region that extends beyond the adaptor construct, and therefore the amplification primers may be at least 20-100 bases in

- 30 -

length. The forward and reverse primers may be of significantly different lengths; for example one may be 20-40 bases, whereas the other one may be 40-100 bases in length. The nucleotide sequences of the adaptor-target specific portions of the forward and reverse primers are selected to achieve specific hybridisation to the adaptor-target sequences to be amplified under the conditions of the annealing steps of the amplification reaction, whilst minimising non-specific hybridisation to any other target sequences present.

Skilled readers will appreciate that it is not strictly required for the adaptor-target specific portion to be 100% complementary, a satisfactory level of specific annealing can be achieved with less than perfectly complementary sequences. In particular, one or two mis-matches in the adaptor-target specific portion can usually be tolerated without adversely affecting specificity for the template. Therefore the term 'adaptor-target specific portion' should not be interpreted as requiring 100% complementarity with the adaptor-target. However, the requirement that the primers do not anneal non-specifically to regions of the adaptor-target other than their respective primer-binding sequences must be fulfilled.

Amplification primers are generally single stranded polynucleotide structures. They may also contain a mixture of natural and non-natural bases and also natural and non-natural backbone linkages, provided that any non-natural modifications do not preclude function as a primer - that being defined as the ability to anneal to a template polynucleotide strand during conditions of the amplification reaction and to act as an initiation point for synthesis of

- 31 -

a new polynucleotide strand complementary to the template strand.

Primers may additionally comprise non-nucleotide chemical modifications, for example phosphorothioates to increase exonuclease resistance, again provided such that modifications do not prevent primer function. Modifications may, for example, facilitate attachment of the primer to a solid support, for example a biotin moiety. Certain modifications may themselves improve the function of the molecule as a primer, or may provide some other useful functionality, such as providing a site for cleavage to enable the primer (or an extended polynucleotide strand derived therefrom) to be cleaved from a solid support.

In an embodiment wherein tags are attached to the adaptors, the amplification can be carried out on either the pooled or unpooled samples. In an embodiment wherein universal adaptors are used, tags must be part of the amplification primers, and therefore, each sample must be amplified independently prior to pooling. The pooled nucleic acid samples can then be processed for sequencing

Preparation of immobilised samples for sequencing

The pooled sample can then be immobilised in preparation for sequencing. Sequencing can be performed as an array of single molecules, or can be amplified prior to sequencing. The amplification can be carried out using one or more immobilised primers. The immobilised primer(s) can be a lawn on a planar surface, or on a pool of beads. The pool of beads can be isolated into an emulsion with a single bead in each 'compartment' of the emulsion. At a concentration of only one template per 'compartment', only a single template is amplified on each bead.

- 32 -

The term 'solid-phase amplification' as used herein refers to any nucleic acid amplification reaction carried out on or in association with a solid support such that all or a portion of the amplified products are immobilised on the solid support as they are formed. In particular, the term encompasses solid-phase polymerase chain reaction (solid-phase PCR) and solid phase isothermal amplification which are reactions analogous to standard solution phase amplification, except that one or both of the forward and reverse amplification primers is/are immobilised on the solid support. Solid phase PCR covers systems such as emulsions, wherein one primer is anchored to a bead and the other is in free solution, and colony formation in solid phase gel matrices wherein one primer is anchored to the surface, and one is in free solution.

Although the invention encompasses 'solid-phase' amplification methods in which only one amplification primer is immobilised (the other primer usually being present in free solution), it is preferred for the solid support to be provided with both the forward and the reverse primers immobilised. In practice, there will be a 'plurality' of identical forward primers and/or a 'plurality' of identical reverse primers immobilised on the solid support, since the amplification process requires an excess of primers to sustain amplification. References herein to forward and reverse primers are to be interpreted accordingly as encompassing a 'plurality' of such primers unless the context indicates otherwise.

As will be appreciated by the skilled reader, any given amplification reaction requires at least one type of forward primer and at least one type of reverse primer specific for the template to be amplified. However, in certain

- 33 -

embodiments the forward and reverse primers may comprise template-specific portions of identical sequence, and may have entirely identical nucleotide sequence and structure (including any non-nucleotide modifications). In other words, it is possible to carry out solid-phase amplification using only one type of primer, and such single-primer methods are encompassed within the scope of the invention. Other embodiments may use forward and reverse primers which contain identical template-specific sequences but which differ in some other structural features. For example one type of primer may contain a non-nucleotide modification which is not present in the other.

In all embodiments of the invention, primers for solid-phase amplification are preferably immobilised by single point covalent attachment to the solid support at or near the 5' end of the primer, leaving the template-specific portion of the primer free to anneal to its cognate template and the 3' hydroxyl group free for primer extension. Any suitable covalent attachment means known in the art may be used for this purpose. The chosen attachment chemistry will depend on the nature of the solid support, and any derivatisation or functionalisation applied to it. The primer itself may include a moiety, which may be a non-nucleotide chemical modification, to facilitate attachment. In a particular embodiment, the primer may include a sulphur-containing nucleophile, such as phosphorothioate or thiophosphate, at the 5' end. In the case of solid-supported polyacrylamide hydrogels (as described below), this nucleophile will bind to a bromoacetamide group present in the hydrogel. A more particular means of attaching primers and templates to a solid support is via 5' phosphorothioate attachment to a hydrogel comprised of polymerised acrylamide

- 34 -

and *N*-(5-bromoacetamidylpentyl) acrylamide (BRAPA), as described fully in WO05065814.

Certain embodiments of the invention may make use of solid supports comprised of an inert substrate or matrix (e.g. glass slides, polymer beads, etc) which has been "functionalised", for example by application of a layer or coating of an intermediate material comprising reactive groups which permit covalent attachment to biomolecules, such as polynucleotides. Examples of such supports include, but are not limited to, polyacrylamide hydrogels supported on an inert substrate such as glass. In such embodiments, the biomolecules (e.g. polynucleotides) may be directly covalently attached to the intermediate material (e.g. the hydrogel), but the intermediate material may itself be non-covalently attached to the substrate or matrix (e.g. the glass substrate). The term "covalent attachment to a solid support" is to be interpreted accordingly as encompassing this type of arrangement.

The pooled samples may be amplified on beads wherein each bead contains a forward and reverse amplification primer. In a particular embodiment, the library of templates prepared according to the first, second or third aspects of the invention is used to prepare clustered arrays of nucleic acid colonies, analogous to those described in US20050100900, US7115400, WO 00/18957 and WO 98/44151, the contents of which are incorporated herein by reference in their entirety, by solid-phase amplification and more particularly solid phase isothermal amplification. The terms 'cluster' and 'colony' are used interchangeably herein to refer to a discrete site on a solid support comprised of a plurality of identical immobilised nucleic acid strands and a plurality of identical immobilised complementary nucleic

- 35 -

acid strands. The term 'clustered array' refers to an array formed from such clusters or colonies. In this context the term 'array' is not to be understood as requiring an ordered arrangement of clusters.

5 The term solid phase, or surface, is used to mean either a planar array wherein primers are attached to a flat surface, for example, glass, silica or plastic microscope slides or similar flow cell devices; beads, wherein either one or two primers are attached to the beads and the beads
10 are amplified; or an array of beads on a surface after the beads have been amplified.

Clustered arrays can be prepared using either a process of thermocycling, as described in WO9844151, or a process whereby the temperature is maintained as a constant, and the
15 cycles of extension and denaturing are performed using changes of reagents. Such isothermal amplification methods are described in patent application numbers WO0246456 and US20080009420 (Isothermal methods for creating clonal single molecule arrays), which are incorporated herein by reference
20 in their entirety. Due to the lower temperatures required in the isothermal process, this is particularly preferred.

Use in sequencing/methods of sequencing

The invention also encompasses methods of sequencing
25 amplified nucleic acids generated by solid-phase amplification. Thus, the invention provides a method of nucleic acid sequencing comprising amplifying a pool of nucleic acid templates using solid-phase amplification as described above and carrying out a nucleic acid sequencing
30 reaction to determine the sequence of the whole or a part of at least one amplified nucleic acid strand produced in the

- 36 -

solid-phase amplification reaction, as well as the sequence of the tag that codes for the identity of the sample.

Sequencing can be carried out using any suitable sequencing technique, wherein nucleotides are added successively to a free 3' hydroxyl group, resulting in synthesis of a polynucleotide chain in the 5' to 3' direction. The nature of the nucleotide added is preferably determined after each nucleotide addition. Sequencing techniques using sequencing by ligation, wherein not every contiguous base is sequenced, and techniques such as massively parallel signature sequencing (MPSS) where bases are removed from, rather than added to the strands on the surface are also within the scope of the invention, as are techniques using detection of pyrophosphate release (pyrosequencing). Such pyrosequencing based techniques are particularly applicable to sequencing arrays of beads where the beads have been amplified in an emulsion such that a single template from the library molecule is amplified on each bead.

The initiation point for the sequencing reaction may be provided by annealing of a sequencing primer to a product of the solid-phase amplification reaction. In this connection, one or both of the adaptors added during formation of the template library may include a nucleotide sequence which permits annealing of a sequencing primer to amplified products derived by whole genome or solid-phase amplification of the template library.

Depending on the embodiment of the invention, the tag sequence and fragment sequence may be determined in a single read from a single sequencing primer, or in multiple reads from two sequencing primers. In the case of two reads from two sequencing primers, the 'tag read' and the 'fragment

- 37 -

read' can be performed in either order, with a suitable
denaturing step to remove the annealed primer after the
first sequencing read is completed. Suitable denaturing
steps may include formamide, hydroxide or heat as described
5 elsewhere.

The products of solid-phase amplification reactions
wherein both forward and reverse amplification primers are
covalently immobilised on the solid surface are so-called
'bridged' structures formed by annealing of pairs of
10 immobilised polynucleotide strands and immobilised
complementary strands, both strands being attached to the
solid support at the 5' end. Arrays comprised of such
bridged structures provide inefficient templates for nucleic
acid sequencing, since hybridisation of a conventional
15 sequencing primer to one of the immobilised strands is not
favoured compared to annealing of this strand to its
immobilised complementary strand under standard conditions
for hybridisation.

In order to provide more suitable templates for nucleic
20 acid sequencing, it is preferred to remove substantially all
or remove or displace at least a portion of one of the
immobilised strands in the 'bridged' structure in order to
generate a template which is at least partially single-
stranded. The portion of the template which is single-
25 stranded will thus be available for hybridisation to a
sequencing primer. The process of removing all or a portion
of one immobilised strand in a 'bridged' double-stranded
nucleic acid structure may be referred to herein as
'linearisation', and is described in further detail in
30 WO07010251, the contents of which are incorporated herein by
reference in their entirety.

- 38 -

Bridged template structures may be linearised by cleavage of one or both strands with a restriction endonuclease or by cleavage of one strand with a nicking endonuclease. Other methods of cleavage can be used as an
5 alternative to restriction enzymes or nicking enzymes, including *inter alia* chemical cleavage (e.g. cleavage of a diol linkage with periodate), cleavage of abasic sites by cleavage with endonuclease (for example 'USER', as supplied by NEB, part number M5505S), or by exposure to heat or
10 alkali, cleavage of ribonucleotides incorporated into amplification products otherwise comprised of deoxyribonucleotides, photochemical cleavage or cleavage of a peptide linker.

It will be appreciated that a linearization step may
15 not be essential if the solid-phase amplification reaction is performed with only one primer covalently immobilised and the other in free solution.

Following the cleavage step, regardless of the method used for cleavage, the product of the cleavage reaction may
20 be subjected to denaturing conditions in order to remove the portion(s) of the cleaved strand(s) that are not attached to the solid support. Suitable denaturing conditions, for example sodium hydroxide solution, formamide solution or heat, will be apparent to the skilled reader with reference
25 to standard molecular biology protocols (Sambrook et al., 2001, *Molecular Cloning, A Laboratory Manual*, 3rd Ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor Laboratory Press, NY; Current Protocols, eds Ausubel et al.). Denaturation results in the production of a sequencing
30 template which is partially or substantially single-stranded. A sequencing reaction may then be initiated by hybridisation of a sequencing primer to the single-stranded portion of the template.

- 39 -

Thus, the invention encompasses methods wherein the nucleic acid sequencing reaction comprises hybridising a sequencing primer to a single-stranded region of a linearised amplification product, sequentially incorporating
5 one or more nucleotides into a polynucleotide strand complementary to the region of amplified template strand to be sequenced, identifying the base present in one or more of the incorporated nucleotide(s) and thereby determining the sequence of a region of the template strand.

10 One preferred sequencing method which can be used in accordance with the invention relies on the use of modified nucleotides having removable 3' blocks, for example as described in WO04018497 and US7057026, the contents of which are incorporated herein by reference in their entirety. Once
15 the modified nucleotide has been incorporated into the growing polynucleotide chain complementary to the region of the template being sequenced there is no free 3'-OH group available to direct further sequence extension and therefore the polymerase can not add further nucleotides. Once the
20 nature of the base incorporated into the growing chain has been determined, the 3' block may be removed to allow addition of the next successive nucleotide. By ordering the products derived using these modified nucleotides, it is possible to deduce the DNA sequence of the DNA template.
25 Such reactions can be done in a single experiment if each of the modified nucleotides has a different label attached thereto, known to correspond to the particular base, to facilitate discrimination between the bases added during each incorporation step. Alternatively, a separate reaction
30 may be carried out containing each of the modified nucleotides separately.

- 40 -

The modified nucleotides may carry a label to facilitate their detection. A fluorescent label, for example, may be used for detection of modified nucleotides. Each nucleotide type may thus carry a different fluorescent label, for example, as described in US Provisional Application No. 60/801,270 (Novel dyes and the use of their labelled conjugates), published as WO07135368, the contents of which are incorporated herein by reference in their entirety. The detectable label need not, however, be a fluorescent label. Any label can be used which allows the detection of an incorporated nucleotide.

One method for detecting fluorescently labelled nucleotides comprises using laser light of a wavelength specific for the labelled nucleotides, or the use of other suitable sources of illumination. The fluorescence from the label on the nucleotide may be detected by a CCD camera or other suitable detection means. Suitable instrumentation for recording images of clustered arrays is described in US Provisional Application No. 60/788,248 (Systems and devices for sequence by synthesis analysis), published as WO07123744, the contents of which are incorporated herein by reference in their entirety.

The invention is not intended to be limited to use of the sequencing method outlined above, as essentially any sequencing methodology which relies on successive incorporation of nucleotides into a polynucleotide chain can be used. Suitable alternative techniques include, for example, Pyrosequencing™, FISSEQ (fluorescent in situ sequencing), MPSS and sequencing by ligation-based methods, for example as described in US6306597.

The nucleic acid sample may be further analysed to obtain a second read from the opposite end of the fragment.

- 41 -

Methodology for sequencing both ends of a cluster are described in co-pending applications WO07010252 and PCTGB2007/003798, the contents of which are incorporated by reference herein in their entirety. In one example, the series of steps may be performed as follows; generate clusters, linearise, hybridise first sequencing primer and obtain first sequencing read. The first sequencing primer can be removed, a second primer hybridised and the tag sequenced. The nucleic acid strand may then be 'inverted' on the surface by synthesising a complementary copy from the remaining immobilised primers used in cluster amplification. This process of strand resynthesis regenerates the double stranded cluster. The original template strand can be removed, to linearise the resynthesised strand that can then be annealed to a sequencing primer and sequenced in a third sequencing run.

In the cases where strand resynthesis is employed, both strands must be immobilised to the surface in a way that allows subsequent release of a portion of the immobilised strand. This can be achieved through a number of mechanisms as described in WO07010251, the contents of which are incorporated herein by reference in their entirety. For example, one primer can contain a uracil nucleotide, which means that the strand can be cleaved at the uracil base using the enzymes uracil glycosylase (UDG) which removes the nucleoside base, and endonuclease VIII that excises the abasic nucleotide. This enzyme combination is available as USER™ from New England Biolabs (NEB part number M5505). The second primer may comprise an 8-oxoguanine nucleotide, which is then cleavable by the enzyme FPG (NEB part number M0240). This design of primers give complete control of which primer is cleaved at which point in the process, and also where in the cluster the cleavage occurs. The primers may also be chemically modified, for example with a disulfide or diol

- 42 -

modification that allows chemical cleave at specific locations.

5

Kits

The invention also relates to kits for use in preparing libraries of tagged polynucleotides using the method of the first, second or third aspects of the invention.

10 Embodiments of the kit comprise at least a supply of a mismatched adaptor as defined herein, plus a supply of at least one amplification primer which is capable of annealing to the mismatched adaptor and priming synthesis of an extension product, which extension product would include any
15 target sequence ligated to the adaptor when the adaptor is in use.

Particular features of the 'mismatch' adaptors for inclusion in the kit are as described elsewhere herein in relation to other aspects of the invention. The structure
20 and properties of amplification primers are well known to those skilled in the art. Suitable primers of appropriate nucleotide sequence for use with the adaptors included in the kit can be readily prepared using standard automated nucleic acid synthesis equipment and reagents in routine use
25 in the art. The kit may include a supply of one single type of primer or separate supplies (or even a mixture) of two different primers, for example a pair of amplification primers suitable for PCR or isothermal amplification of templates modified with the mismatched adaptor in solution
30 phase and/or on a suitable solid support (i.e. solid-phase amplification). The kit may comprise a mismatch double stranded adapter for ligation to a sample of interest, plus

- 43 -

at least two different amplification primers that carry a different tag sequence, where the tag sequence does not hybridise to the adapter. This kit can be used to amplify at least two different samples where each sample is amplified
5 using a single tagged primer, and then pooled after the individual amplification reactions.

In one embodiment the kit may include supplies of different primer-pairs for use in solution phase and solid phase PCR and more particularly isothermal amplification. In
10 this context the 'different' primer-pairs may be of substantially identical nucleotide sequence but differ with respect to some other feature or modification, such as for example surface-capture moieties, etc. In other embodiments the kit may include a supply of primers for use in an
15 initial primer extension reaction and a different primer-pair (or pairs) for solution and/or solid phase amplification.

Adaptors and/or primers may be supplied in the kits ready for use, or more preferably as concentrates requiring
20 dilution before use, or even in a lyophilised or dried form requiring reconstitution prior to use. If required, the kits may further include a supply of a suitable diluent for dilution or reconstitution of the primers. Optionally, the kits may further comprise supplies of reagents, buffers,
25 enzymes, dNTPs, etc., for use in carrying out PCR or isothermal amplification. Suitable (but non-limiting) examples of such reagents are as described in the Materials and Methods sections of the accompanying Examples. Further components which may optionally be supplied in the kit
30 include 'universal' sequencing primers suitable for sequencing templates prepared using the mismatched adaptors and primers.

- 44 -

The invention will be further understood with reference to the following non-limiting experimental examples:

5 **Examples**

The following are examples of general techniques which may be applied in carrying out the method of the invention.

10 **Example 1: Preparation and conventional sequencing of tagged libraries:**

The following experimental details describe the complete exposition of one embodiment of the invention as described above.

Two libraries were made, one using purified human BAC
15 DNA (140k human chromosome 6 insert cloned into a pTARBAC vector) and the other purified PhiX174 RF1 DNA (NEB, N3021L). Each library was prepared separately. The DNA was first prepared for ligation to forked adaptors by:
fragmentation of the DNA by nebulisation, end repair of the
20 DNA ends to render them blunt-ended and phosphorylated, then the addition of a single 'A' nucleotide onto the 3' ends of the DNA fragments. The ligation reaction was performed with the prepared fragmented DNA and adaptors pre-formed by annealing 'Oligo A' and 'Oligo B' (sequences given below).
25 The same adaptors were used for each library. The product of the reaction was isolated/purified from unligated adaptor by gel electrophoresis. Finally, the product of the ligation reaction was subjected to cycles of PCR to selectively amplify ligated product that contained genomic DNA with
30 adaptor at both ends of the fragments. During the PCR a unique tag was added to each library using a unique PCR

- 45 -

primer, so that the BAC library was indexed with the unique sequence tag 'ATC' and the PhiX library was indexed with the unique sequence tag 'CGA'. See Figure 1 for general schematic of protocol.

5

Materials and Methods

Step 1) Nebulization

Materials:

- 0.5 ug/ul Human BAC DNA (140k human chromosome 6 insert
10 cloned into a pTARBAC vector)
- 1ug/ul PhiX174 RF1 DNA (NEB, N3021L)
- Nebulization Buffer (53.1 ml glycerol, 42.1 ml water,
3.7 ml 1 M Tris HCl pH7.5, 1.1 ml 0.5 M EDTA)
- TE
- 15 • Nebulizers (Invitrogen, K7025-05)
- PCR purification kit columns (Qiagen, 28104)

Procedure:

Mixed 5 µl (5 µg) of PhiX174 DNA with 45 µl of TE and
20 700 µl of nebulization buffer. Mixed 10 µl (5 µg) of BAC
DNA with 40 µl of TE and 700 µl of nebulization buffer.
Chilled DNA solutions were each fragmented in a nebulizer on
ice for 6 minutes under 32 pounds per square inch (psi) of
pressure. The recovered volumes were each purified with a
25 Qiagen PCR purification kit column and eluted in 30 µl of
EB.

Step 2) End-repair

Materials:

- 30 • Nebulized DNA (from Step 1)
- Water

- 46 -

- T4 DNA ligase buffer with 10mM ATP (10x) (NEB, B0202S)
- dNTPs mix (10mM each) (NEB, N0447S)
- T4 DNA Polymerase (3U/ul) (NEB, M0203L)
- *E. coli* DNA Pol I large fragment (Klenow) (5U/ul) (NEB, M0210S)
- T4 polynucleotide kinase (10U/ul) (NEB, M0201L)
- PCR purification kit columns (Qiagen, 28104)

Procedure:

10 End repair mix was assembled as follows:

	Nebulized DNA	30ul
	Water	45ul
	T4 DNA ligase buffer with 10mM ATP	10ul
15	dNTPs	4ul
	T4 DNA pol	5ul
	Klenow DNA pol	1ul
	T4 PNK	5ul
		100ul total

20

The reaction was incubated for 30 minutes at room temperature. The DNA was purified on a Qiagen column, eluting in 30ul EB.

25 Step 3) A- tailing Reaction

Materials:

- End repaired DNA (from Step 2)
- Water
- NEB buffer 2 (10x) (NEB, B7002S)
- 30 • dATP (1mM) (Amersham-Pharmacia, 272050)
- Klenow fragment (3' to 5' exo minus) (5U/ul) (NEB, M0212B)

- 47 -

- * Hot block or PCR machine
- * MinElute PCR purification kit column (Qiagen, 28004)

Procedure:

5 The following reaction mix was assembled:

End repaired DNA	30µl
Water	2µl
NEB [®] buffer 2	5µl
dATP	10µl
10 Klenow fragment (3' to 5' exc minus)	3µl
	50µl total

The reaction was incubated for 30 minutes at 37°C, then the DNA purified on a Qiagen MinElute column, eluting in 10µl

15 EB.

Step 4) Annealed Adaptors

Materials:

- * Oligo A: 5'ACACTCTTTCCCTACACGACGCTCTTCCGATCXT (x =
20 phosphorothioate bond) (SEQ ID No. 1)
- * Oligo B: 5'Phosphate-GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
(SEQ ID No. 2)
- * 50mM Tris/50mM NaCl pH7.0
- * PCR machine

25

Procedure:

The oligos were mixed together to a final concentration of 15µM each, in 10mM Tris/10mM NaCl pH 7.0. The adaptor strands were annealed in a PCR machine programmed as
30 follows: Ramp at 0.5 °C/sec to 97.5 °C; Hold at 97.5 °C for 150 sec; then a step of 97.5 °C for 2 sec with a temperature drop of 0.1 °C/cycle for 775 cycles.

- 48 -

In order to ensure a sequencing primer is attached to both ends of the adapter for the paired end method, the sequence of oligonucleotide B was changed to:

5 5' Phosphate-GATCGGAAGAGCACACGTCT 3' (SEQ ID No. 3)

Step 5) Ligation

Materials:

- 10 • A-tailed genomic DNA (from Step 3)
- Quick ligase buffer (2x) (NEB, B2200S)
- Annealed adaptor (15uM) (from 4.)
- Quick Ligase (1U/ul) (NEB, M2200L)
- PCR purification kit columns (Qiagen, 28104)

15

Procedure:

Reaction mix was assembled as follows:

A-tailed genomic DNA	10ul
Quick ligase buffer	25ul
20 Annealed adaptor	10ul
Quick Ligase	5ul
	50ul total

The reaction was incubated for 15 minutes at room
25 temperature, then the DNA purified on a Qiagen column,
eluting in 30ul EB.

Step 6) Gel purification

Materials:

- 30 • Ligation reaction (from Step 5)
- Agarose (Biorad, 161-3107)
- TAE (50x)

- 49 -

- Distilled water
- Ethidium bromide (Sigma, E1510)
- Loading buffer (4x) (50 mM Tris pH8, 40 mM EDTA, 40% w/v sucrose)
- 5 • Low molecular weight ladder (NEB, N3233L)
- Gel trays and tank. Electrophoresis unit
- Dark reader transilluminator (Clare Chemical Research, D195M)
- Gel extraction kit columns (Qiagen, 28704)

10

Procedure:

The entire sample from the purified ligation reaction was loaded into one lane of a 2% agarose gel containing ethidium bromide and run at 120V for 60 minutes. The gel was
15 then viewed on a 'White-light' box and fragments from 120bp to 170bp excised and purified with a gel extraction column, eluting in 30µl elution buffer (EB).

Step 7) Exonuclease I Treatment of PCR Primers

20 Materials:

- Exonuclease I (E. coli) (20 U/ul) (NEB, M0293S)
- Exonuclease I Reaction Buffer (10x) (NEB, M0293S)
- Water
- DNA Primers with a phosphorothioate at the n-1 position
- 25 • P6 Bio-Rad columns (Bio-Rad, 732-6221)

Procedure:

DNA Primers with a phosphorothioate at the n-1 position (5 x 85 µl of each Primer (approx 25 µM)) were aliquoted
30 into Eppendorf tubes. 10 µl of 10 X Exonuclease I Reaction Buffer and 5 µl of Exonuclease I was added to each tube.

- 50 -

Each Eppendorf tube was placed in a rack and stored in an oven set at 37 °C for 16 hours. After 16 hr, the tubes were placed on a hotblock set at 80 °C for 2 minutes.

Subsequently, the solutions from the Eppendorfs were passed
5 through P6 Bio-Rad columns and spun in a centrifuge at 2000 rpm for 2 minutes. An extra 20 µl of H₂O was added to the columns and the columns respun. The filtered solutions were placed into a SpeedVac® and evaporated until each was at 20 µl, and the fractions combined. The pooled fractions were
10 injected into a reverse phase HPLC system, and the main peak was collected. The collected fractions were evaporated to dryness in a SpeedVac®, 50 µl of water was added and the fraction was subjected again to evaporation to dryness. The resulting pellets were dissolved in 50 µl of water, pooled
15 and a UV measurement taken to determine the concentration of the oligonucleotide.

Step 8) PCR

Materials:

- * Gel purified DNA (from Step 6)
- 20 * Water
- * Phusion master mix (2x) (NEB, F-531L)
- * Exonuclease treated universal PCR primer (25µM): 5'
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGA
TCxT (SEQ ID No. 4) 3', where x= phosphorothioate bond
25 (from Step 7)
- * Exonuclease treated indexing PCR primer 1 (25µM): 5'
CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTC
CGATCxT (SEQ ID No. 5), where x= phosphorothioate bond
(from Step 7)
- 30 * Exonuclease treated indexing PCR primer 2 (25µM): 5'
CAAGCAGAAGACGGCATACGATCGCGGTCTCGGCATTCCTGCTGAACCGCTCTTC

- 51 -

CGATC×T (SEQ ID No. 6) , where x= phosphorothioate bond
(from Step 7)

- PCR machine
- PCR purification kit columns (Qiagen, #28104)

5

Procedure:

The PCR reaction was prepared as follows:

BAC Library

	Gel purified BAC DNA	1µl
10	Phusion mastermix	25µl
	Universal PCR primer	1µl
	Indexing PCR primer 1	1µl
	Water	<u>22µl</u>
		50µl total

15

PhiX174 Library

	Gel purified PhiX174 DNA	1µl
	Phusion mastermix	25µl
	Universal PCR primer	1µl
20	Indexing PCR primer 2	1µl
	Water	<u>22µl</u>
		50µl total

Thermocycling was carried out in a PCR machine under
the following conditions:

- 25
- 30 secs @ 98°C
 - [10 sec @ 98°C, 30 sec @ 65°C, 30 sec @ 72°C] 18 cycles
 - 5 minutes @ 72°C
 - Hold @ 4°C

30 PCR products were purified on a Qiagen column, eluting
in 30µl EB. The resulting DNA libraries were ready for

- 52 -

mixing and amplification on a surface amplification platform.

In the case of paired end runs, the human and Phi-X sample
5 were split into 6 portions, and each portion was amplified
with the two universal primers plus one of the 12 primers
shown below.

Universal LPX primer 5'
10 GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCXT (SEQ ID No. 7) 3'
Lpx primer 5'
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCXT
(SEQ ID No. 8)
Tag1 primer (CGTGAT) 5'
15 CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTC (SEQ ID No. 9)
3'
Tag2 primer (ACATCG) 5'
CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTC (SEQ ID No. 10)
3'
20 Tag3 primer (GCCTAA) 5'
CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTC (SEQ ID No. 11)
3'
Tag4 primer (TGGTCA) 5'
CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTC (SEQ ID No. 12)
25 3'
Tag5 primer (CACTGT) 5'
CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTC (SEQ ID No. 13)
3'
Tag6 primer (ATTGGC) 5'
30 CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTC (SEQ ID No. 14)
3'

- 53 -

- Tag7 primer (GATCTG) 5'
CAAGCAGAAGACGGCATACGAGATGATCTGGTGAAGTGGAGTTC (SEQ ID No. 15)
3'
- Tag8 primer (TCAAGT) 5'
5 CAAGCAGAAGACGGCATACGAGATTCAAGTGTGAAGTGGAGTTC (SEQ ID No. 16)
3'
- Tag9 primer (CTGATC) 5'
CAAGCAGAAGACGGCATACGAGATCTGATCGTGAAGTGGAGTTC (SEQ ID No. 17)
3'
- 10 Tag10 primer (AAGCTA) 5'
CAAGCAGAAGACGGCATACGAGATAAGCTAGTGAAGTGGAGTTC (SEQ ID No. 18)
3'
- Tag11 primer (GTAGCC) 5'
CAAGCAGAAGACGGCATACGAGATGTAGCCGTGAAGTGGAGTTC (SEQ ID No. 19)
15 3'
- Tag12 primer (TACAAG) 5'
CAAGCAGAAGACGGCATACGAGATTACAAGGTGAAGTGGAGTTC (SEQ ID No. 20)
3'
- 20 The two universal primers contain a phosphorothioate group
(x) and were exonucleases treated and purified prior to
amplification. The amplification conditions were as
described above, and the 6 tagged human samples, and 6
tagged Phi-X sample were pooled together for sequencing on
25 clusters.

Validation of Libraries by Conventional Sanger Sequencing

- Four (4) µl of the libraries were cloned into a plasmid
vector (Zero Blunt TOPO PCR cloning kit, Invitrogen #K2800-
30 20) and plated out on agar, according to the manufacturer's
instructions. Colonies were picked, mini-prepped and the

- 54 -

cloned fragments sequenced by conventional Sanger sequencing.

16 clones from BAC Library (tag)

- 5 1 (204bp) Insert: E.coli 85bp
AATGATACGGCGACCACCGAGATCTACACTCTTTCCGTACACGACGGCTCTTC
CGATCTACTGATTTTCATTGCAGCCAAAGGCAAACCTTTGGCTGCATCGTTTAA
CAGTCGCCATAAGCCTTTCCTCTGTAAACCGCCTTCTGAGATCGGAAGAGCG
GGTTCAGCAGGAATGCCGAGACCGTCGTATGCCGTCTTCTGCTTC
10 (SEQ ID No. 21)
- 2 (214bp) Insert: BAC 95bp
AATGATACGGCGACCACCGAGATCTACACTCTTTCCGTACACGACGGCTCTTC
CGATCTATCAATATTGTGAAAATGACCATACTGCCAAAAAACTACAAATT
15 CAATGCAATTTTCATCAAAATACCATCATCATTCTTCACAATATTGATAGATC
GGAAGAGCGGTTCAGCAGGAATGCCGAGACCGTCGTATGCCGTCTTCTGCTTC
(SEQ ID No. 22)
- 3 (215bp) Insert: BAC 96bp
20 AATGATACGGCGACCACCGAGATCTACACTCTTTCCGTACACGACGGCTCTTC
CGATCTCTCACTCCTGGCAGAGGGACGTGTGACTAGCCATGGGCCCTAG
GTCTCCAGTTCCTGGGTAGCTTGTATTTTGAACATCTCCTGTATATTAGTTA
GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGTCGTATGCCGTCTTCTGCTTC
(SEQ ID No. 23)
- 25 4 (147bp) Insert: BAC 28bp
AATGATACGGCGACCACCGAGATCTACACTCTTTCCGTACACGACGGCTCTTC
CGATCTAGTGTAGTTGAGATCTGCCTTAGCAGCAAGATCGGAAGAGCGGTTC
CAGCAGGAATGCCGAGACCGTCGTATGCCGTCTTCTGCTTC (SEQ ID
30 No. 24)

- 55 -

5 (183bp) Insert: BAC 64bp

AATGATACGGCGACCACCGAGATCTACACTCCTTCCCTACACGACGCTCTTCC
CGATCTAACACATTTCAAAGTTTGGGGCCCTCCTCCTCCCCAAAAACAAAC
CACAAAAACAAACAAAAAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
5 ACCGATTCGTATGCCGTCTTCTGCTTG (SEQ ID No. 25)

6 (170bp) Insert: BAC 59bp

GGCGACCACCGAGATCTACACTCCTTCCCTACACGACGCTCTTCCGATCTG
AATGCCTTTTATAGCATTTAATTTTCCTAAGTATAATTACCAAATAAAAATTG
10 TATAAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATTCGTATG
CCGTCTTCTGCTTG (SEQ ID No. 26)

7 (180bp) Insert: BAC 61bp

AATGATACGGCGACCACCGAGATCTACACTCCTTCCCTACACGACGCTCTTCC
15 CGATCTTGGGCCCCGGGAGGAGTTTGCCGGGGAGGAGTGGGTTTGAATCG
GGGTAAAGGAAAGAGAAAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
ACCGATTCGTATGCCGTCTTCTGCTTG (SEQ ID No. 27)

8 (190bp) Insert: BAC 73bp

20 TGATACGGCGACCACCGAGATCTACACTCCTTCCCTACACGACGCTCTTCC
GATCTAAGATCTATTTCAAATGGACTGTAGATCTAAGTATAAAAGGTAAGAGA
ATAATTATTCTAGAAAGTAAATGTAAGATCGGAAGAGCGGTTCAGCAGGAAT
GCCGAGACCGATTCGTATGCCGTCTTCTGCTTG (SEQ ID No. 28)

25 9 (192bp) Insert: BAC 74bp

AATGATACGGCGACCACCGAGATCTACACTCCTTCCCTACACGACGCTCTTCC
GATCTGGGAGGCCAAGGTGGGTGGATCACCTGAGATCAGGAGTTCGAGAC
CAGCTGCCCAACATGATGAACTCTGTCTAGATCGGAAGAGCGGTTCAGCA
GGAATGCCGAGACCGATTCGTATGCCGTCTTCTGCTTG (SEQ ID No.
30 29)

- 56 -

10 (185bp) Insert: BAC 66bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTTGACCATTGTAACCATTAAATGTAGACTGCAATGATATGCACTATTTA
CAACCTTTTTTAAGACTCTAGATCGGAAGAGCGGTTCAGCAGGAATGCCGA
5 GACCG TCGTATGCCGTCTTCTGCTTG (SEQ ID No. 30)

11 (199bp) Insert: BAC 80bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTCTTTGAAGAGCTGGCAGTAGAAGATAAACAGGCTGGGGAAGAAGA
10 GAAAGTGCTCAAGGAGAAGGAGCAGCAGCAGCAGC
AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCG TCGTATGCCGT
CTTCTGCTTG (SEQ ID No. 31)

12 (212bp) Insert: BAC 93bp

15 AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTAGTATTCAACAAGTCTGTCTTTTCCAAGTGTCTTTAAAGACCAGAAA
TACCTGTTTTTAACACACAGGGTTGCAAAATTCAGAGGAGATTGGCAGATCG
GAAGAGCGGTTCAGCAGGAATGCCGAGACCG TCGTATGCCGTCTTCTG
CTTG (SEQ ID No. 32)

20

13 (247bp) Insert: E.coli 128bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTCTTGAGATGAGTGATGACGGCGCGCTGGAAGTTGCTCGTCGCGCT
CGCGGTACGCCGCGCATTGCCAACCGTCTGCTGCGTCGAGTGGTGATTTC
25 GCCGAAGTGAAGCACGATGGCACCATCTCAAGAGATCGGAAGAGCGGTTC
GCAGGAATGCCGAGACCG TCGTATGCCGTCTTCTGCTTG (SEQ ID No.
33)

14 (202bp) Insert: BAC 83bp

30 AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTGGGGTTGGTGAACCCAGATGCCTCCAGGATTGGTGGGCCCTG

- 57 -

TGGCACTTGTACCTGCTGTTGCTGTTGCTGCTGCTGCTGAGATCGGAAGAG
CGGTTGAGCAGGAATGCCGAGACCGATCGTATGCCGTCTTCTGCTTG (SEQ
ID No. 34)

5 15 (166bp) Insert: BAC 47bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTATGATAAGGAGCAGGTTTACAGATCATAAGTGCAAAAGCGGGCGA
GAAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCGTATGCC
GTCTTCTGCTTG (SEQ ID No. 35)

10

16 (147bp) Insert: BAC 31bp

GATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG
ATCTCTGATACTGTTGTAACCACCCAATTGGTTCAAGATCGGAAGAGCGGT
CAGCAGGAATGCCGAGACCGATCGTATGCCGTCTTCTGCTTG (SEQ ID

15 No. 36)

16 clones from PhiX174 Library (GAT tag)

1 (183bp) Insert: PhiX 67bp

20 AATGATACGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGA
TCTTAAAGCTACCACTTATATGGCTGTTGGTTTCTATGTGGCTAAATACGTTA
ACAAAAAGTCAGATATGAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGA
CCGCGATCGTATGCCGTCTTCTGCTTG (SEQ ID No. 37)

25 2 (174bp) Insert: PhiX 58bp

GATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG
ATCTCGACCACTCGCGATTCAATCATGACTTCGTGATAAAAGATTGAGTGTG
AGGTTATAACAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGCGAT
CGTATGCCGTCTTCTGCTTG (SEQ ID No. 38)

30

3 (179bp) Insert: PhiX 62bp

- 58 -

TGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC
GATCTTCTATGAAGGATGTTTTCCGTTCTGGTGATTCTGCTAAGAAGTTTAAAG
ATTGCTGAGGGTCAAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCG
GCGATCGTATGCGCGTCTTCTGCTTG (SEQ ID No. 39)

5

4 (169bp) Insert: PhiX 50bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC
CGATCTGGTTGCGCCGCCAAAACGTCGGCTACAGTAACTTTTCCCAGCCTC
AATCTAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGCGCGTATG
CGCGTCTTCTGCTTG (SEQ ID No. 40)

10

5 (166bp) Insert: PhiX 50bp

TGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGA
TCTAACAACCTGAACGGAAGTGGAAACACTGGTCATAATCATGGTGCGGAATAA
GAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGCGCGTATGCG
GCTCTCTGCTTG (SEQ ID No. 41)

15

6 (164bp) Insert: PhiX 45bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC
CGATCTAAATACGTTAACAAAAAGTCAGATATGGACCTTGCTGCTAAAGGTA
GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGCGCGTATGCGCGT
CTTCTGCTTG (SEQ ID No. 42)

20

7 (194bp) Insert: PhiX 75bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC
CGATCTGTCAATATAACCAGTAGTGTTAACAGTCGGGAGAGGAGTGCCATTA
ACACCATCCTTCATGAACCTAATCCACTGAGATCGGAAGAGCGGTTCAGCAG
GAATGCCGAGACCGCGCGTATGCGCGTCTTCTGCTTG (SEQ ID No. 43)

25

30 8 (198bp) Insert: PhiX 79bp

- 59 -

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTGATAGTTTGACGGTTAATGCTGGTAATGGTGGTTTTCTTCATTGCAT
TCAGATGGATACATCTGTCAACGCCGCTAATCAGATCGGAAGAGCGGTTC
GCAGGAATGCCGAGACCGCGATCGTATGCCGTCTTCTGCTTG (SEQ ID

5 No. 44)

9 (179bp) Insert: PhiX 60bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTTGAAACCAACATAAACATTATTGCCCGGCGTACGAGGAAGGACGT
10 CAATAGTCACACAGTAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACG
GCGATCGTATGCCGTCTTCTGCTTG (SEQ ID No. 45)

10 (198bp) Insert: PhiX 79bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
15 CGATCTCCAGCTTCTTCGGCACCTGTTTTACAGACACCTAAAGCTACATCGT
CAACGTTATATTTTGATAGTTTGACGGTTAATAAGATCGGAAGAGCGGTTC
GCAGGAATGCCGAGACCGCGATCGTATGCCGTCTTCTGCTTG (SEQ ID

No. 46)

20 11 (151bp) Insert: PhiX 32bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTCAGTTTTGCCGCAAGCTGGCTGCTGAACGCCCAGATCGGAAGAGC
GGTTCAGCAGGAATGCCGAGACCGCGATCGTATGCCGTCTTCTGCTTG (SE
Q ID No. 47)

25 12 (173bp) Insert: PhiX 54bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTTTTGCTGATGAACTAAGTCAACCTCAGCACTAACCTTGCGAGTCAT
TTCTTTGAAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGCGATCG
30 TATGCCGTCTTCTGCTTG (SEQ ID No. 48)

- 60 -

13 (151bp) Insert: PhiX 32bp

AATGATACGGCGACCACCGAGATCTACACTGTTTCCCTACACGAGCGCTCTTC
 CGATCTAATAGACGCAACGCGAGCAGTAGACTCCTTCTAGATCGGAAGAGCG
 GGTTCAGCAGGAATGCCGAGACCGCGATCGTATGCCGTCTTCTGCTTG (SE

5 Q ID No. 49)

14 (168bp) Insert: PhiX 49bp

AATGATACGGCGACCACCGAGATCTACACTGTTTCCCTACACGAGCGCTCTTC
 CGATCTGGAGGCCTCCAGCAATCTTGAACACTCATCCTTAATACCTTTCTTTT
 10 TGAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGCGATCGTATGCC
 GTCTTCTGCTTG (SEQ ID No. 50)

15 (181bp) Insert: PhiX 62bp

AATGATACGGCGACCACCGAGATCTACACTGTTTCCCTACACGAGCGCTCTTC
 15 CGATCTACTCCTAAGCAGAAAACCTACCGCGCTTCGCTTGGTCAACCCCTCA
 GCGGCAAAAATTAAAAAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGAC
 CGCGATCGTATGCCGTCTTCTGCTTG (SEQ ID No. 51)

16 (182bp) Insert: E.coli 63bp

20 AATGATACGGCGACCACCGAGATCTACACTGTTTCCCTACACGAGCGCTCTTC
 CGATCTGAGCCGACATCGAGGTGCCAAACACCGCCGTCGATATGAACTCTT
 GGC CGGTATCAGCCTGTTAGATCGGAAGAGCGGTTCAGCAGGAATGCCGA
 GACCGCGATCGTATGCCGTCTTCTGCTTG (SEQ ID No. 52)

25 These results confirm that the library preparation
 method produces 2 libraries of 'sequenceable' DNA templates.
 Each library contained a plurality of genomic inserts, each
 of which was flanked by the two adaptors

(AATGATACGGCGACCACCGAGATCTACACTGTTTCCCTACACGAGCGCTCTTCGATCT

30 (SEQ ID No. 53) and

AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGNNNNCGATGCGGTCTTCTGCTT

(SEQ ID No. 54)), required for cluster formation and SBS

- 61 -

sequencing. The insert DNA from each of the libraries aligned to the correct reference, either BAC or PhiX, with a small amount of E.coli contamination in each library. The correct TAG (■■■ for PhiX or ■■■ for BAC) was present in
5 each library.

Example of Cluster preparation using the mixed PHiX and BAC libraries

Example 2): Acrylamide coating of glass chips

10 The solid supports used are typically 8-channel glass chips such as those provided by Silex Microsystems (Sweden). However, the experimental conditions and procedures are readily applicable to other solid supports.

Chips were washed as follows: neat Decon for 30
15 minutes, milliQ H₂O for 30 minutes, NaOH 1N for 15 minutes, milliQ H₂O for 30 minutes, HCl 0.1N for 15 minutes, milliQ H₂O for 30 minutes.

Polymer solution preparation:

20 For 10 ml of 2% polymerisation mix.

- 10 ml of 2% solution of acrylamide in milliQ H₂O
- 165 µl of a 100mg/ml N-(5-bromoacetamidylpentyl) acrylamide (BRAPA) solution in DMF (23.5 mg in 235µl DMF)
- 25 - 11.5 µl of TEMED
- 100 µl of a 50 mg/ml solution of potassium persulfate in milliQ H₂O (20mg in 400µl H₂O)

The 10 ml solution of acrylamide was first degassed
30 with argon for 15 minutes. The solutions of BRAPA, TEMED and potassium persulfate were successively added to the

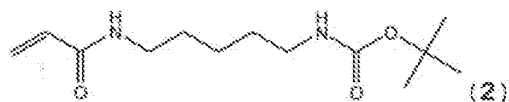
- 62 -

acrylamide solution. The mixture was then quickly vortexed and immediately used. Polymerization was then carried out for 1h 30 minutes at RT. Afterwards the channels were washed with milliQ H₂O for 30 minutes and filled with 0.1 M
5 potassium phosphate buffer for storage until required.

Example 3) Synthesis of N-(5-bromoacetamidylpentyl) acrylamide (BRAPA) (1)



N-Boc-1,5-diaminopentane toluene sulfonic acid was obtained from Novabiochem. The bromoacetyl chloride and acryloyl chloride were obtained from Fluka. All other reagents were
15 Aldrich products.



To a stirred suspension of N-Boc-1,5-diaminopentane
20 toluene sulfonic acid (5.2 g, 13.88 mmol) and triethylamine (4.83 ml, 2.5 eq) in THF (120 ml) at 0°C was added acryloyl chloride (1.13 ml, 1 eq) through a pressure equalized dropping funnel over a one hour period. The reaction mixture was then stirred at room temperature and the progress of the
25 reaction checked by TLC (petroleum ether:ethyl acetate 1:1). After two hours, the salts formed during the reaction were filtered off and the filtrate evaporated to dryness. The residue was purified by flash chromatography (neat petroleum ether followed by a gradient of ethyl acetate up to 60%) to

- 63 -

yield 2.56 g (9.98 mmol, 71 %) of product 2 as a beige solid. ^1H NMR (400 MHz, d_6 -DMSO) : 1.20-1.22 (m, 2H, CH_2), 1.29-1.43 (m, 13H, tBu, $2\times\text{CH}_2$), 2.86 (q, 2H, $J = 6.8$ Hz and 12.9 Hz, CH_2), 3.07 (q, 2H, $J = 6.8$ Hz and 12.9 Hz, CH_2), 5.53 (dd, 1H, $J = 2.3$ Hz and 10.1 Hz, CH), 6.05 (dd, 1H, $J = 2.3$ Hz and 17.2 Hz, CH), 6.20 (dd, 1H, $J = 10.1$ Hz and 17.2 Hz, CH), 6.77 (t, 1H, $J = 5.3$ Hz, NH), 8.04 (bs, 1H, NH). Mass (electrospray+) calculated for $\text{C}_{13}\text{H}_{24}\text{N}_2\text{O}_3$ 256, found 279 (256+ Na^+).



Product 2 (2.56g, 10 mmol) was dissolved in trifluoroacetic acid:dichloromethane (1:9, 100 ml) and stirred at room temperature. The progress of the reaction was monitored by TLC (dichloromethane:methanol 9:1). On completion, the reaction mixture was evaporated to dryness, the residue co-evaporated three times with toluene and then purified by flash chromatography (neat dichloromethane followed by a gradient of methanol up to 20%). Product 3 was obtained as a white powder (2.43 g, 9 mmol, 90%). ^1H NMR (400 MHz, D_2O): 1.29-1.40 (m, 2H, CH_2), 1.52 (quint., 2H, $J = 7.1$ Hz, CH_2), 1.61 (quint., 2H, $J = 7.7$ Hz, CH_2), 2.92 (t, 2H, $J = 7.6$ Hz, CH_2), 3.21 (t, 2H, $J = 6.8$ Hz, CH_2), 5.68 (dd, 1H, $J = 1.5$ Hz and 10.1 Hz, CH), 6.10 (dd, 1H, $J = 1.5$ Hz and 17.2 Hz, CH), 6.20 (dd, 1H, $J = 10.1$ Hz and 17.2 Hz, CH). Mass (electrospray+) calculated for $\text{C}_8\text{H}_{15}\text{N}_2\text{O}$ 156, found 179 (156+ Na^+).

To a suspension of product 3 (6.12 g, 22.64 mmol) and triethylamine (6.94 ml, 2.2 eq) in THF (120 ml) was added

- 64 -

bromoacetyl chloride (2.07 ml, 1.1eq), through a pressure equalized dropping funnel, over a one hour period and at -60°C (cardice and isopropanol bath in a Dewar). The reaction mixture was then stirred at room temperature overnight and the completion of the reaction was checked by TLC (dichloromethane:methanol 9:1) the following day. The salts formed during the reaction were filtered off and the reaction mixture evaporated to dryness. The residue was purified by chromatography (neat dichloromethane followed by a gradient of methanol up to 5%). 3.2 g (11.55 mmol, 51 %) of the product **1** (BRAPA) were obtained as a white powder. A further recrystallization performed in petroleum ether:ethyl acetate gave 3g of the product **1**. ¹H NMR (400 MHz, d₆-DMSO) : 1.21-1.30 (m, 2H, CH₂), 1.34-1.48 (m, 4H, 2xCH₂), 3.02-3.12 (m, 4H, 2xCH₂), 3.81 (s, 2H, CH₂), 5.56 (d, 1H, J = 9.85 Hz, CH), 6.07 (d, 1H, J = 16.9 Hz, CH), 6.20 (dd, 1H, J = 10.1 Hz and 16.9 Hz, CH), 8.07 (bs, 1H, NH), 8.27 (bs, 1H, NH). Mass (electrospray+) calculated for C₁₀H₁₇BrN₂O₂ 276 or 278, found 279 (278+H⁺), 299 (276+Na⁺).

Example 4) Grafting primers onto surface of SFA coated chip

An SFA coated chip is placed onto a modified MJ-Research thermocycler and attached to a peristaltic pump. Grafting mix consisting of 0.5 μM of a forward primer and 0.5 μM of a reverse primer in 10 mM phosphate buffer (pH 7.0) is pumped into the channels of the chip at a flow rate of 60 μl/minutes for 75 s at 20 °C. The thermocycler is then heated to 51.6 °C, and the chip is incubated at this temperature for 1 hour. During this time, the grafting mix undergoes 18 cycles of pumping: grafting mix is pumped in at 15 μl/minutes for 20 s, then the solution is pumped back and

- 65 -

forth (5 s forward at 15 µl/minutes, then 5 s backward at 15 µl/minutes) for 180 s. After 18 cycles of pumping, the chip is washed by pumping in 5xSSC/5mM EDTA at 15 µl/minutes for 300 s at 51.6 °C. The thermocycler is then cooled to 20 °C.

5 The primers are typically 5'-phosphorothioate oligonucleotides incorporating any specific sequences or modifications required for cleavage. Their sequences and suppliers vary according to the experiment for which they are to be used, and in this case were complementary to the
10 5'-ends of the template duplex. The DNA sequence used in this process was the pool of the two libraries, which have ends complementary to the grafted primers. The library mix was denatured using sodium hydroxide treatment followed by snap dilution as described.

15 For some of the experiments detailed, the amplified clusters contained a diol linkage in one of the grafted primers. Diol linkages can be introduced by including a suitable linkage into one of the primers used for solid-phase amplification. Synthesis of the diol phosphoramidite
20 is described in Example 4 below. Products containing such diol linkages can be cleaved using periodate and propanolamine as described, and the resulting single stranded polynucleotides hybridised as described.

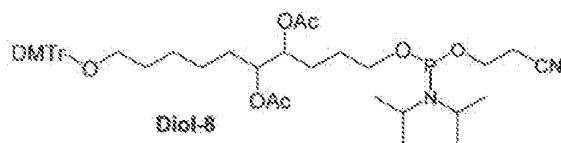
 The grafted primers contain a sequence of T bases at
25 the 5'-end to act as a spacer group to aid linearisation and hybridization. The sequences of the two primers grafted to the chip are as follows:

30 P5diol: 5' PS-TTTTTTTTTT-diol-AATGATACGGCGACCACCGA (SEQ ID No. 55)
P7: 5' PS-TTTTTTTTTTCAAGCAGAAGACGGCATACGA (SEQ ID No. 56)

- 66 -

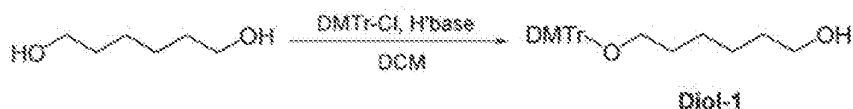
Example 5: Preparation of diol-phosphoramidite for DNA coupling

5



Step 1:

10



1,6-Hexanediol (Sigma Aldrich 99 %) (14.6 g, 124 mmol), *N,N*-diisopropylethylamine (Hünig's base; Sigma Aldrich; redistilled) (21.6 mL, 124 mmol) is dissolved in anhydrous DCM/ DMF (250/ 50 mL) under N₂. The solution is cooled to 0°C and the first portion of 4, 4'-dimethoxytrityl chloride (DMTr-Cl; Sigma-Aldrich 95 %) (10.5 g, 31 mmol) is added.

20 The reaction mixture is then warmed to room temperature. After stirring for 1 h, the reaction mixture is cooled to 0°C again and the second portion of DMTr-Cl (10.5 g, 31 mmol) is added and then allowed to stir at room temperature for another 2 hours. TLC (EtOAc: petroleum ether 4:6)

25 analysis indicates ca. 95% consumption of starting material derivative (DMTr-OH). The reaction is concentrated under reduced pressure and Aq. NaHCO₃ (sat.) solution (500 mL) is poured into the residue. The resulting mixture is extracted with petroleum ether/ EtOAc (2:1) (3 x 1000 mL). The

- 67 -

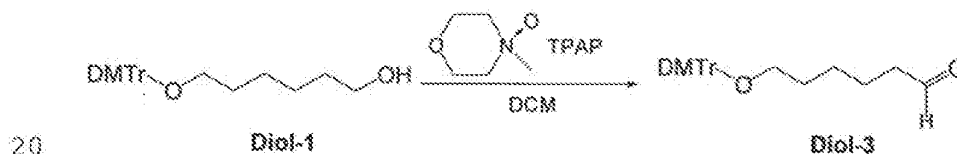
combined organic layers are dried over MgSO_4 , and concentrated under vacuum. The residue is co-evaporated with xylene (2 x 100 mL) to remove DMF. The reaction mixture, is pre-absorbed on silica gel and subjected to

5 flash chromatography using solvents containing 1 % Et_3N petroleum ether to petroleum ether/ EtOAc (7:3) as eluent. The yield of pale yellow oil is 16.58 g, 64 %, with a further 7.8g (17%) of the bis-tritylated by-product.

10 TLC: R_f : 0.35 (diol-1); R_f : 0.7 (bis-tritylated by-product) (petroleum ether/ EtOAc 6:4).

^1H NMR (400 MHz, CDCl_3): δ 1.32-1.44 (m, 4H, 2x CH_2), 1.54-1.68 (m, 4H, 2x CH_2), 3.06 (t, J = 6.6 Hz, 2H, CH_2O), 3.62-3.68 (m, 2H, CH_2OH), 3.81 (s, 6H, 2x MeO), 6.83-6.85 (m, 4H, Ph), 7.24-7.35 (m, 7H, Ph), 7.45-7.47 (m, 2H, Ph).

15

Step 2:

To a solution of Diol-1 (16.6 g, 39.5 mmol) in anhydrous DCM (200 mL), tetrapropylammonium perruthenate (TPAP; Sigma Aldrich 97 %) (277 mg, 0.79 mmol) is added

25 under N_2 atmosphere. The solution is cooled to 0 °C and *N*-methylmorpholine *N*-oxide (Sigma Aldrich 97 %) (2.7 g, 23 mmol) is added. The reaction is warmed to room temperature. After 1 hour, the other three portions of *N*-methylmorpholine *N*-oxide (3 x 2.0 g, 51.2 mmol) are added within a period of

- 68 -

four hours. TLC (EtOAc: petroleum ether 4:6) indicates the reaction goes to completion. The reaction is quenched with aq. NaHCO₃ (sat.) (1000 mL) and extracted to CH₂Cl₂ (4 x 1000 mL). The combined organic layers are dried over MgSO₄. The solution is concentrated under reduced pressure. Diol-3, 9.9 g, 60 %, is isolated by flash chromatography using solvents containing 1% Et₃N from petroleum ether to petroleum ether/ EtOAc (6:4) as eluent, as a pale yellow oil.

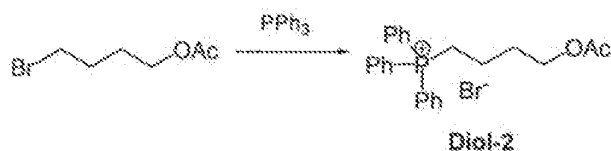
10

TLC: R_f: 0.7 (petroleum ether/ EtOAc 6:4).

¹H NMR (400MHz, CDCl₃): δ 1.30-1.37 (m, 2H, CH₂), 1.48-1.57 (m, 4H, 2x CH₂), 2.34 (td, J = 1.7 and 7.4 Hz, 2H, CH₂CHO), 2.97 (s, 2H, CH₂O), 3.72 (s, 6H, 2x MeO), 6.73-6.76 (m, 4H, Ph), 7.10-7.26 (m, 7H, Ph), 7.34-7.36 (m, 2H, Ph), 9.67 (t, J = 1.7, 1H, CHO).

20

Step 3:



A solution of triphenylphosphine (Sigma-Aldrich 99 %, ReagentPlus™), (39.3 g, 150 mmol) and 4-bromobutyl acetate (Sigma-Aldrich) (26 mL, 180 mmol) in anhydrous toluene (300 mL) is heated under reflux for 36 hours under N₂ in an oil-bath (140 °C). During the reflux, oil is precipitated out. The reaction mixture is cooled to room temperature. TLC (petroleum ether/ EtOAc 7:3) analysis of the toluene

25

- 69 -

solution showed that there is still triphenylphosphine (R_f : 0.8) left. The supernatant is decanted into another round-bottomed flask and concentrated down to the approximate volume of 30 mL. The solution is heated under reflux again
5 for another 12 hours. The supernatant is decanted. The portions of oil are combined together, dissolved in water (500 mL) and extracted with EtOAc (2 x 500 mL). The combined organic layers are back-extracted with water (150 mL). Two
10 lots of aqueous layers are combined, and evaporated under reduced pressure. The resulting residue is co-evaporated with acetonitrile (2 x 100 mL) to give 78.4 g, 95 % yield of a pale yellow oil. NMR indicates that the product is pure, and is used for the next step reaction without further purification.

15

TLC: R_f : 0.0 (petroleum ether/ EtOAc 7:3).

^1H NMR (400 MHz, CDCl_3): δ 1.63-1.73 (m, 2H, CH_2), 1.94 (s, 3H, 2x CH_3), 2.06-2.16 (m, 2H, CH_2), 3.97-4.05 (m, 2H, CH_2P),
20 4.11 (t, $J = 6.0$, 2H, CH_2O), 7.69-7.95 (m, 15H, Ph).

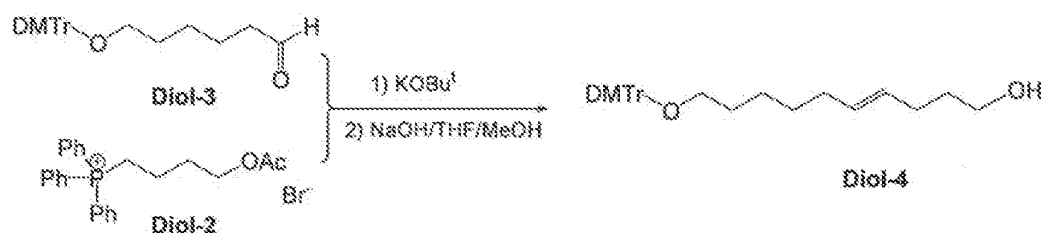
^{31}P NMR (162 MHz, CDCl_3): 25.9 ppm.

Mass spec details: LC-MS (Electrospray positive): (M^+) 377.

25

Step 4:

- 70 -



Diol-2 (10.34 g, 22.7 mmol) is weighed into a round-bottomed flask and dissolved with DCM (20 mL). The solution is then evaporated under reduced pressure until it gives a white foam. The flask is then subjected to high vacuum for 24 h. To this flask, anhydrous THF (180 mL) is added under N₂. The resulting suspension is cooled to -78 °C with an acetone-dry ice bath. With vigorous stirring, KOBu^t (3.3 g, 29.5 mmol) is added under N₂. Slowly the colour of the suspension turns orange, and white solids are gradually precipitated out. To this suspension, a solution of diol-3 (dried at 60 °C under high vacuum for 1 h before the reaction), (9.5 g, 22.7 mmol) in THF (50 mL) is added drop wise over half an hour. Acetone-dry ice bath is then removed. The reaction mixture is slowly warmed to room temperature and stirred for another hour. The colour of the reaction mixture turns yellow after the addition of diol-3. The reaction mixture is concentrated down under reduced pressure. The resulting residue is partitioned between DCM (800 mL) and aq. NaCl (sat.) (800 mL). The aqueous layer is extracted with an additional DCM (2 x 800 mL). The organic extractions are combined, dried over MgSO₄, filtered, and evaporated under reduced pressure to give yellow oil. The oil is dissolved in THF/MeOH (125/100 mL) and cooled to 0°C. To this solution, NaOH (1M in H₂O, 25 mL) is added. After allowing the reaction to stir for 1 hour, TLC analysis

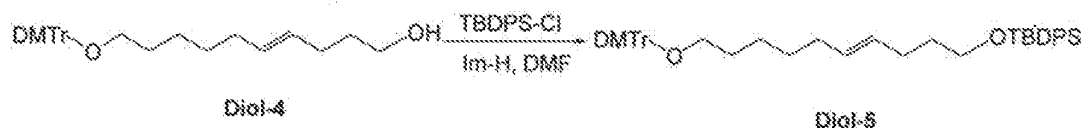
- 71 -

indicates full consumption of starting material. The reaction mixture is neutralized with acetic acid (1.5 mL). The reaction mixture is concentrated down under reduced pressure. The resulting residue is partitioned between DCM (800 mL) and aq. NaHCO₃ (sat.) (800 mL). The aqueous layer is extracted with additional DCM (2 x 800 mL). The organic extractions are combined, dried over MgSO₄, filtered, and evaporated to give a pale yellow oil. Diol-4, 6.45 g, 60 % is isolated by flash chromatography using solvents containing 1% Et₃N from petroleum ether to petroleum ether/EtOAc (5:5) as eluent, as a light yellow oil.

TLC: R_f = 0.45 (petroleum ether/ EtOAc 6:4).

¹H NMR (400 MHz, CDCl₃) δ 1.24-1.32 (m, 4H, 2x CH₂), 1.54-1.57 (m, 4H, 2x CH₂), 1.93-1.96 (m, 2H, CH₂), 2.02-2.07 (m, 2H, CH₂), 2.96 (t, J = 6.6 Hz, 2H, CH₂O), 3.54-3.59 (m, 2H, CH₂OH), 3.72 (s, 6H, 2x MeO), 5.29-5.32 (m, 2H, 2x =CH), 6.73-6.77 (m, 4H, Ph), 7.11-7.27 (m, 7H, Ph), 7.36-7.38 (m, 2H, Ph).

Step 5:



25

To a solution of Diol-4 (5.68 g, 12 mmol) and imidazole (Sigma Aldrich, 99 %), (1.63 g, 24 mmol) in anhydrous DMF (100 mL), *t*-butyldiphenylsilyl chloride (Sigma Aldrich, 98

30

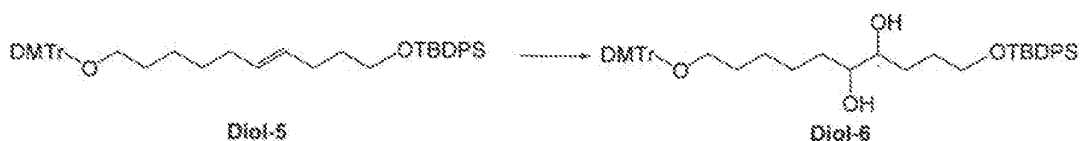
- 72 -

5 %), (4.05 mL, 15.6 mmol) is added drop wise under N₂ atmosphere at room temperature. The reaction is stirred for 1 hour. TLC (petroleum ether/ EtOAc 8:2) indicates that the starting material is fully consumed. A saturated aq. NaHCO₃ solution (500 mL) is added to quench the reaction. The resulting mixture is extracted with petroleum ether/EtOAc (2:1) (3 x 500 mL). The organic layers are combined, dried over MgSO₄, filtered, and evaporated to give a yellow oil. Diol-5, 8.14 g, 95 % is isolated by flash chromatography using solvents containing 1% Et₃N from petroleum ether to petroleum ether/ EtOAc (9:1) as eluent, as a colourless oil.

TLC: R_f = 0.7 (petroleum ether : EtOAc 8:2).

15 ¹H NMR (400 MHz, CDCl₃): δ 0.97 (s, 9H, 3xMe), 1.19-1.30 (m, 4H, 2x CH₂), 1.48-1.55 (m, 4H, 2x CH₂), 1.91-1.95 (m, 2H, CH₂), 2.01-2.06 (m, 2H, CH₂), 2.95 (t, J = 6.6 Hz, 2H, CH₂O), 3.58 (t, J = 6.3 Hz, 2H, CH₂O), 3.70 (s, 6H, 2x MeO), 5.24-5.27 (m, 2H, 2x =CH), 6.72-6.75 (m, 4H, Ph), 7.11-7.37 (m, 15H, Ph), 7.57-7.60 (m, 4H, Ph).

Step 6:



A mixture of diol-5 (9.27 g, 13 mmol), AD-mix-α (Sigma Aldrich), (18.2 g), methanesulfonamide (Sigma Aldrich, 97 %), (1.23 g, 13 mmol), t-BuOH (65 mL) and water (65 mL) is stirred together vigorously at 55 °C for 14 h. The TLC

- 73 -

analysis (petroleum ether : EtOAc 6:4) indicates ca. 95 % consumption of the starting material. The reaction mixture is cooled to room temperature, treated with sodium sulfite (15.3 g, 12 mmol), then further stirred for 30 minutes. A saturated aq. NaHCO₃ solution (500 mL) is added to the reaction. The resulting mixture is extracted with EtOAc (3 x 500 mL). The organic layers are combined, dried over MgSO₄, filtered, and evaporated to give yellow oil. Diol-6, 7.96 g, 82 %, is isolated by flash chromatography (silica gel, Fluka, 70-230 mesh) using solvents containing 1% Et₃N from petroleum ether to petroleum ether/ EtOAc (1:1) as elutant, as a white solid.

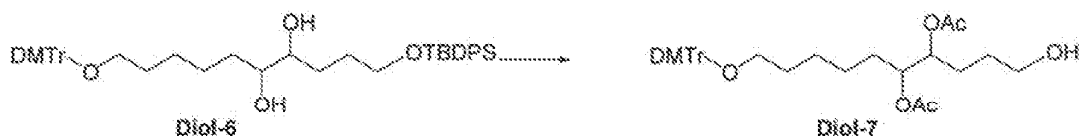
TLC: R_f = 0.3 (petroleum ether : EtOAc 6:4).

15

¹H NMR (400 MHz, CDCl₃) δ 1.07 (s, 9 H, 3xMe), 1.41-1.7 (m, 12 H, 6x CH₂), 1.94 (d, J = 4.3 Hz, 1H, OH), 2.94-2.95 (m, 1H, OH), 3.06 (t, J = 6.6 Hz, 2H, CH₂O), 3.61-3.63 (m, 2 H, 2x CHOH), 3.73 (t, J = 5.6 Hz, 2H, CH₂O), 3.81 (s, 6H, 2x MeO), 5.24-5.27 (m, 2H, 2x =CH), 6.82-6.85 (m, 4H, Ph), 7.21-7.47 (m, 15 H, Ph), 7.57-7.60 (m, 4 H, Ph).

20

Step 7:



25

To a solution of diol-6 (7.96 g, 13 mmol) and DMAP (Sigma-Aldrich ReagentPlus™, 99%) (260 mg, 2.13 mmol) in a

30

- 74 -

mixture of pyridine (15 mL) and DCM (30 mL), acetic anhydride (Fluka 99%), (2.5 mL, 26.68 mmol) is added at room temperature. TLC analysis (petroleum ether:EtOAc 6:4) indicates full consumption of the starting material after 1
5 h. The reaction is quenched by saturated aq. NaHCO₃ solution (500 mL). After 5 minutes, the mixture is extracted with DCM (3 x 500 mL). The organic layers are combined, dried over MgSO₄, filtered, and evaporated. The residue is co-evaporated with toluene (2 x 100 mL). The resulting yellow
10 oil is subjected to a plug of silica gel (50 g, Fluka, 70-230 mesh) to remove DMAP using eluents containing 0.1% Et₃N from petroleum ether to petroleum ether/ EtOAc (7:3) (250 mL each). The combined fractions of product are concentrated to dryness. The resulting colourless oil is dissolved in THF
15 (100 mL) and treated with TBAF (Sigma-Aldrich; 5% wt water), (1 M in THF, 15 mL) at 0 °C. The reaction solution is slowly warmed to room temperature and stirred for 2 hours. TLC analysis (petroleum ether:EtOAc 6:4) indicates that desilylation is completed. The volatile solvent (THF) is
20 evaporated under reduced pressure at low temperature. A saturated aq. NaHCO₃ solution (500 mL) is added to the residue. The resulting mixture is extracted with EtOAc (3 x 500 mL). The organic layers are combined, dried over MgSO₄, filtered, and evaporated to give yellow oil. Diol-7, 4.2 g,
25 66 %, is isolated by flash chromatography using solvents containing 1% Et₃N from petroleum ether to petroleum ether/ EtOAc (1:1) as eluent, as a white solid.

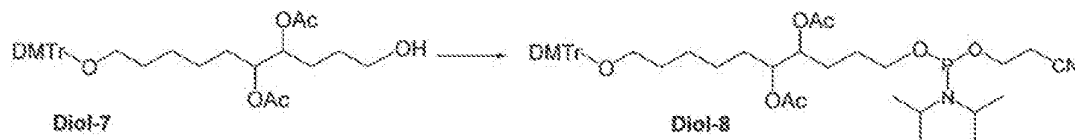
TLC: R_f = 0.45 (petroleum ether : EtOAc 1:1).

30

¹H NMR (400 MHz, CDCl₃) δ 1.29-1.33 (m, 4H, 2x CH₂), 1.47-1.63 (m, 8 H, 4x CH₂), 1.99, 2.01 (2s, 6H, 2 MeC(O)), 3.00

(t, $J = 6.5$ Hz, 2H, CH₂O), 3.60-3.64 (m, 2 H, CH₂O), 3.75 (s, 6H, 2x MeO), 4.92-4.97 (m, 2H, 2x CHOAc), 6.76-6.80 (m, 4H, Ph), 7.15-7.29 (m, 7 H, Ph), 7.38-7.40 (m, 2 H, Ph).

Step 3:



To a solution of diol-7 (2.08 g, 3.5 mmol) and diisopropylethylamine (Sigma Aldrich), (1.53 ml, 8.75 mmol) in DCM (17 mL), 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (1.0 g, 4.2 mmol) is added drop wise at room temperature under N_2 . After stirring for 1 hour, TLC analysis (petroleum ether:EtOAc 4:6) indicates the full consumption of the starting material. The solvent (THF) is concentrated under reduced pressure. The resulting residue is subjected to chromatography directly. Diol-8, 2.5 g, 90 %, is isolated by flash chromatography using solvents containing 1% Et_3N from petroleum ether to petroleum ether/EtOAc (1:1) as eluent, as a colourless syrup.

TLC: $R_f = 0.55$ (petroleum ether : EtOAc 4:6).

25 ¹H NMR (400 MHz, CDCl₃) δ 1.09, 1.10, 1.11, 1.12 (4 s, 12 H, N(CHMe₂)₂), 1.26-1.31 (m, 4H, 2x CH₂), 1.45-1.56 (m, 8 H, 4x CH₂), 1.95, 1.969, 1.971, 1.98 (4 s, 6H, 2 MeCO), 2.56 (t, *J* = 6.5 Hz, 2H, CH₂CN), 2.95 (t, *J* = 6.5 Hz, 2H, CH₂O), 3.49-3.55 (m, 4 H, CH₂O), 3.72 (s, 6H, 2x MeO), 4.89-4.92 (m, 2H,

- 76 -

2x CH₃CO₂), 6.74-6.76 (m, 4H, Ph), 7.13-7.25 (m, 7 H, Ph),
7.34-7.37 (m, 2 H, Ph).

³¹P NMR (162 MHz, CDCl₃): 148.67, 148.69 ppm.

5

Example 6) Preparation of clusters by isothermal
amplification:

Step 1: Hybridisation and amplification

The DNA sequence used in the amplification process is
10 the mixture of the two libraries prepared in Example 1,
which have ends complementary to the grafted primers. The
duplex DNA (1 nM) is denatured using 0.1 M sodium hydroxide
treatment followed by snap dilution to the desired 0.2-2 pM
'working concentration' in 'hybridization buffer' (5 x SSC /
15 0.1% Tween).

Surface amplification was carried out by isothermal
amplification using an MJ Research thermocycler, coupled
with an 8-way peristaltic pump Ismatec IPC ISM931 equipped
with Ismatec tubing (orange/yellow, 0.51 mm ID). A schematic
20 of the instrument is shown in figure 10.

The single stranded template (denatured as indicated
above) is hybridised to the grafted primers immediately
prior to the amplification reaction, which thus begins with
25 an initial primer extension step rather than template
denaturation. The hybridization procedure begins with a
heating step in a stringent buffer to ensure complete
denaturation prior to hybridisation. After the
hybridization, which occurs during a 20 minutes slow cooling
30 step, the flowcell was washed for 5 minutes with a wash
buffer (0.3 x SSC / 0.1% Tween).

- 77 -

A typical amplification process is presented in Table 1, which details the flow volumes per channel.

Table 1. Template Hybridization and 1st Extension

Step	Description	T (°C)	Time (sec)	Flow rate (µl/min)	Pumped V (µl)
1	Pump Hybridization pre-mix	20	120	60	120
2	Pump Hybridization mix	98.5	300	15	75
3	Remove bubbles	98.5	10	100	16.7
4	Stop flow and hold T	98.5	30	static	0
5	Slow cooling	98.5-40.2	19.5 min	static	0
6	Pump wash buffer	40.2	300	15	75
7	Pump amplification pre-mix	40.2	200	15	50
8	Pump amplification mix	40.2	75	60	75
9	First Extension	74	90	static	0
10	cool to room temperature	20	0	static	0

Table 2. Isothermal Amplification

Step	Description	T (°C)	Time (sec)	Flow rate (µl/min)	Pumped V (µl)
(1) This sequence 35 times	Pump Formamide	60	75	60	75
	Pump Amplification pre-mix	60	75	60	75
	Pump Bst mix	60	95	60	95
	Stop flow and hold T	60	180	static	0
2	Pump wash buffer	60	120	60	120

Details are as follows: Hybridisation pre mix (buffer) = 5 x SSC / 0.1% Tween; Hybridisation mix = 0.1 M hydroxide DNA sample, diluted in hybridisation pre mix; Wash buffer = 0.3 x SSC / 0.1% Tween; Amplification pre mix = 2 M betaine, 20 mM Tris, 10 mM Ammonium Sulfate, 2 mM Magnesium sulfate, 0.1% Triton, 1.3% DMSO, pH 8.8; Amplification mix = 2 M

- 78 -

betaine, 20 mM Tris, 10 mM Ammonium Sulfate, 2 mM Magnesium sulfate, 0.1% Triton, 1.3% DMSO, pH 8.8 plus 200 μ M dNTP's and 25 units/mL of Taq polymerase (NEB Product ref M0273L);
 Bst mix = 2 M betaine, 20 mM Tris, 10 mM Ammonium Sulfate, 2
 5 mM Magnesium sulfate, 0.1% Triton, 1.3% DMSO, pH 8.8 plus
 200 μ M dNTP's and 80 units/mL of Bst polymerase (NEB Product ref M0275L)

Step 2: Blocking extendable 3'-OH groups

10 To prepare the blocking pre-mix, 1530 μ L of water and
 170 μ L of 10X blocking buffer (NEB buffer 4; product number B7004S) are mixed for a final volume of 1700 μ L. To prepare
 the blocking mix, 1065.13 μ L of blocking pre-mix, 21.12 μ L
 of 125 μ M ddNTP mix, and 13.75 μ L of TdT terminal
 15 transferase (NEB; part no M0252S) are mixed for a final
 volume of 1100 μ L.

To block the nucleic acid within the clusters formed in
 the flow cell channels, the computer component of the
 instrumentation flows the appropriate blocking buffer
 20 through the flow cell, and controls the temperature as shown
 in the exemplary embodiments below in Table 3.

Step	Description	T (°C)	Time (sec)	Flow rate (μ L/min)	Pumped V (μ L)
1	Pump Blocking pre-mix	20	200	15	50
2	Pump Blocking mix	37.7	300	15	75
3	Stop flow and hold T	37.7	20	static	0
4	Cyclic pump Blocking mix and	37.7	8 x (20+180)	15 / static	45

- 79 -

	wait				
5	Pump wash buffer	20	300	15	75

Example 7: Linearisation and hybridization of a sequencing primer

5 To prepare the linearization mix, 1429 μL of water, 64 mg of sodium periodate, 1500 μL of formamide, 60 μL of 1M Tris pH8, and 6011.4 μL of 3-aminopropanol are mixed for a final volume of 3 mL. The periodate is first mixed with the water while the Tris is mixed with the formamide. The two
10 solutions are then mixed together and the 3-aminopropanol is added to that mixture.

To linearize the nucleic acid within the clusters formed within the flow cell channels, 300 μL per channel of linearisation mix is flowed in at 15 $\mu\text{L}/\text{minutes}$ at 20 $^{\circ}\text{C}$;
15 followed by 75 μL of water at the same flow rate.

To prepare the primer mix, 895.5 μL of hybridization buffer and 4.5 μL of sequencing primer (100 μM) are mixed to a final volume of 900 μL . The sequence of the sequencing primer used in the first reaction was:

20 5'-ACACTCTTTCCCTACACGACGCTCTTCCGATC (SEQ ID No. 57)

To denature the nucleic acid within the clusters and to hybridize the sequencing primer, the following reagents are flowed through the cell as shown in Table 4:

Step	Description	T ($^{\circ}\text{C}$)	Time (sec)	Flow rate ($\mu\text{L}/\text{min}$)	Pumped V (μL)
1	Pump 0.1 M NaOH	20	300	15	75
2	Pump TE	20	300	15	75
3	Pump Primer mix	20	300	15	75

- 80 -

4	Hold at 60C	60	900	0	0
5	Pump wash buffer	40.2	300	15	75

Example 8: Sequencing from the target fragment

Sequencing of the clusters from the above illustrative
 5 protocol was carried out using modified nucleotides prepared
 as described in International patent application WO
 2004/018493, and labeled with four spectrally distinct
 fluorophores, as described in US application number
 60/801270; filed May 18th 2006. Sequencing of clusters is
 10 described in more detail in patent WO06064199. The contents
 of the above-listed three documents are incorporated herein
 by reference in their entireties.

A mutant 9thN polymerase enzyme (an exo- variant
 including the triple mutation L408Y/Y409A/P410V and C223S)
 15 was used for the nucleotide incorporation steps.

Incorporation mix, Incorporation buffer (50 mM Tris-HCl
 pH 8.0, 6 mM MgSO₄, 1 mM EDTA, 0.05% (v/v) Tween -20, 50mM
 NaCl) plus 110nM YAV exo- C223S, and 1μM each of the four
 labeled modified nucleotides, was applied to the clustered
 20 templates, and heated to 45°C.

Templates were maintained at 45°C for 30 minutes,
 cooled to 20°C and washed with Incorporation buffer, then
 with 5x SSC/0.05% Tween 20. Templates were then exposed to
 Imaging buffer (100mM Tris pH7.0, 30mM NaCl, 0.05% Tween 20,
 25 50mM sodium ascorbate, freshly dissolved). Templates were
 scanned in 4 colors at room temperature. Templates were
 then exposed to sequencing cycles of Cleavage and
 Incorporation as follows:

- 81 -

Cleavage: Prime with Cleavage buffer (0.1M Tris pH 7.4, 0.1M NaCl and 0.05% Tween 20), 125 μ L/channel; 60 μ L/minutes.

Heat to 60 °C.

- 5 Treat the clusters with Cleavage mix (100 mM TCEP in Cleavage buffer), 75 μ L/channel; 60 μ L/minutes.

Wait for a total of 15 minutes in addition to pumping fresh cleavage mix, 25 μ L/channel; 60 μ L/minutes, every 4 minutes.

Cool to 20 °C.

- 10 Wash with Enzymology buffer.

Wash with 5XSSC/0.05% Tween 20.

Prime with Imaging buffer.

Scan in 4 colors at RT.

- 15 Incorporation: Prime with Incorporation buffer, 125 μ L/channel; 60 μ L/minutes, Heat to 60°C

Treat with Incorporation mix, 75 μ L/channel; 60 μ L/minutes.

Wait for a total of 15 minutes in addition to pumping fresh Incorporation mix, 25 μ L/channel; 60 μ L/minutes, every 4

- 20 minutes.

Cool to 20 °C.

Wash with Incorporation buffer, 75 μ L/channel; 60 μ L/minutes.

Wash with 5XSSC/0.05% Tween 20, 75 μ L/channel; 60 μ L/minutes

- 25 Prime with imaging buffer, 100 μ L/channel; 60 μ L/minutes

- 82 -

Scan in 4 colors at RT.

Repeat the process of Incorporation and Cleavage for as many cycles as required.

- 5 Incorporated nucleotides were detected using a Total Internal Reflection based fluorescent CCD imaging apparatus described in "Systems and Devices for Sequence by Synthesis Analysis," USSN 60/788,248, filed March 31, 2006.

10 Example 9) Sequencing the tag

The first sequencing primer can be removed, and a second one applied using the same protocol as described in Example 7. The sequence of the second sequencing primer was as follows:

- 15 5'-AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCG (SEQ ID No. 58)

- The sequencing cycles described in Example 8 can be repeated to sequence the tag. In this case, only three cycles of sequencing were performed to analyse the tag. The data in
- 20 Figure 4 show that all the clusters sequenced aligned against a known sample (PhiX, BAC or E. coli contamination), and that for the samples that aligned correctly, 98% showed the correct tag. The 0.1% errors derive from mis-alignments of the target read, rather than errors in the tag sequences.

25

Example 10) 12-plex paired reads from a library of fragments

Step 1) Grafting primers onto surface of SFA coated chip

- 83 -

An SEA coated chip is placed onto a modified MJ-Research thermocycler and attached to a peristaltic pump. Grafting mix consisting of 0.5 μ M of a forward primer and 0.5 μ M of a reverse primer in 10 mM phosphate buffer (pH 7.0) is pumped into the channels of the chip at a flow rate of 60 μ l/minutes for 75 s at 20 °C. The thermocycler is then heated to 51.6 °C, and the chip is incubated at this temperature for 1 hour. During this time, the grafting mix undergoes 18 cycles of pumping: grafting mix is pumped in at 15 μ l/minutes for 20 s, then the solution is pumped back and forth (5 s forward at 15 μ l/minutes, then 5 s backward at 15 μ l/minutes) for 180 s. After 18 cycles of pumping, the chip is washed by pumping in 5xSSC/5mM EDTA at 15 μ l/minutes for 300 s at 51.6 °C. The thermocycler is then cooled to 20 °C.

The primers are typically 5'-phosphorothioate oligonucleotides incorporating any specific sequences or modifications required for cleavage. Their sequences and suppliers vary according to the experiment for which they are to be used, and in this case were complementary to the 5'-ends of the template duplex. The DNA sequence used in this process was the pool of the twelve libraries, which have ends complementary to the grafted primers. The library mix was denatured using sodium hydroxide treatment followed by snap dilution as described.

The grafted primers contain a sequence of T bases at the 5'-end to act as a spacer group to aid linearisation and hybridization. The sequences of the three primers grafted to the chip are as follows:

Oligo A: 5'-PS-TTTTTTTTTTAAATGATACGGCGACCACCGAGAUCTACAC-3'
where U = 2-deoxyuridine; (SEQ ID No. 59)

- 84 -

Oligo B: 5'-PS-TTTTTTTTTTCAAGCAGAAGACGGCATAACGoxoA-3', where Goxo = 8-oxoguanine). (SEQ ID No. 60)

5 Step 2) Preparation of clusters by isothermal amplification:

The DNA sequence used in the amplification process is the mixture of the twelve libraries prepared in Example 1, which have ends complementary to the grafted primers. The
10 duplex DNA (1 nM) is denatured using 0.1 M sodium hydroxide treatment followed by snap dilution to the desired 0.2-2 pM 'working concentration' in 'hybridization buffer' (5 x SSC / 0.1% Tween).

Surface amplification was carried out by isothermal
15 amplification using a commercially available Solexa/Illumina cluster station as described in WO08002502. The cluster station is essentially a hotplate and a fluidics system for controlled delivery of reagents to a flow cell.

The single stranded template (denatured as indicated
20 above) is hybridised to the grafted primers immediately prior to the amplification reaction, which thus begins with an initial primer extension step rather than template denaturation. The hybridization procedure begins with a heating step in a stringent buffer to ensure complete
25 denaturation prior to hybridisation. After the hybridization, which occurs during a 20 minute slow cooling step, the flowcell was washed for 5 minutes with a wash buffer (0.3 x SSC / 0.1% Tween).

During template hybridization and first extension, a
30 number of solutions/buffers are typically employed, e.g., a solution comprising the DNA samples, a hybridization buffer (5 x SSC / 0.1% tween), a wash buffer (0.3 x SSC / 0.1%

- 85 -

tween), a 2M sodium hydroxide solution, a cluster buffer (200 mM Tris, 100 mM Ammonium Sulfate, 20 mM Magnesium sulfate, 1% Triton, 1.3% DMSO, pH 8.8); an amplification additive (5 M betaine), DNA polymerase, and 10 mM dNTP mix.

5 To prepare the hybridization mixes, a 0.2 ml strip sample tube and the hybridization buffer are pre-chilled. Using 1.7 ml Eppendorf tube(s), the DNA template(s) are then diluted to 1nM in buffer EB (Qiagen). 1 µL of 2 M NaOH is added to 19 µL of template, vortexed briefly and incubated
10 for 5 minutes at room temperature to denature the DNA template into single strands. The denatured DNA is diluted to working concentration (0.2-2 pM) in pre-chilled hybridization buffer (e.g. for 1 mL of 1pM Hybridization mix, 1 µL of denatured DNA is diluted into 1 mL of pre-
15 chilled hybridization buffer). The volume required depends on the number of channels used - at least 120 µL of hybridization mix per channel is optionally used. Thus, 1 mL of hybridization mix is enough for 8 channels. The samples are vortexed briefly, spun down and aliquoted into
20 the pre-chilled 0.2 ml strip tubes (with no bubbles in the bottom of the tubes) and used immediately.

 To prepare the Amplification pre-mix (of sufficient volume for the first extension and 35 cycles of isothermal amplification), 35 mL of H₂O (milliQ), 7 mL of Cluster
25 buffer (200 mM Tris, 100 mM Ammonium Sulfate, 20 mM Magnesium sulfate, 1% Triton, 1.3% DMSO, pH 8.8), and 28 mL of Amplification additive (5 M betaine solution) are mixed to achieve a final volume of 70 mL.

 To prepare the first extension Taq mix, 780 µL of
30 Amplification pre-mix, 16 µL of 10 mM dNTPs, and 4 µL of Taq

- 86 -

DNA polymerase are mixed together for a final volume of 800 μ l.

A typical amplification process is detailed in the following table (Table 1), detailing the flow volumes per channel, controlled automatically by the computer component of the invention.

Table 1. Template hybridization and first extension.

Step	Description	T (°C)	Time (sec)	Flow rate (μ l/min)	Pumped V (μ l)
1	Pump Hybridization pre-mix	20	120	60	120
2	Pump Hybridization mix	96	300	15	75
3	Remove bubbles	96	6	100	10
4	Stop flow and hold T	96	30	static	0
5	Slow cooling	96-40	1120	static	0
6	Pump wash buffer	40	300	15	75
7	Pump amplification pre-mix	40	280	15	70
8	Pump amplification mix	40	95	60	95
9	First Extension	74	90	static	0
10	cool to room temperature	20	0	static	0

Isothermal amplification at 60 °C using formamide as denaturant

The copied DNA can be isothermally amplified into clusters at 60 °C using formamide as a denaturant. The isothermal amplification (including both temperature control

- 87 -

and reagent control) is overseen by the computer component. Table 2 gives outlines of exemplary script controls. After the isothermal amplification, and optional washing step occur, the nucleic acid of the clusters is ready to be linearized (see below).

Table 2. Isothermal amplification

Step	Description	T (°C)	Time (sec)	Flow rate (µl/min)	Pumped V (µl)
(1) This sequence 35 times	Pump Formamide	60	55	30	28
	Pump Amplification pre-mix	60	55	30	28
	Pump Bst mix	60	72	30	36
2	Pump wash buffer	60	280	30	140
3	Pump Storage Buffer	20	380	15	95

10 Wash buffer = 0.3 x SSC / 0.1% Tween

Amplification pre mix = 2 M betaine, 20 mM Tris, 10 mM Ammonium Sulfate, 2 mM Magnesium sulfate, 0.1% Triton, 1.3% DMSO, pH 8.8

Bst mix = 2 M betaine, 20 mM Tris, 10 mM Ammonium Sulfate, 2
 15 mM Magnesium sulfate, 0.1% Triton, 1.3% DMSO, pH 8.8 plus
 200 µM dNTPs and 80 units/mL of Bst polymerase (NEB Product ref M0275L)
 Storage Buffer = 5X SSC.

20

Step 3) Preparation of clusters for first sequencing read

The preparation for read one was performed on the Illumina cluster station. All volumes used in the protocol were 95 µl
 25 per lane unless otherwise stated. Linearisation of A-type

- 88 -

surface immobilised oligonucleotides was achieved by incubation with USER enzyme mix (cocktail of Uracil DNA Glycosylase and Endonuclease VIII, NEB #M5505, 10 U/ml, 10 mM KCl, 20 mM Tris pH 8.8, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 0.1% Triton X-100, 37 °C, 30 minutes). All exposed 3'-OH termini of DNA, either from the extended template or unextended surface oligonucleotides were blocked by dideoxy chain termination using a cocktail of terminal transferase (0.25 U/μl) and a modified polymerase (SBS polymerase as described below) (0.015 mg/ml, 100 μM ddNTP, 50 mM tris, 50 mM NaCl, 6 mM MgSO₄, 1 mM EDTA, 0.05 % Tween 20). Blocking was achieved in a two stage protocol, initial incubation at 37 °C for 30 minutes followed by a ramping to 60 °C and incubating the flowcell for a further 15 minutes). Linearised and blocked clusters were washed with 0.3X SSC and storage buffer prior to denaturation with 0.1N NaOH. Denatured clusters were neutralised with TE buffer (10 mM Tris pH 8.0, 1 mM EDTA) and washed with storage buffer. The read 1 specific sequencing primer (5'- ACACTCTTTCCTACACGACGCTCTTCCGATC -3' (SEQ ID No. 61), 0.5 μM in hybridisation buffer) was annealed to the clusters by incubation at 60 °C for 15 minutes, followed by a 0.3X SSC wash at 40 °C (ramp rate 1 °C / sec). The flowcell was finally flushed with storage buffer (at 20 °C). Processed flowcells were transferred to the Illumina Genome Analyser for sequencing read 1.

Step 4) Sequencing from the target fragment

- 89 -

Sequencing of the clusters from the above illustrative protocol was carried out using modified nucleotides prepared as described in International patent application WO 2004/018493, and labeled with four spectrally distinct
5 fluorophores, as described in PCT application number PCT/GB2007/001770, published as WO07135368. Sequencing of clusters is described in more detail in patent WO06064199. The contents of the above-listed documents are incorporated herein by reference in their entireties.

10 A mutant 9°N polymerase enzyme (an exo- variant including the triple mutation L408Y/Y409A/P410V and C223S) (SBS polymerase) was used for the nucleotide incorporation steps.

All processes were conducted as described in the Illumina
15 Genome Analyser operating manual. The flowcell was mounted to the analyser, primed with sequencing reagents: position #1 = incorporation mix (1 µM dNTP mix, 0.015 µg/ml SBS polymerase, 50 mM Tris pH 9.0, 50 mM NaCl, 6 mM MgSO₄, 1 mM EDTA, 0.05 % Tween 20); position # 2 = spare (MilliQ water
20 only); position # 3 = scan mix (100 mM Tris pH 7.0, 50 mM sodium ascorbate); position #4 = High salt wash (5x SSC, 0.05 % Tween 20); position # 5 = incorporation buffer (50 mM Tris pH 9.0, 50 mM NaCl, 1 mM EDTA, 0.05 % Tween 20);
25 position #6 = cleavage mix (100 mM TCEP, 100 mM Tris pH 9.0, 100 mM NaCl, 50 mM sodium ascorbate, 0.05% Tween 20);
position #7 = cleavage buffer (100 mM Tris pH 9.0, 100 mM NaCl, 0.05% Tween 20); position #8 = spare (single reads) or connected to PE module outlet (paired read experiments). Flowcells were sequenced using standard sequencing recipes
30 for either 27- or 37-cycle experiments. Data was analysed using the standard analysis pipeline.

- 90 -

Incorporation: Prime with Incorporation buffer, 125 μ L/channel; 60 μ L/minutes, Heat to 60°C

Treat with Incorporation mix, 75 μ L/channel; 60 μ L/minutes.

- 5 Wait for a total of 15 minutes in addition to pumping fresh Incorporation mix, 25 μ L/channel; 60 μ L/minutes, every 4 minutes.

Cool to 20 °C.

- Wash with Incorporation buffer, 75 μ L/channel; 60 μ L/minutes.
- 10

Wash with 5XSSC/0.05% Tween 20, 75 μ L/channel; 60 μ L/minutes

Prime with imaging buffer, 100 μ L/channel; 60 μ L/minutes

Scan in 4 colors at RT.

- 15 Cleavage: Prime with Cleavage buffer (0.1M Tris pH 7.4, 0.1M NaCl and 0.05% Tween 20), 125 μ L/channel; 60 μ L/minutes.

Heat to 60 °C.

- Treat the clusters with Cleavage mix (100 mM TCEP in Cleavage buffer), 75 μ L/channel; 60 μ L/minutes.
- 20

Wait for a total of 15 minutes in addition to pumping fresh cleavage mix, 25 μ L/channel; 60 μ L/minutes, every 4 minutes.

Cool to 20 °C.

Wash with Enzymology buffer.

- 25 Wash with 5XSSC/0.05% Tween 20.

- 91 -

Repeat the process of Incorporation and Cleavage for as many cycles as required.

Incorporated nucleotides were detected using the
5 Illumina genome analyzer, a Total Internal Reflection based
fluorescent CCD imaging apparatus described in "Systems and
Devices for Sequence by Synthesis Analysis," USSN
60/788,248, filed March 31, 2006 and corresponding PCT
application PCT/US07/07991, published as WO07123744, the
10 contents of which are incorporated herein by reference in
their entirety.

Step 5) Sequencing of the Tag

Following the successful completion sequencing the
15 first read on the Genome Analyser, flowcells remained
mounted and were prepared for tag sequencing in situ, using
the Illumina Paired End module. Temperature control was
achieved by using the Genome Analyser peltier. All flow
rates were 60 μ l/min and 75 μ l per lane unless otherwise
20 stated. Clusters were denatured with 0.1 M NaOH to remove
the extended sequencing primer from read 1. Hybridisation of
the tag sequencing primer

(5'-GATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3') (SEQ ID No. 62),
0.5 μ M in hybridisation buffer) as described for read 1.
25 Sequencing of the tag was performed on the Illumina Genome
Analyser.

- 92 -

Step 6) Preparation of the clusters for the paired sample read

Following the successful completion sequencing of two
5 reads on the Genome Analyser, flowcells remained mounted and
were prepared for paired read 2 *in situ*, using the Illumina
Paired End module. Temperature control was achieved by using
the Genome Analyser peltier. All flow rates were 60 µl/min
and 75 µl per lane unless otherwise stated. Clusters were
10 denatured with 0.1 M NaOH to remove the extended sequencing
primer from the tag sequencing. Clusters were 3'-
dephosphorylated using T4 polynucleotide kinase (Invitrogen
#18004-010, 200 U/ml, 50 mM imidazole pH 6.4, 12 mM MgCl₂,
70 µM ADP, 1 mM 2-mercaptoethanol, 37 °C, 30 minutes), prior
15 to re-synthesis of the A-strand achieved using 15 cycles of
60 °C isothermal amplification (same reagents and conditions
as described for cluster creation except conducted at 30
µl/min). Clusters were washed before and after resynthesis
with 0.3X SSC (150 µl and 245 µl respectively).
20 Linearisation of the B-strand of the re-synthesised clusters
was achieved by the excision of 8-oxoguanine from the B-type
oligo using Fpg (formamidopyrimidine DNA glycosylase, NEB
#M0240, 80 U/ml, 10 mM Bis Tris propane pH 7.0, 10 mM MgCl₂,
1mM dithiothreitol, 37 °C, 30 minutes). Blocking was
25 performed as described for read 1 using the same blocking
cocktail. Linearised and blocked clusters were denatured
prior to hybridisation of the read 2 specific sequencing
primer
(5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC-3') (SEQ ID No. 63),
30 0.5 µM in hybridisation buffer) as described for read 1.

- 93 -

Read 2 of the processed flowcells was subsequently sequenced on the Illumina Genome Analyser.

Data generated from the 12 samples is shown in figure 8. In each case the tag sequence could unambiguously
5 determine the source of the sample. Each sample from the 12 was represented by a reasonable number of clusters, and for each tag read, the large majority (> 90%) of the reads aligned against the source genome identified by the tag.

While the foregoing invention has been described in
10 some detail for purposes of clarity and understanding, it will be clear to one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention. For example, all the techniques and apparatus
15 described above may be used in various combinations. All publications, patents, patent applications, or other documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication, patent, patent
20 application, or other document were individually indicated to be incorporated by reference for all purposes.

- 94 -

Claims:

1. A method for sequencing nucleic acid sequences on an array and identifying subsets of nucleic acid sequences on an array, wherein each subset of nucleic acid sequences is isolated from a different source, said method comprising the steps of:
- 5 an array, wherein each subset of nucleic acid sequences is isolated from a different source, said method comprising the steps of:
- (a) providing at least two samples of randomly fragmented double stranded nucleic acid targets, wherein each of said randomly fragmented double stranded nucleic acid targets is isolated from a different source;
- 10 (b) ligating a universal adaptor to the ends of each target fragment of each sample to generate adaptor-target-adaptors of each sample, wherein each of said adaptor-target-adaptors comprises a target fragment flanked by universal adaptor sequences and said universal adaptor comprises a region of double stranded nucleic acid and at least one region of single stranded nucleic acid;
- 15 (c) amplifying adaptor-target-adaptors of each sample with two or more sample specific amplification primers to generate amplified nucleic acids, wherein one of said amplification primers comprises a sample specific tag sequence, and wherein amplified nucleic acids of each sample comprise said sample specific tag sequence and said sample specific tag sequence differentiates amplified nucleic acids originating from different samples;
- 20 (d) pooling the amplified nucleic acids of different samples;
- (e) immobilising the amplified nucleic acids of different samples on an array to generate an array of immobilised fragments;
- 25 30

- 25 -

(f) sequencing the immobilised fragments on the array to determine a sequence read of each immobilised target fragment and identify the sample specific tag sequence of each immobilised fragment, thereby determining both a
5 nucleic acid sequence of said immobilized fragment and identifying the immobilized fragment as a member of a subset of nucleic acids on the array.

2. A method for sequencing nucleic acid sequences on an
10 array and identifying subsets of nucleic acid sequences on an array, wherein each subset of nucleic acid sequences is isolated from a different source, said method comprising the steps of:

(a) providing at least two samples of randomly fragmented
15 double stranded nucleic acid targets, wherein each of said randomly fragmented double stranded nucleic acid targets is isolated from a different source;

(b) ligating a sample specific tagged adaptor to the ends of each target fragment of each sample to generate adaptor-
20 target-adaptors of each sample, wherein each of said adaptor-target-adaptors comprises a target fragment flanked by tagged adaptor sequences and said tagged adaptor comprises a region of double stranded nucleic acid and at least one region of single stranded nucleic acid and a
25 sample specific tag that differentiates adaptor-target-adaptors originating from different samples;

(c) pooling the adaptor-target-adaptors of different samples to generate a pooled sample of adaptor-target-adaptors;
30 (d) amplifying the pooled sample of adaptor-target-adaptors with a pair of universal amplification primers to generate an amplified pooled sample;

- 96 -

(e) immobilising the amplified pooled sample on an array to generate an array of immobilised fragments;
(f) sequencing the array to determine a sequence read of each immobilised target fragment and identify the sample specific tag sequence of each immobilised fragment, thereby
5 determining both a nucleic acid sequence of said immobilized fragment and identifying the immobilized fragment as a member of a subset of nucleic acids on the array.

- 10 3. A method for sequencing nucleic acid sequences on an array and identifying subsets of nucleic acid sequences on an array, wherein each subset of nucleic acid sequences is isolated from a different source, said method comprising the steps of:
- 15 (a) providing at least two samples of randomly fragmented double stranded nucleic acid targets, wherein each of said randomly fragmented double stranded nucleic acid targets is isolated from a different source;
- (b) ligating a sample specific tagged adaptor to the ends of
20 each target fragment of each sample to generate adaptor-target-adaptors of each sample, wherein each of said adaptor-target-adaptors comprises a target fragment flanked by tagged adaptor sequences and said tagged adaptor comprises a region of double stranded nucleic acid and at
25 least one region of single stranded nucleic acid and a sample specific tag that differentiates adaptor-target-adaptors originating from different samples, and wherein the sample specific tag sequence is attached directly to the target fragment in each adaptor-target-adaptor with no
30 intervening nucleotides between the sample specific tag sequence and target fragment sequences;

- 97 -

(c) amplifying adaptor-target-adaptors of each sample with a pair of amplification primers complementary to the sample specific tag sequence to generate amplified nucleic acids from each sample;

5 (d) pooling the amplified nucleic acids from each sample to generate a pool of amplified nucleic acid samples;

(e) immobilising the pool of amplified nucleic acid samples on an array to generate an array of immobilised fragments;

10 (f) sequencing the array to determine a sequence read of each immobilised target fragment and identify the sample specific tag sequence of each immobilised fragment, thereby determining both a nucleic acid sequence of said immobilized fragment and identifying the immobilized fragment as a member of a subset of nucleic acids on the array.

15

4. The method according to claim 1 or 2, wherein said sequencing is performed in two separate reads.

20 5. The method according to claim 3, wherein said sequencing of the target fragment and the sample specific tag sequence is performed in a single sequencing read.

25 6. The method according to claims 1 to 3, wherein the pooled samples are amplified on a solid surface prior to sequencing.

7. The method according to claim 6, wherein said solid surface is a collection of beads.

30 8. The method according to claim 7, wherein said beads are isolated as single beads in an emulsion.

- 98 -

9. The method according to claims 1 to 3, wherein the array of immobilized fragments is amplified prior to sequencing to generate an amplified array.

5 10. The method according to claim 9, wherein said amplified array is a clustered array of amplified single molecules.

11. The method according to claim 10, wherein the clustered array is formed by solid-phase nucleic acid amplification.

10

12. The method according to claim 11, wherein the clustered array is formed by isothermal solid-phase nucleic acid amplification.

15 13. The method according to any preceding claim, wherein said sequencing involves cycles of ligation with labelled oligonucleotides.

14. The method according to any preceding claim, wherein
20 said sequencing involves cycles of addition of nucleotides.

15. The method according to claim 14, wherein said nucleotides are labelled.

25 16. The method according to claims 13 or 15, wherein said labels are fluorophores.

17. The method according to any preceding claim, wherein the sequencing step is carried out on an array of immobilised
30 beads.

- 99 -

18. The method according to claim 1 wherein the nucleic acid target fragments can undergo a sequencing reaction from both ends to obtain a paired end read.

5 19. The method according to claim 18 wherein both ends of the nucleic acid fragment are sequenced before the sequence of the tag is determined.

20. The method according to claim 18 wherein the read from
10 the second end of the fragment occurs after the tag is determined.

21. The method according to claim one that uses two
amplification primers per sample, one of which carries the
15 tag.

22. The method according to claim one that uses three
amplification primers per sample, one of which carries the
tag.

20

23. A kit comprising a double stranded nucleic acid adapter
and at least two different primer polynucleotides that carry
a different tag sequence that does not hybridise to the
adapter.

25

24. A kit according to claim 23 further comprising a
nucleotide polymerase and four nucleoside triphosphates.

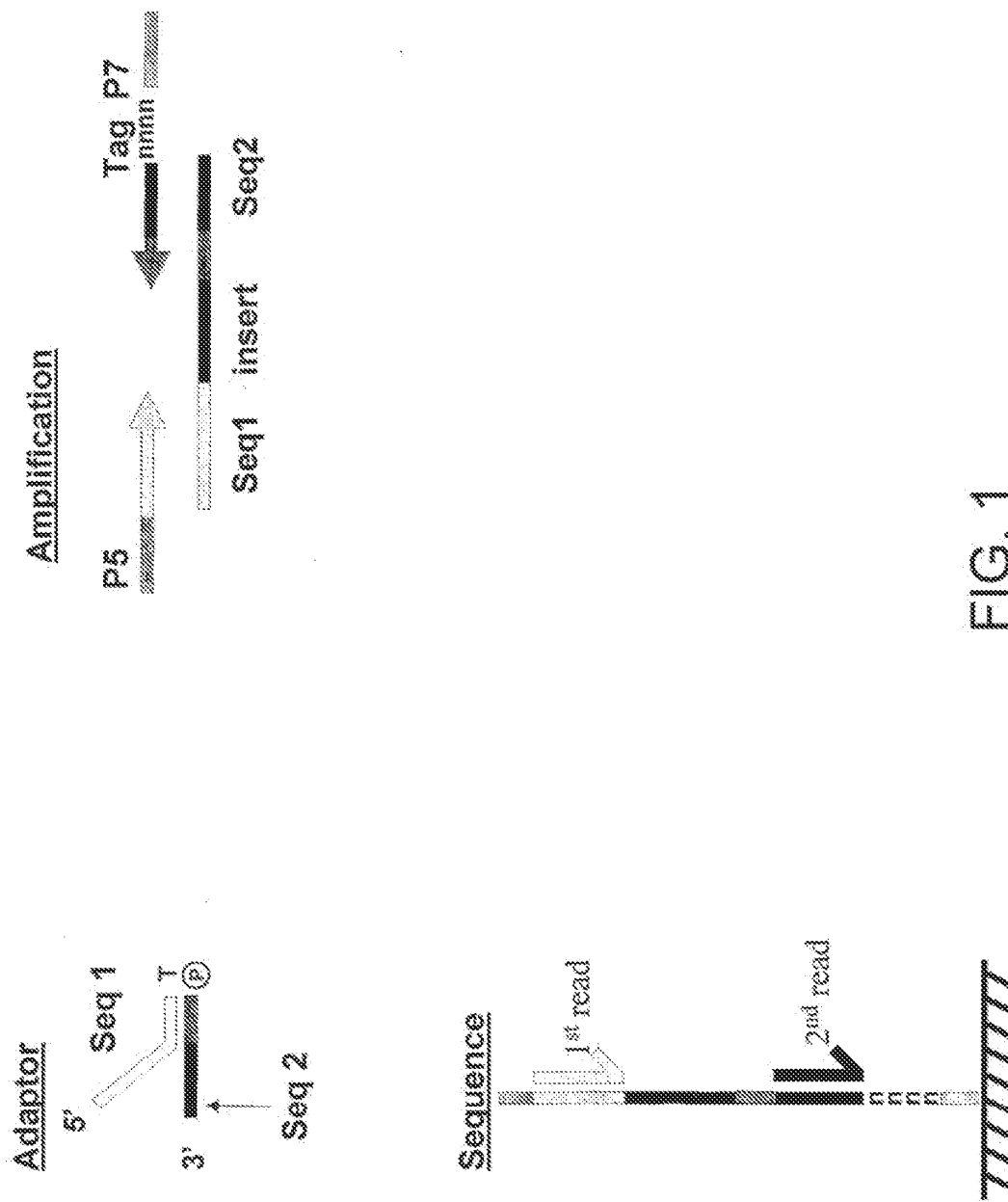


FIG. 1

2 / 11

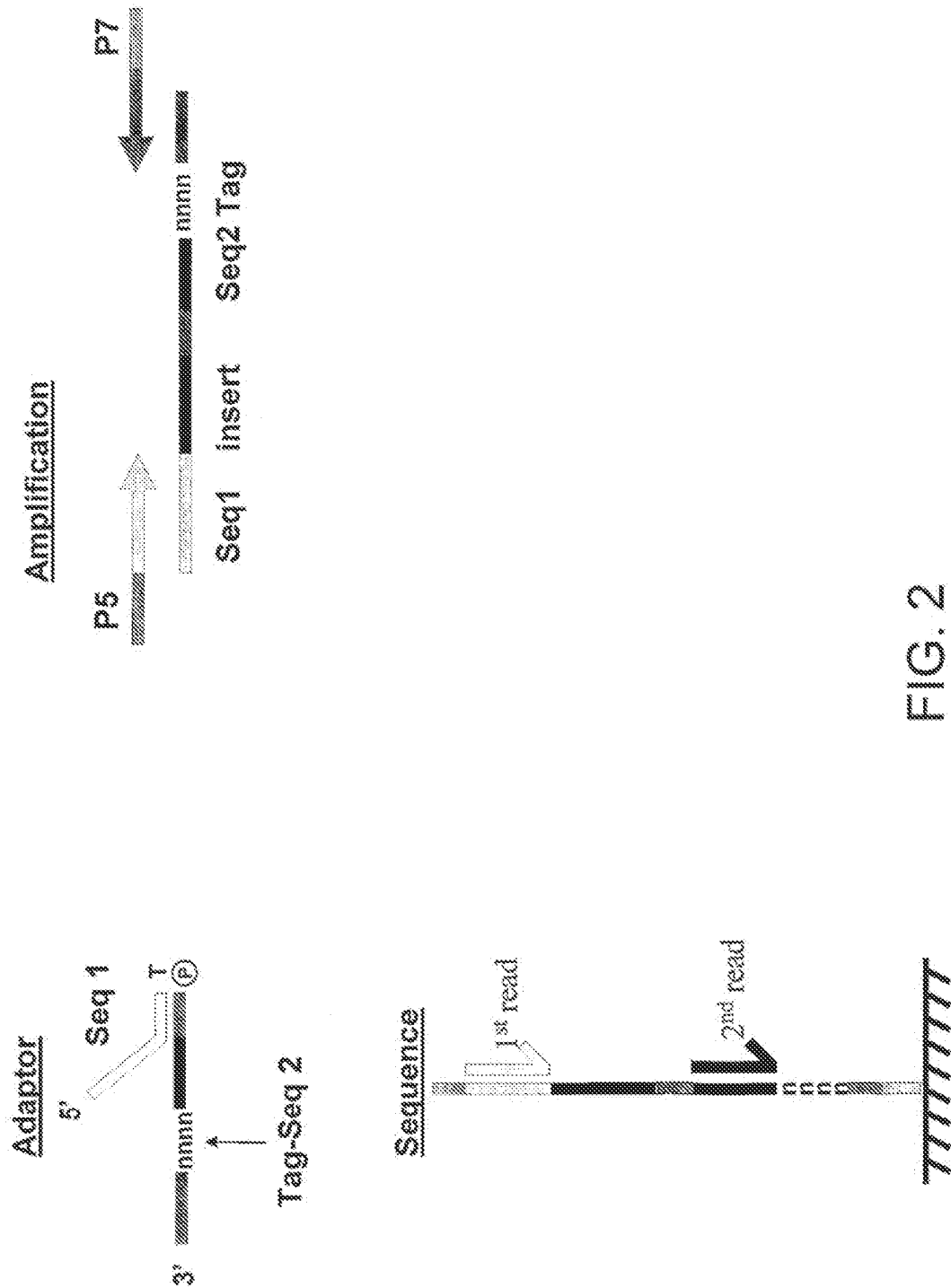


FIG. 2

SUBSTITUTE SHEET (RULE 26)

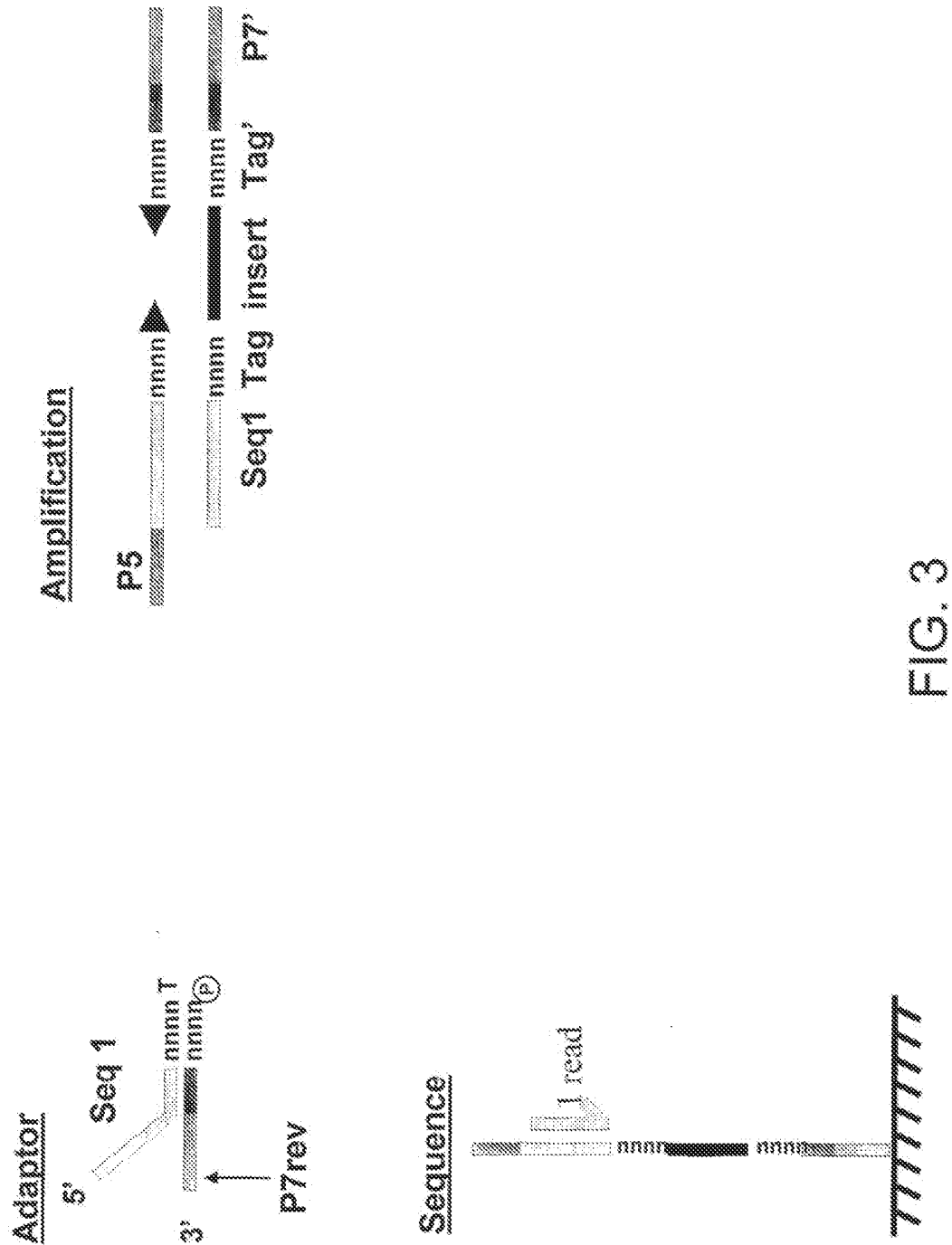


FIG. 3

SUBSTITUTE SHEET (RULE 26)

- Mixed two libraries and sequenced by SBS in one channel
- Data from a single tile

<u>Reference Sequences</u>	<u>% of Clusters Aligning to Reference</u>
PhiX	68
BAC	29
E.coli	3

	<u>PhiX TAG</u>	<u>BAC TAG</u>	<u>Other TAGs</u>
PhiX	98 %	0.1 %	2 %
BAC	0.1 %	98 %	2 %

- Majority of BAC and PhiX inserts were sequenced with the correct TAG

FIG. 4

SUBSTITUTE SHEET (RULE 26)

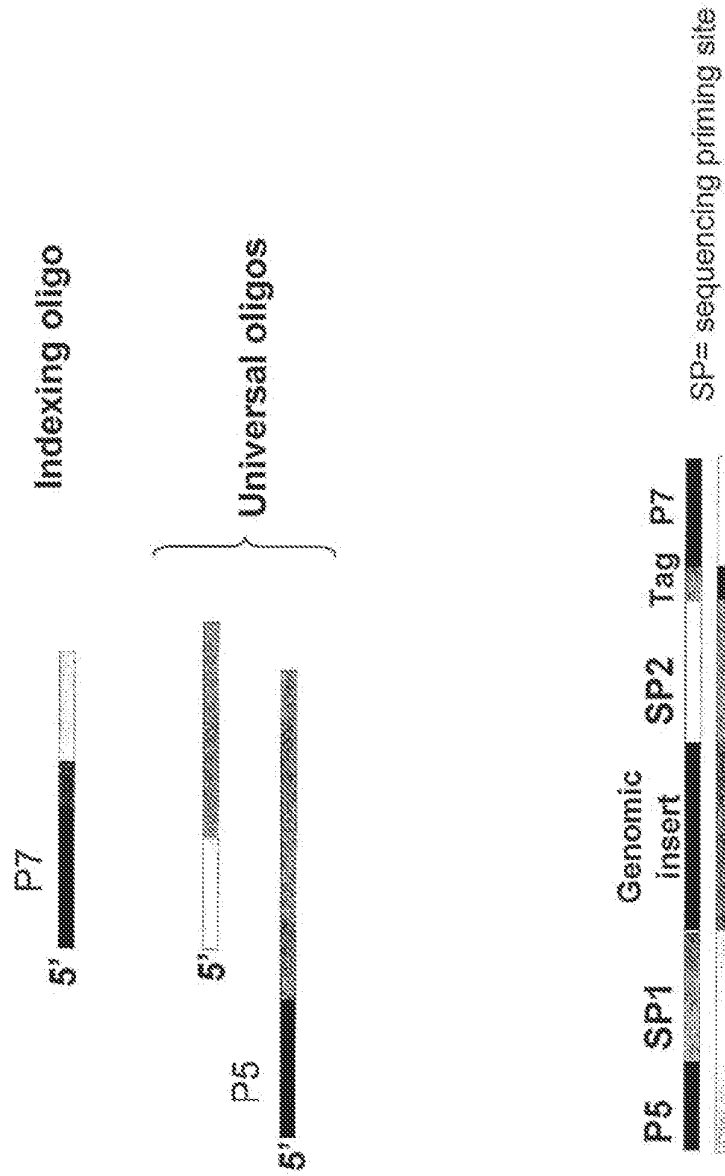
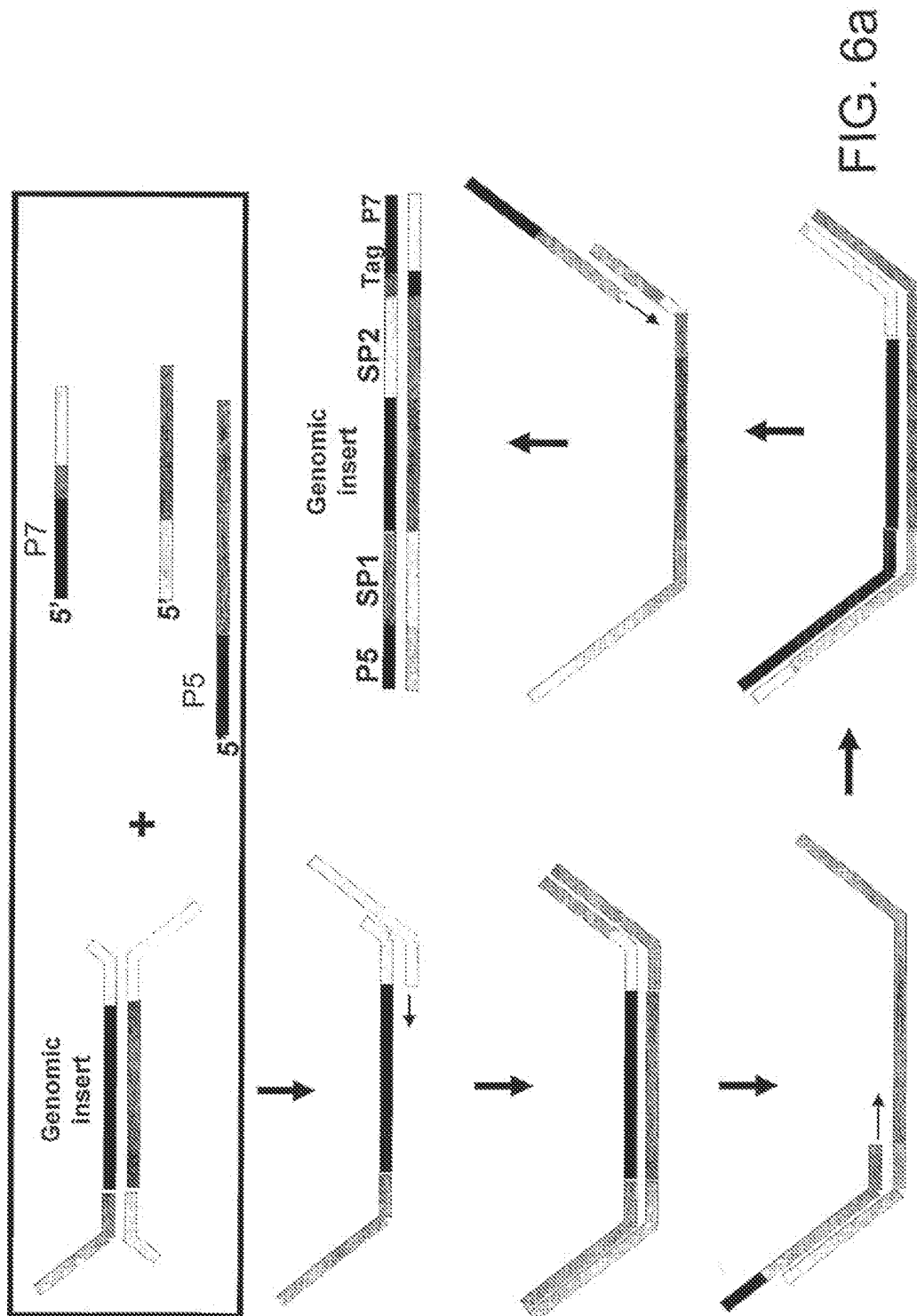
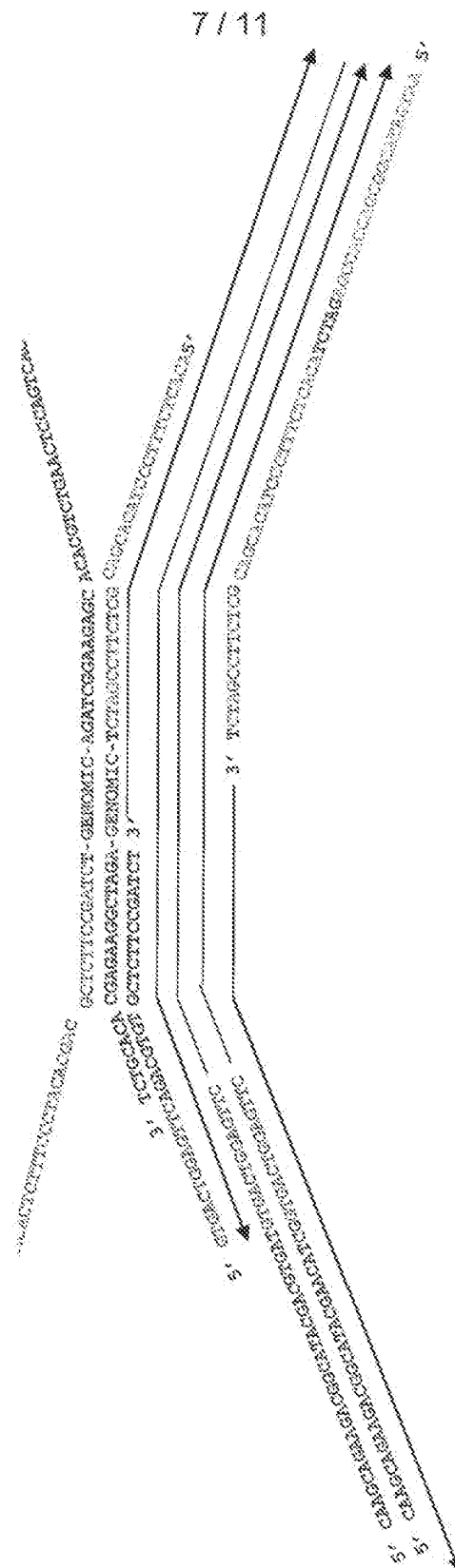


FIG. 5

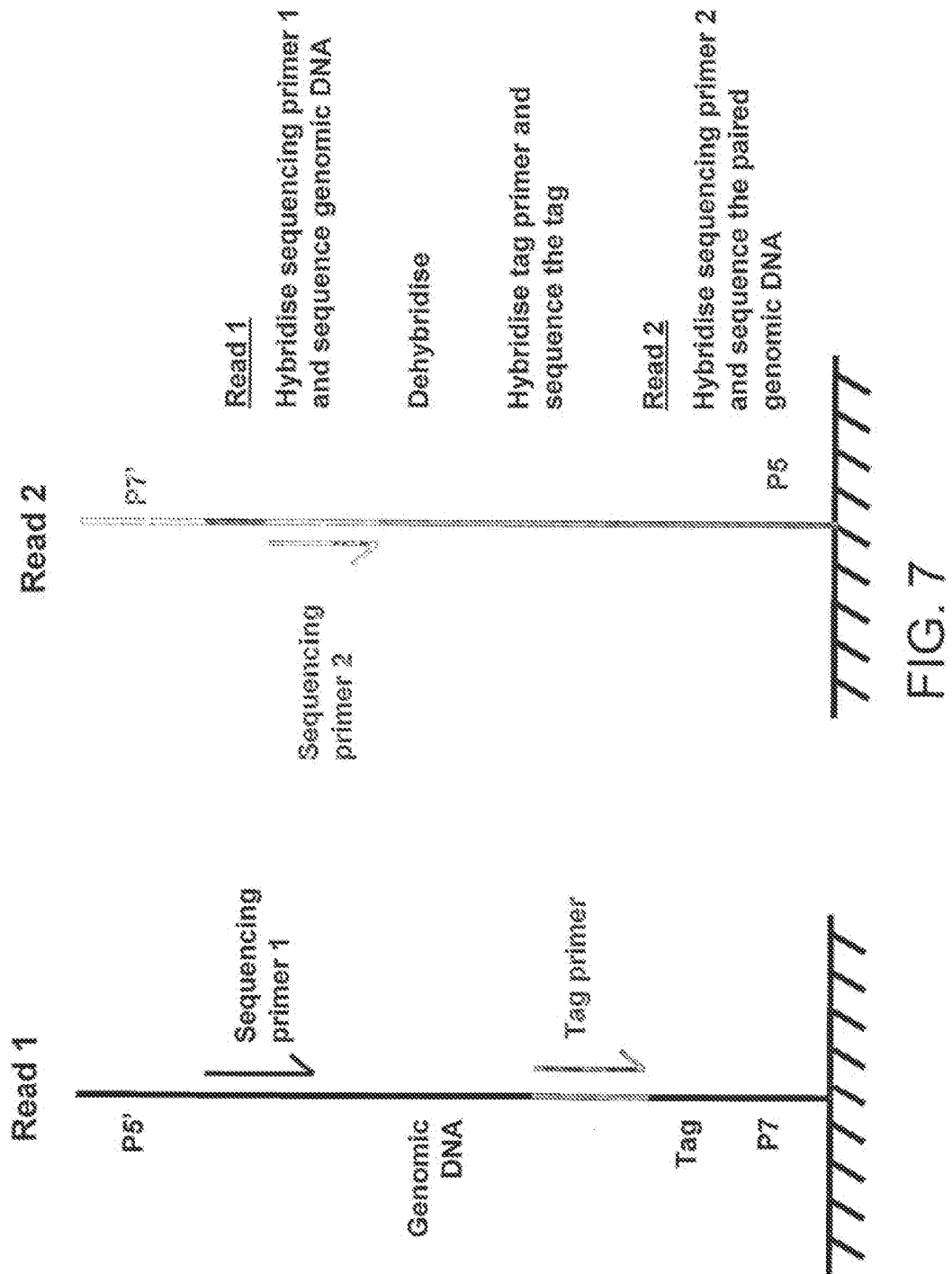
SUBSTITUTE SHEET (RULE 26)



SUBSTITUTE SHEET (RULE 26)

[illegible]

SUBSTITUTE SHEET (RULE 26)



SUBSTITUTE SHEET (RULE 26)

9 / 11

Tag	Genome	Clus No	BAC R1	BAC R2	PhiX R1	PhiX R2
ATCAGG	PhiX	6522	0%	0%	96%	92%
CGATGT	PhiX	3548	0%	0%	96%	92%
TTAGGC	PhiX	10394	0%	0%	97%	93%
TGACCA	PhiX	6842	0%	0%	97%	93%
ACAGTG	PhiX	6677	0%	0%	97%	93%
CAGATC	PhiX	11231	0%	0%	96%	93%
GCCAAT	BAC	6174	97%	92%	0%	0%
ACTIGA	BAC	5375	97%	93%	0%	0%
GATCAG	BAC	5404	97%	93%	0%	0%
TAGCTT	BAC	10239	97%	94%	0%	0%
GGCTAC	BAC	9734	97%	93%	0%	0%
CTTGTA	BAC	9704	97%	93%	0%	0%

Expected Tag	Observed Tag	Genome	Clus No	BAC R1	BAC R2	PhiX R1	PhiX R2
TGACCA	TGACAA	PhiX	8	0%	0%	100%	88%
GCCAAT	TCCAAT	BAC	34	97%	94%	0%	0%
GATCAG	TATCAG	BAC	28	100%	96%	0%	0%

FIG. 8

SUBSTITUTE SHEET (RULE 26)

10/11

Forked Adaptor Primers SBS3/SBS12 (indexing/paired ends)
 Adaptor oligo 2 5' ACACTCTTTCCCTACACGACGCTCTTCCGATCT 3'
 Adaptor oligo 1 (short) 5' P-GATCGAAGAGCACACGCTCT 3'

5' ACACTCTTTCCCTACACGAC GCTCTTCCGATCT 3'
 3' TCTCCACGACGAGGCTAG 5'

PCR Primer
 Universal LPX primer 5' GTGACTGGAGTTCAGACGGTGTGCTCTTCCGATCT 3'
 Lpx primer 5' AATGTAACGGCGACACCGAGATCTACATCTTTCCCTACACGACGCTCTTCCGATCT
 tag1 primer (CGTGAT) 5' CAAGCAGAAGACGGCATACGAGATCGTGTATGTGACTGGAGTTC 3'
 tag2 primer (ACATCG) 5' CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTC 3'
 tag3 primer (GCCTAA) 5' CAAGCAGAAGACGGCATACGAGATGCTTAAGTGAAGTGGAGTTC 3'
 tag4 primer (TCGTCA) 5' CAAGCAGAAGACGGCATACGAGATTTGTTCACTGACTGGAGTTC 3'
 tag5 primer (CACTGT) 5' CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTC 3'
 tag6 primer (ATTGGC) 5' CAAGCAGAAGACGGCATACGAGATATTGTCGTGACTGGAGTTC 3'
 tag7 primer (GATCTG) 5' CAAGCAGAAGACGGCATACGAGATGATCTGCTGACTGGAGTTC 3'
 tag8 primer (TCAAGT) 5' CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTC 3'
 tag9 primer (CTGATC) 5' CAAGCAGAAGACGGCATACGAGATCTGATCTGCTGACTGGAGTTC 3'
 tag10 primer (AAGCTA) 5' CAAGCAGAAGACGGCATACGAGATAAGCTAGTGAAGTGGAGTTC 3'
 tag11 primer (GTAGCC) 5' CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTC 3'
 tag12 primer (TACAA) 5' CAAGCAGAAGACGGCATACGAGATTACAGTGTGACTGGAGTTC 3'

Sequencing primers

SBS12 5' GTGACTGGAGTTCAGACGGTGTGCTCTTCCGATC 3'
 SBS12rev 5' GATCGAAGAGCACACGCTCTGAACTCCAGTCAC 3'
 SBS3 5' ACACTCTTTCCCTACACGACGCTCTTCCGATC 3'

FIG. 9

SUBSTITUTE SHEET (RULE 26)

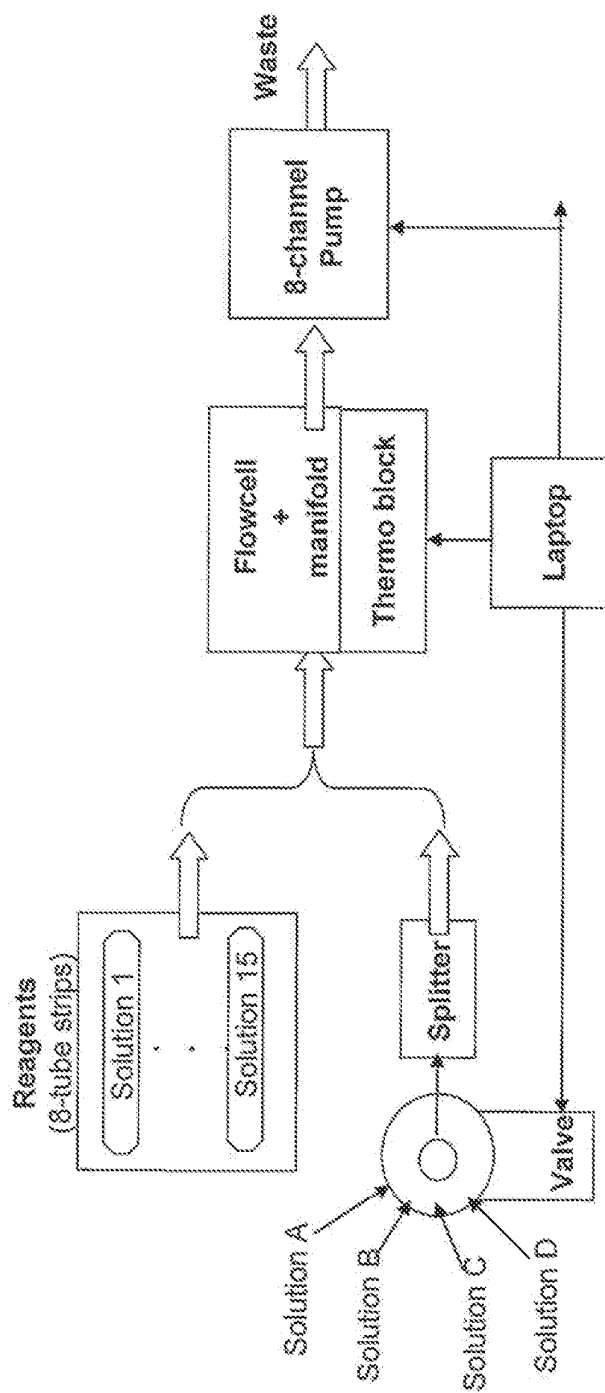


FIG. 10

Electronic Patent Application Fee Transmittal				
Application Number:		14540917		
Filing Date:		13-Nov-2014		
Title of Invention:		COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ		
First Named Inventor/Applicant Name:		Douglas AMORESE		
Filer:		Jeffrey R. Bolland		
Attorney Docket Number:		NUGN-001/01US		
Filed as Large Entity				
Filing Fees for Utility under 35 USC 111(a)				
Description	Fee Code	Quantity	Amount	Sub-Total in USD(\$)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD(\$)
Miscellaneous:				
Submission- Information Disclosure Stmt	1806	1	180	180
Total in USD (\$)				180

Electronic Acknowledgement Receipt	
EFS ID:	23247867
Application Number:	14540917
International Application Number:	
Confirmation Number:	7076
Title of Invention:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ
First Named Inventor/Applicant Name:	Douglas AMORESE
Customer Number:	58249
Filer:	Jeffrey R. Bolland
Filer Authorized By:	
Attorney Docket Number:	NUGN-001/01US
Receipt Date:	19-AUG-2015
Filing Date:	13-NOV-2014
Time Stamp:	13:00:37
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	Deposit Account
Payment was successfully received in RAM	\$ 180
RAM confirmation Number	9414
Deposit Account	501283
Authorized User	COOLEY LLP
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:	

File Listing:					
Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1		NUGN-001_01US_Response_to _NFOA.pdf	142680 06d7c5bc36f60c5c36d831ff70f6a6be2af32cfe	yes	11
	Multipart Description/PDF files in .zip description				
	Document Description		Start	End	
	Amendment/Req. Reconsideration-After Non-Final Reject		1	1	
	Claims		2	5	
	Applicant Arguments/Remarks Made in an Amendment		6	11	
Warnings:					
Information:					
2	Information Disclosure Statement (IDS) Form (SB08)	NUGN-001_01US_IDS_August_2015.pdf	69742 2fd9cb51a7792dde42a6323b9a8f7f2464642d3c	no	4
Warnings:					
Information:					
This is not an USPTO supplied IDS fillable form					
3	Non Patent Literature	Myers_Hudon_Alpha_2013.pdf	3679773 86c0dd39853c16b157dd64e7f5757fe376d744d4	no	15
Warnings:					
Information:					
4	Foreign Reference	WO2014144092-PAMPH-164_.pdf	10419986 e6684bdc4b929d13bd60222f106a3e1be79>4tb/	no	68
Warnings:					
Information:					
5	Foreign Reference	WO2012061832-PAMPH-249_.pdf	11652152 828cdb94517cf80709698e82478825c0075afbcd	no	103
Warnings:					
Information:					

6	Foreign Reference	WO2008093098-PAMPH-183_.pdf	13606793 79b481541cfac705dde070df4aec986c1f1e32c6	no	111
Warnings:					
Information:					
7	Non Patent Literature	NUGN-001_01WO_ISR_WO.pdf	2767104 8d3a06bfffdd70945b91edc36d24d78e690d08045	no	17
Warnings:					
Information:					
8	Fee Worksheet (SB06)	fee-info.pdf	30795 e616816b9b263bc57dc608a80ff285c2ecc8396a	no	2
Warnings:					
Information:					
Total Files Size (in bytes):			42369025		
<p>This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.</p> <p><u>New Applications Under 35 U.S.C. 111</u> If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.</p> <p><u>National Stage of an International Application under 35 U.S.C. 371</u> If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.</p> <p><u>New International Application Filed with the USPTO as a Receiving Office</u> If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.</p>					

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)	Application Number		14540917	
	Filing Date		2014-11-13	
	First Named Inventor	AMORESE, Douglas		
	Art Unit	1637		
	Examiner Name	Teresa E. Strzelecka		
	Attorney Docket Number	NUGN-001/01US 301735-2003		

U.S.PATENTS								
Examiner Initial*	Cite No	Patent Number	Kind Code ¹	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear		
	1							
If you wish to add additional U.S. Patent citation information please click the Add button.								
U.S.PATENT APPLICATION PUBLICATIONS								
Examiner Initial*	Cite No	Publication Number	Kind Code ¹	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear		
	1	20120208705	A1	2012-08-16	STEEMERS et al.			
If you wish to add additional U.S. Published Application citation information please click the Add button.								
FOREIGN PATENT DOCUMENTS								
Examiner Initial*	Cite No	Foreign Document Number ³	Country Code ²	Kind Code ⁴	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T ⁵
	1	2008093098	WO	A1	2008-08-07	SOLEXA LIMITED		☐
	2	2012061832	WO	A1	2012-05-10	ILLUMINA, INC.		☐
	3	2014144092	WO	A1	2014-09-18	NUGEN TECHNOLOGIES, INC.		☐

**INFORMATION DISCLOSURE
STATEMENT BY APPLICANT**
(Not for submission under 37 CFR 1.99)

Application Number	14540917
Filing Date	2014-11-13
First Named Inventor	AMORESE, Douglas
Art Unit	1637
Examiner Name	Teresa E. Strzelecka
Attorney Docket Number	NUGN-001/01US 301735-2003

If you wish to add additional Foreign Patent Document citation information please click the Add button

NON-PATENT LITERATURE DOCUMENTS

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T ⁵
	1	MYERS, R. M. Myers lab microRNA-seq Protocol, Hudson Alpha Institute for Biotechnology, May 2, 2013. p. 1-15. Retrieved from ,https://www.encodeproject.org/documents/9fb8e06c-d145-447a-a606-e7f3dc44b549/@@download/attachment/HA_Myers_miRNA-seq_Protocol_2013-05-02.pdf> on February 28, 2015.	☐
	2	International Searching Authority. Written Opinion and International Search Report for PCT/US2014/065530. Date of mailing is March 5, 2015.	☐

If you wish to add additional non-patent literature document citation information please click the Add button

EXAMINER SIGNATURE

Examiner Signature		Date Considered	
--------------------	--	-----------------	--

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

¹ See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. ² Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). ³ For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. ⁴ Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. ⁵ Applicant is to place a check mark here if English language translation is attached.

**INFORMATION DISCLOSURE
STATEMENT BY APPLICANT**
(Not for submission under 37 CFR 1.99)

Application Number	14540917
Filing Date	2014-11-13
First Named Inventor	AMORESE, Douglas
Art Unit	1637
Examiner Name	Teresa E. Strzelecka
Attorney Docket Number	NUGN-001/01US 301735-2003

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

- ☐ That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

- ☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

- ☐ See attached certification statement.
- ☒ The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.
- ☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/Jeffrey R. Bolland/	Date (YYYY-MM-DD)	2015-08-19
Name/Print	Jeffrey R. Bolland	Registration Number	73,759

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these records.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

IN THE CLAIMS:

1. (Currently Amended) A method for detecting a duplicate sequencing read from a population of sample sequencing reads comprising:
 - a) ligating an adaptor to a 5' end of each nucleic acid fragment of a plurality of nucleic acid fragments from one or more samples, wherein the adaptor comprises:
 - (i) an indexing primer binding site;
 - (ii) an indexing site;
 - (iii) an identifier site comprising at least 8 nucleotides; and
 - (iv) a target sequence primer binding site;
 - b) amplifying the adapter-nucleic acid fragment ligated products;
 - c) generating a population of sequencing reads from amplified adapter-nucleic acid fragment ligated products; and
 - d) detecting the population of sequencing reads comprising a sequencing read with a duplicate identifier site and target sequence;

wherein the indexing site is unique amongst a subset of the plurality of nucleic acid fragments and is an index for multiple polynucleotides; and wherein the sequence of the identifier site is not conserved across all adaptor sequences.
2. (Original) The method of claim 1, wherein the method further comprises removing from the population of sequence reads the sequencing read with a duplicate identifier site and target sequence.
3. (Original) The method of claim 1, wherein the identifier site is sequenced with the indexing site.

4. (Withdrawn) The method of claim 1, wherein the identifier site is sequenced separately from the indexing site.
5. (Withdrawn) The method of claim 1, wherein the identifier site is sequenced with the target sequence.
6. (Withdrawn) The method of claim 1, wherein the identifier site is sequenced separately from the target sequence.
7. (Original) The method of claim 1, wherein the adaptor comprises from 5' to 3':
 - (i) the indexing primer binding site;
 - (ii) the indexing site;
 - (iii) the identifier site; and
 - (iv) the target sequence primer binding site.
8. (Withdrawn) The method of claim 1, wherein the adaptor comprises from 5' to 3':
 - (i) the indexing primer binding site;
 - (ii) the indexing site;
 - (iii) the target sequence primer binding site; and
 - (iv) the identifier site.
9. (Original) The method of claim 1, wherein the plurality of nucleic acid fragments is generated from more than one sample.
10. (Original) The method of claim 9, wherein the nucleic acid fragments from each sample has the same indexing site.
11. (Original) The method of claim 10, wherein the sequencing reads are separated based on the indexing site.

12. (Original) The method of claim 11, wherein the separation of sequencing reads is performed prior to step d).
13. (Original) The method of claim 1, wherein the nucleic acid fragments are DNA fragments, RNA fragments, or DNA/RNA fragments.
14. (Original) The method of claim 13, wherein the nucleic acid fragments are genomic DNA fragments or cDNA fragments.
15. (Original) The method of claim 1, wherein the indexing site is between 2 and 8 nucleotides in length.
16. (Original) The method of claim 1, wherein the indexing site is about 6 nucleotides in length.
17. (Currently Amended) The method of claim 1, wherein the identifier site is ~~between 1 and 8~~ at least 10 nucleotides in length.
18. (Original) The method of claim 1, wherein the identifier site is about 8 nucleotides in length.
19. (Original) The method of claim 1, wherein the indexing primer binding site is a universal indexing primer binding site.
20. (Original) The method of claim 1, wherein the target sequence primer binding site is a universal target sequence primer binding site.
21. (Currently Amended) A kit comprising a plurality of double-stranded adaptors, wherein each adaptor comprises:
 - (i) an indexing primer binding site;
 - (ii) an indexing site;
 - (iii) an identifier site comprising at least 8 nucleotides; and

(iv) a target sequencing primer binding site;

wherein the kit further comprises one or more oligonucleotide primers, reagents for amplification, and instructions for employing the kit; wherein the plurality of adaptors, one or more oligonucleotide primers, and reagents for amplification are separately contained; wherein the indexing site is unique amongst a subset of the adaptors and is an index for multiple polynucleotides; and wherein the sequence of the identifier site is not conserved across all adaptor sequences.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In Re Application of: AMORESE, Douglas, Confirmation No.: 7076
et al.

Application No.: 14/540,917 Group Art Unit: 1637

Filed: November 13, 2014 Examiner: Teresa E. Strzelecka

FOR: **COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE
SEQUENCING READ**

Mail Stop Amendment

Commissioner for Patents

P.O. Box 1450

Alexandria, VA 22313-1450

AMENDMENT/RESPONSE TO OFFICE ACTION

In response to the Office Action dated May 20, 2015, this paper is timely filed within the 3-month period for reply ending August 20, 2015. Please amend the above-identified patent application in the following manner:

Amendments to the Claims are reflected in the listing of the claims which begins on page 2 of this paper.

Remarks/Arguments begin on page 6 of this paper.

REMARKS

I. Status

Claims 1-21 are pending; claims 4-6 and 8 have been withdrawn from consideration; and claims 1-3, 7, and 9-21 have been examined on the merits.

Claims 1, 17, and 21 have been amended. Support for the amendment of claim 1 can be found in the specification in at least paragraphs [0029], [0030], and paragraphs [0084] – [0091] which describe the random or semi-random nature of each of the identifier sequences/sites. Support for the amendment of claim 17 can be found in at least paragraph [0030]. Support for the amendment of claim 21 can be found in the specification in at least paragraphs [0028] – [0030], [0071], [0072], [0074], and [0084] – [0091].

No new matter is introduced by way of amendment. Accordingly, entry thereof is respectfully requested. Upon entry of this amendment, claims 1-3, 7, and 9-21 are under examination.

II. Restriction/Election Requirement

The Examiner is thanked for the rejoinder and consideration of claim 21, which is directed to a kit comprising adaptors of the presently claimed subject matter.

III. 35 U.S.C. § 101

Claim 21 is rejected as allegedly directed to a judicial exception. The Examiner asserts that the claim is drawn to a “plurality of adaptors, i.e., nucleic acid sequences, comprising an indexing primer binding site, an indexing site, an identifier site and a target sequencing binding site.” Office Action, p. 4. The Examiner then alleges that because no specific sequences are

claimed, “the claim is drawn to a plurality of oligonucleotides which have corresponding sequences occurring in nature.” *Id.* The Examiner also states that the claim does not “include additional elements that are sufficient to amount to significantly more than the judicial exception because no other structural elements are claimed.” Office Action, p. 5. Applicant respectfully traverses this rejection.

Solely for expediting prosecution, claim 21 has been amended to further recite oligonucleotide primers, reagents for amplification, instructions, and the limitations of the plurality of adaptors of instant claim 1. Instant claim 21 is drawn to a unique grouping of specific oligonucleotides. The plurality of adaptors claimed in the kit comprise an indexing primer binding site, an indexing site, an identifier site comprising at least 8 nucleotides, and a target sequencing primer binding site. The indexing site is unique amongst a subset of the adaptors and is an index for multiple polynucleotides, and the sequence of the identifier site is not conserved across all adaptor sequences. The instantly claimed plurality of adaptors cannot be equated to corresponding sequences occurring in nature. Furthermore, the amended claim now reads on a kit that cannot simply be interpreted as any composition of oligonucleotides. Accordingly, withdrawal of the rejection is respectfully requested.

V. 35 U.S.C. § 102(a)(1)

The Examiner has rejected claim 21 as allegedly being anticipated by both Grothues *et al.* (1993. *Nucl. Acids Res.* 21:1321-1322) and McCloskey *et al.* (2007. *Biochem. Genet.* 45:761-767). Applicant respectfully traverses this rejection.

The Examiner states Grothues *et al.* discloses a plurality of tagged random primers, comprising 9-15 bp random 3' component and a 17 bp constant sequences at the 5' end, and

alleges these primers are the same as the adaptors as claimed, and thus, anticipates the instant claims. Office Action, p. 5. Applicant respectfully disagrees. The instant claims are not anticipated by Grothues *et al.* as the primers disclosed by Grothues *et al.* are single stranded and the adaptors of the instant claims are double-stranded. In addition, Grothues *et al.* does not disclose, teach, or suggest instructions for employing the kit; separately contained adaptors, reagents for amplification, and oligonucleotide primers; the indexing site as unique amongst only a subset of the adaptors and as an index for multiple polynucleotides; and the identifier site as not being conserved across all adaptor sequences. Accordingly, Grothues *et al.* fails to teach each of the elements of instant claim 21 and does not anticipate claim 21.

The Examiner states McCloskey *et al.* discloses a plurality of randomly-barcode oligonucleotides containing, from the 3' end, a target-specific sequence, batch-stamp sequence, a random 7 bp barcode, another primer binding site, and another sequence, and alleges these oligonucleotides are the adaptors as claimed, and thus, anticipates the instant claims. Office Action, p. 6. Applicant respectfully disagrees. The instant claims are not anticipated by McCloskey *et al.* as McCloskey *et al.* does not disclose, teach, or suggest instructions for employing the kit; separately contained adaptors, reagents for amplification, and oligonucleotide primers; the indexing site as unique amongst only a subset of the adaptors and as an index for multiple polynucleotides; and the identifier site as comprising at least 8 nucleotides. Accordingly, McCloskey *et al.* fails to teach each of the elements of instant claim 21 and does not anticipate claim 21.

Withdrawal of the rejection is respectfully requested.

VI. 35 U.S.C. § 103

The Examiner has rejected claim 1-3, 7, and 9-20 as allegedly being unpatentable over Samuels *et al.* (US20120220494A1, filed 02/16/12 and published 08/30/12) and McCloskey *et al.* Applicant respectfully traverses this rejection.

The Examiner states “Samuels *et al.* teach a method for detecting a duplicate sequencing read from a population of sample sequence reads.” Office Action at p. 6. However, the Examiner also states “Samuels *et al.* teach detection of PCR-introduced errors in sequencing, but do not specifically teach detecting duplicate sequencing reads.” Office Action, p. 8 (emphasis added). Samuels *et al.* discloses a method of providing barcode sequences to targets such as a nucleic acid sequences, wherein universal priming sites may be present in barcode libraries [205], [0216]-[0216]. The barcode sequences and universal priming sites may be attached to the nucleic acid sequences and amplified in a PCR reaction [0288]-[0289]. The amplification products may then be sequenced as a whole or in part [0326-0327]. Sequencing errors versus single nucleotide polymorphisms may be detected through the use of barcodes and sequencing of the amplification products [0339] – [0341].

Samuels *et al.* do not teach (1) the inclusion of an identifier site comprising at least 8 nucleotides, wherein the identifier site is not conserved across all adaptor sequences; (2) the indexing site as unique amongst only a subset of the nucleic acid fragments; (3) a target sequence primer binding site; and (3) the detection and removal of duplicate sequencing reads.

McCloskey *et al.* disclose a method of identifying and removing duplicate reads through incorporating an adaptor comprising a batch stamp, a barcode, a primer-binding site for PCR, and extra bases that serve to minimize nonspecific annealing, wherein the barcodes comprise 7 random nucleotides (p. 763, Fig. 1; p. 764, para 2; p. 766, para 4-5).

McCloskey *et al.* does not teach the method wherein the identifier site (barcode) comprises at least 8 nucleotides or a target sequence primer binding site.

The Examiner alleges McCloskey *et al.* accounts for the deficiencies of Samuels *et al.* Applicant respectfully disagrees. Samuels *et al.* and McCloskey *et al.*, alone or in combination fail to teach an indexing site that is at least 8 nucleotides in length, as recited in the instant claims. McCloskey *et al.* only discloses barcodes that are 7 nucleotides in length, which limit the total number of adaptor sequences comprising unique barcodes that can be generated. The instant claims recite indexing sites of at least 8 nucleotides in length, which allows for a significantly greater population of adaptor sequences comprising unique barcodes that can be generated.

Withdrawal of the rejection is respectfully requested.

CONCLUSION

In view of the foregoing, Applicant respectfully submits that no further impediments exist to the allowance of this application and, therefore, requests an indication of allowability. However, the Examiner is requested to call the undersigned if any questions or comments arise.

The Director is hereby authorized to charge any appropriate fees under 37 C.F.R. §§1.16, 1.17, and 1.21 that may be required by this paper, and to credit any overpayment, to Deposit Account No. 50-1283.

Dated: August 19, 2015

Respectfully submitted,
COOLEY LLP

COOLEY LLP
ATTN: Patent Group
1299 Pennsylvania Avenue NW, Suite 700
Washington, DC 20004

By: /Jeffrey R. Bolland/
Jeffrey R. Bolland, Ph.D.
Reg. No. 73,759

Tel: (202) 728-7017
Fax: (202) 842-7899

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PATENT APPLICATION FEE DETERMINATION RECORD Substitute for Form PTO-875					Application or Docket Number 14/540,917		Filing Date 11/13/2014		☐ To be Mailed	
---	--	--	--	--	---	--	----------------------------------	--	---------------------------------------	--

ENTITY: ☒ LARGE ☐ SMALL ☐ MICRO									
APPLICATION AS FILED – PART I									
		(Column 1)		(Column 2)					
FOR		NUMBER FILED		NUMBER EXTRA		RATE (\$)		FEE (\$)	
☐ BASIC FEE (37 CFR 1.16(a), (b), or (c))		N/A		N/A		N/A			
☐ SEARCH FEE (37 CFR 1.16(k), (l), or (m))		N/A		N/A		N/A			
☐ EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))		N/A		N/A		N/A			
TOTAL CLAIMS (37 CFR 1.16(i))		minus 20 =		*		X \$ =			
INDEPENDENT CLAIMS (37 CFR 1.16(h))		minus 3 =		*		X \$ =			
☐ APPLICATION SIZE FEE (37 CFR 1.16(s))		If the specification and drawings exceed 100 sheets of paper, the application size fee due is \$310 (\$155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).							
☐ MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))									
* If the difference in column 1 is less than zero, enter "0" in column 2.						TOTAL			

APPLICATION AS AMENDED – PART II										
		(Column 1)		(Column 2)		(Column 3)				
AMENDMENT	08/19/2015		CLAIMS REMAINING AFTER AMENDMENT				HIGHEST NUMBER PREVIOUSLY PAID FOR		PRESENT EXTRA	
	Total (37 CFR 1.16(i))		* 21		Minus		** 21		= 0	
	Independent (37 CFR 1.16(h))		* 2		Minus		*** 3		= 0	
	☐ Application Size Fee (37 CFR 1.16(s))									
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))									
						TOTAL ADD'L FEE		0		

		(Column 1)		(Column 2)		(Column 3)				
AMENDMENT			CLAIMS REMAINING AFTER AMENDMENT				HIGHEST NUMBER PREVIOUSLY PAID FOR		PRESENT EXTRA	
	Total (37 CFR 1.16(i))		*		Minus		**		=	
	Independent (37 CFR 1.16(h))		*		Minus		***		=	
	☐ Application Size Fee (37 CFR 1.16(s))									
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))									
						TOTAL ADD'L FEE				

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.
 ** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".
 *** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".
 The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1.

This collection of information is required by 37 CFR 1.16. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**
 If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)	Application Number		14540917	
	Filing Date		2014-11-13	
	First Named Inventor	AMORESE, Douglas		
	Art Unit	1637		
	Examiner Name	Teresa E. Strzelecka		
	Attorney Docket Number	NUGN-001/01US 301735-2003		

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /TS/

U.S.PATENTS								
Examiner Initial*	Cite No	Patent Number	Kind Code ¹	Issue Date	Name of Patentee or Applicant of cited Document	Pages, Columns, Lines where Relevant Passages or Relevant Figures Appear		
	1							
If you wish to add additional U.S. Patent citation information please click the Add button.								
U.S.PATENT APPLICATION PUBLICATIONS								
Examiner Initial*	Cite No	Publication Number	Kind Code ¹	Publication Date	Name of Patentee or Applicant of cited Document	Pages, Columns, Lines where Relevant Passages or Relevant Figures Appear		
	1	20120208705	A1	2012-08-16	STEEMERS et al.			
If you wish to add additional U.S. Published Application citation information please click the Add button.								
FOREIGN PATENT DOCUMENTS								
Examiner Initial*	Cite No	Foreign Document Number ³	Country Code ²	Kind Code ⁴	Publication Date	Name of Patentee or Applicant of cited Document	Pages, Columns, Lines where Relevant Passages or Relevant Figures Appear	T ⁵
	1	2008093098	WO	A1	2008-08-07	SOLEXA LIMITED		☐
	2	2012061832	WO	A1	2012-05-10	ILLUMINA, INC.		☐
	3	2014144092	WO	A1	2014-09-18	NUGEN TECHNOLOGIES, INC.		☐

**INFORMATION DISCLOSURE
STATEMENT BY APPLICANT**
(Not for submission under 37 CFR 1.99)

Application Number	14540917
Filing Date	2014-11-13
First Named Inventor	AMORESE, Douglas
Art Unit	1637
Examiner Name	Teresa E. Strzelecka
Attorney Docket Number	NUGN-001/01US 301735-2003

If you wish to add additional Foreign Patent Document citation information please click the Add button

NON-PATENT LITERATURE DOCUMENTS

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T ⁵
	1	MYERS, R. M. Myers lab microRNA-seq Protocol, Hudson Alpha Institute for Biotechnology, May 2, 2013. p. 1-15. Retrieved from ,https://www.encodeproject.org/documents/9fb8e06c-d145-447a-a606-e7f3dc44b549/@download/attachment/HA_Myers_miRNA-seq_Protocol_2013-05-02.pdf> on February 28, 2015.	☐
	2	International Searching Authority. Written Opinion and International Search Report for PCT/US2014/065530. Date of mailing is March 5, 2015.	☐

If you wish to add additional non-patent literature document citation information please click the Add button

EXAMINER SIGNATURE

Examiner Signature	/Teresa Strzelecka/	Date Considered	09/21/2015
--------------------	---------------------	-----------------	------------

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

¹ See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. ² Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). ³ For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. ⁴ Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. ⁵ Applicant is to place a check mark here if English language translation is attached.



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
14/540,917	11/13/2014	Douglas AMORESE	NUGN-001/01US	7076
58249 7590 09/28/2015 COOLEY LLP ATTN: Patent Group 1299 Pennsylvania Avenue, NW Suite 700 Washington, DC 20004			EXAMINER STRZELECKA, TERESA E	
			ART UNIT 1637	PAPER NUMBER
			NOTIFICATION DATE 09/28/2015	DELIVERY MODE ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

zpatcdocketing@cooley.com

Office Action Summary	Application No. 14/540,917	Applicant(s) AMORESE ET AL.	
	Examiner TERESA E. STRZELECKA	Art Unit 1637	AIA (First Inventor to File) Status Yes

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1) ☒ Responsive to communication(s) filed on 8/19/15.
☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on ____.

2a) ☒ This action is **FINAL**. 2b) ☐ This action is non-final.

3) ☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on ____; the restriction requirement and election have been incorporated into this action.

4) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims*

5) ☒ Claim(s) 1-21 is/are pending in the application.
5a) Of the above claim(s) ____ is/are withdrawn from consideration.

6) ☐ Claim(s) ____ is/are allowed.

7) ☒ Claim(s) 1-21 is/are rejected.

8) ☐ Claim(s) ____ is/are objected to.

9) ☐ Claim(s) ____ are subject to restriction and/or election requirement.

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to PPHfeedback@uspto.gov.

Application Papers

10) ☐ The specification is objected to by the Examiner.

11) ☐ The drawing(s) filed on ____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a) ☐ All b) ☐ Some** c) ☐ None of the:

1. ☐ Certified copies of the priority documents have been received.

2. ☐ Certified copies of the priority documents have been received in Application No. ____.

3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)

2) ☒ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b)
Paper No(s)/Mail Date 8/19/15.

3) ☐ Interview Summary (PTO-413)
Paper No(s)/Mail Date. ____.

4) ☐ Other: ____.