CTCTACTAAAAATACAAAAAAATTAGCCGGGCGTGGTAGTGAGCACCTAGATCGGAAGAG
CTCGTATGCCGTCTTCTGCTTG

Clone 14 (SEQ ID NO:18)

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGC
AATTGGTAAAACAGTAAGCAATGAAACAGACACTTCTCAAATATTCCAAGATGGTACACG
CTTTTCAGTGTGTATGATCCAATAAAGCCATTGGAAGTAGGCTTTAATAGTCAAAAAAGA
CTATTCAGTTAGATAGGAACTATTTGCCTATAACTATTGGCCAAAAATAGGTTAAAAAAT
TGTTTTAAATTTGTGCTTTACAAAACATGTGGACTTTTTTAGAAAATGTGTCAAATTTCA
AAAGAAATATAGACATTATGGAAAGGTCAGTTAAGCACAGCCCTAATCCTGAAAACATAA
CTATGAAAGATACTAGCTGTTACTTGTAACCAAAAGGAAAAAAAAGATATTAGTAACCAA
TAATTAGCAAACAATGCCCATATATTTCCTTTTTTTTTTTTTTTTTTGAGACAGGGGCT
TACTCAGGCTGGATGTGATCATGAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG

Clone 15 (SEQ ID NO:19)

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTT
AGAAGCTCTATTATACTGGAAAAGAGATATGAGACCCTTCCTACTTTAAGAATCAATGAA
GCCGGGTGTGGTGGCTCACGCCTGTACTCCCAGCATTTTGAGAGGCCAAGCTGGGCAGAT
CACCTGAGGTCGGGAGTTTGAGACCAGCCTAGCCAACATGATGAAATCCTGTTTCTACTA
ATAACACAAAAATTAGCCGGGTGTGGTGGCGCACATCTGGAATCCCAGCTACTCCAGAGG
CTGAGGCAGGAAAATTGCTTGAAGCTGGGAAGCAGAAGTTGCAGTGAAGATCGGAAGAGC
TCGTATGCCGTCTTCTGCTTG

Clone 16 (SEQ ID NO:20)

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGG
CTGGACTGAATAGGATAGCCTTAGCTGTAAAATTGGGCTGATCTTTCAAATGGACTCATG
CTTGCCGAATGACTCACGCTCCTGTTTACAAATCAGCTCTGTGAAGAAATGCAGAGTGGG
AGGCTCTGCTTGCCAGACGGAGACCTTAGACCTCCAGGGGCGGAGAACGGAGTACTTCCT
CTGGTGCTCGGCTTCCCTTCCTGGGGGCAGATCTCTCAGCTTCTGGTTGGTGGCTCTCAA
AATCCAGACACAAGGTCAGCTGCAGCCAGCGTGGGCCCTGGAGTAGCTCCAGTTATGGGG
CAGCAATGGCCCCCTCTCATTTTGAGAGCTCACTTTGCCTGTGGATGGTTTTAATCCATC
TGGATAAACTTGAGGCCCATGGGAATACCATATACTATGGTAACCATGTACACTGCTCTA

AAGATGTGGCTGCTGTTGTATAACTTTTTCCTTTATTTTTGTCAATTTCCTATTTTCCAG
AGTCTTGCATACCCACTATGTCTACTGTGATAGTGAACGTAAAAACATACAAGATGTTGG
TGTTATCCTCAATCTCAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG

Clone 17 (SEQ ID NO:21)
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCT
GTGAGAAATCAATGTCTGCTGTTTATAAGCCGCCGGGCTGTGATATCCTGTGAGAGTGGC
CCCAGTGGATGAAGACAGATGCTCTCAAGGAGCGCAGATGACGCGGGTTCCGAAGGACTC
GGCACCCAGCCCGGAGGCCGGCAACATGGGCAAGGGGCCTCTCACGGCTGACCTGTTTCC
TCATCAGCACATCAGGACAATAAGAGCTCCCACTTCACAGGTGGTGAAGAGCCAACGTGG
TGAAGAATGAATAAAGCAGCTCGTGGAAAGTGCTGTGCATGAGGCCTGGCAACCGGTCCC
TGCTCTGAGGTCACCTGCCACGGAGCTGCTGACAGGACCATTAAAAACACAATTGTGCAA
GTGCTCACCCACATTCACAGCAGCAGAATCTCCACCAGCCAAGCATTGGAGACGATCCTT
GCATCCATAGACATGAACAGATGAGCAAAACGTGGTCTATACGGACGATGAAATAGCACT
CAGCCCTAAGAAGAAATAAAATCCCGACAGAGAGATCGGAAGAGCTCGTATGCCGTCTTC
TGCTT

Clone 18 (SEQ ID NO:22)
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGC
TGAGAGGAGAGGAGAGGCTGGCAGGGGGCTGATGCAGGAGGTGATAGGGCTCCCGGTGAT
AAGAGGTGAGAAGAACAGTCTCTGTGTGCCTAGAGAAGAGATTACCAGAAGTCTGCTATC
TGTTTGTTCGCGGATGTCGGACAGGCAGGATCGGTGATGGCAGGTCTTGGGGGAAGGATT
ATCAGGAGCTAAAAGCTGTCTTCACCTTGGCTGCTAAGAACTCATCTCGGATCTTCTTAG
AATTCCAAATCGGACTTTTCTCCTAGCAGTGGCTACATCCTTAACCTCAAAAATACCCGT
ATTAGCAGATCTACCTCCATGAAATAGACAATTCTTGACAAACTAAGATCGGAAGAGCTC
GTATGCCGTCTTCTGCTTG

Clone 19 (SEQ ID NO:23)
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGT
TATTTTTCATACTGATGATTCTAGGGCGGTTCCCTGGACTTGGCTGGGACCAGAACTGTC
TCACAGCTGCCAGCGCTCCCCACCCCCATGAGAACCATCGTGCTGAAGATGAAGAGCCAA
AGCCAAACTGACCAGCCTCTGCAACCCCAAGCATCACCGTGTGAGGACAGCGTGATGTGG

AGCCGCGGAGACCCCACGCAGTGCTGCGAGGGGCAACCACATGCAGGGGGGCAGAGGTGT
GGGGGAGCAAGCAGATGCCACCGGCATGCCTAGGAGCCACAGGGAGCATGCGGGCTGGGC
AGTGATGGGAATGAACGTGAACTCCAGTCCTGCCCCAAGAAGTCCTCCGGCCCCTCCTTC
TCAATTCAGGGCACAAAGTGGTAACTGCAGCATGCAGAGGAGTCGAGGAGTCTTCCCTCG
TCCCAGGAGCAGCACCTCGGGCACAGTCTCGGTCCCACAGAACAGCCAAGTGTGGGTTGG
TGGTCTAGAGACCTCCGAAGATCCAGTGGGGGAAGGATGGGCAGCAGAGGGTCTACTCTC
TGAAAATAAGGGGAAGGGATTTTCCCTCCCCACTGCCAAGGTCCCAGCTACTGGACGTGG
GTGGAGATCGGAAGAGCTCGTATGCCGTCTTCTGCCTT

These results confirm that the library preparation method produces a library of "sequenceable" DNA templates containing a mixture of fragments of different sequence. The insert DNA from each of the 19 clones sequenced was found to align to a human genome reference sequence (alignment not shown), illustrating that the method produces a library of templates which truly reflect the sequence composition of the starting target fragments (i.e. the clones contain human genomic fragments rather than "junk" sequence).

Claims:

1. A method of generating a library of template polynucleotide molecules which have common sequences at their 5' ends and common sequences at their 3' ends, the method comprising:

ligating identical mismatched adapter polynucleotides to both ends of each of one or more target polynucleotide duplexes to form one or more adapter-target constructs, wherein each mismatched adapter is formed from two annealed polynucleotide strands that form a bimolecular complex comprising at least one double-stranded region and an unmatched region, and

carrying out an initial primer extension reaction in which a primer oligonucleotide is annealed to an adapter portion of each of the adapter-target constructs and extended by sequential addition of nucleotides to form extension products complementary to at least one strand of each of the adapter-target constructs,

wherein the extension products, and optionally amplification products derived therefrom, collectively provide a library of template polynucleotide molecules which have common sequences at their 5' ends and common sequences at their 3' ends.

2. A method according to claim 1 wherein the initial primer extension reaction comprises annealing an oligonucleotide primer to an adapter portion of each strand of each individual adapter-target construct and extending

the primers to form extension products complementary to each strand of each individual adapter-target construct.

3. A method according to claim 1 or claim 2 wherein the initial primer extension reaction comprises:
a) annealing a primer oligonucleotide to an adapter portion of each of the adapter-target constructs,
b) extending the primers by sequential addition of nucleotides to form extension products complementary to at least one strand of each of the adapter-target constructs, and
c) subjecting the products obtained in step b) to denaturing conditions, thereby separating the extension products from strands of the adapter-target constructs.

4. A method according to claim 3 wherein steps a) to c) are carried out once and the extension products collected to provide a library of template polynucleotide molecules which have common sequences at their 5' ends and common sequences at their 3' ends.

5. A method according to claim 3 wherein steps a) to c) are repeated one or more times and all the extension collected to provide a library of template polynucleotide molecules which have common sequences at their 5' ends and common sequences at their 3' ends.

6. A method according to any one of claims 1 to 5 wherein the primers used in the initial primer extension reaction are annealed to an unmatched region of the adapters.

7. A method according to any one of claims 1 to 3 wherein the initial primer extension reaction is carried out as part of a polymerase chain reaction and the amplification products of the PCR reaction are collected to provide a library of template polynucleotide molecules which have common sequences at their 5' and 3' ends.

8. A method according to claim 7 wherein the polymerase chain reaction is carried out using first oligonucleotide primers capable of annealing to adapter portions of each of the adapter-target constructs and second oligonucleotide primers capable of annealing to a region of the extended strands produced by extension of the first oligonucleotide primers, this region being complementary to an adapter portion of the adapter-target constructs.

9. A method according to claim 8 wherein the first and second oligonucleotide primers have different nucleotide sequences.

10. A method according to claim 9 wherein the first oligonucleotide primers are capable of annealing to the unmatched region in a first strand of the adapter-target constructs and the second oligonucleotide primers are capable of annealing to an extension product complementary to the unmatched region in a second strand of the adapter-target constructs.

11. A method according to claim 8 wherein the first and second oligonucleotide primers have substantially identical nucleotide sequences.

12. A method according to claim 11 wherein the first and second oligonucleotide primers are capable of annealing to a double-stranded region in the adapter portions of the adapter-target constructs.

13. A method according to any one of claims 1 to 12 wherein the mismatched adapter polynucleotides are forked adapters formed by annealing of partially complementary first and second polynucleotide strands, wherein a sequence of 5 or more consecutive nucleotides at 3′ end of the first strand is complementary to a sequence of 5 or more consecutive nucleotides the 5′ end of the second strand such that a double-stranded region of 5 or more consecutive base pairs is formed by annealing of the two strands and wherein a sequence of at least 10 consecutive nucleotides at the 5′ end of the first strand and a sequence of at least 10 consecutive nucleotides at the 3′ end of the second strand are not complementary such that an unmatched region of at least 10 consecutive nucleotides on each strand remains in single stranded form when the double-stranded region is annealed.

14. A method according to claim 13 wherein the double-stranded region formed when the two strands are annealed is from 5 to 20 consecutive base pairs in length.

15. A method according to claim 13 or claim 14 wherein the unmatched region consists of from 10 to 50 consecutive unpaired nucleotides on each strand.

16. A method according to any one of claims 1 to 15 wherein the target polynucleotide duplexes are fragments of genomic DNA.

17. A method according to claim 16 wherein the target polynucleotide duplexes are fragments of a whole genome.

18. A method according to any one of claims 1 to 15 wherein the target polynucleotide duplexes are double-stranded cDNAs.

19. A method of solid-phase nucleic acid amplification of template polynucleotide molecules which comprises: preparing a library of template polynucleotide molecules which have common sequences at their 5' ends and common sequences at their 3' ends using the method according to any one of claims 1 to 18 and carrying out a solid-phase nucleic acid amplification reaction wherein said template polynucleotide molecules are amplified.

20. A method according to claim 19 wherein said template polynucleotide molecules are amplified by solid-phase PCR using forward and reverse amplification primers.

21. A method according to claim 20 wherein said library of template polynucleotide molecules is prepared using the method of any one of claims 7 to 12 and solid-phase PCR is carried out using primers identical those used in preparing the library of template polynucleotide molecules.

22. A method according to claim 20 wherein said library of template polynucleotide molecules is prepared using the

method of any one of claims 7 to 12 and solid-phase PCR is carried out using primers which are not identical to those used in preparing the library of template polynucleotide molecules.

23. A method according to claim 20 wherein said library of template polynucleotide molecules is prepared using the method of any one of claims 1 to 6 and solid-phase PCR is carried out using forward amplification primers capable of annealing to an adapter portion of the extension products formed in the initial primer extension reaction and reverse primers which are capable of annealing to extension products formed by extension of forward amplification primers annealed to the extension products formed in the initial primer extension reaction.

24. Use of a library of template polynucleotide molecules prepared according to the method of any one of claims 1 to 18 as a template for solid-phase PCR amplification.

25. Use of a library of template polynucleotide molecules prepared according to the method of claim 17 as a template for whole genome amplification.

26. A method of whole genome amplification which comprises: preparing a library of template polynucleotide molecules which have common sequences at their 5' ends and common sequences at their 3' ends using the method according to claim 17 and carrying out a nucleic acid amplification reaction wherein said template polynucleotide molecules are amplified.

27. A kit for use in preparing a library of template polynucleotide molecules which have common sequences at their 5' ends and common sequences at their 3' ends, the kit comprising mismatched adapter polynucleotides as defined in any one of claims 1 to 18 and oligonucleotide primers capable of annealing to the mismatched adapter polynucleotides.

28. A kit according to claim 27 wherein the oligonucleotide primers are capable of annealing to an unmatched region of the mismatched adapter polynucleotides.

29. A kit according to claim 28 wherein the mismatched adapter polynucleotides are forked adapters as defined in any one of claims 13 to 15.

Fig. 1

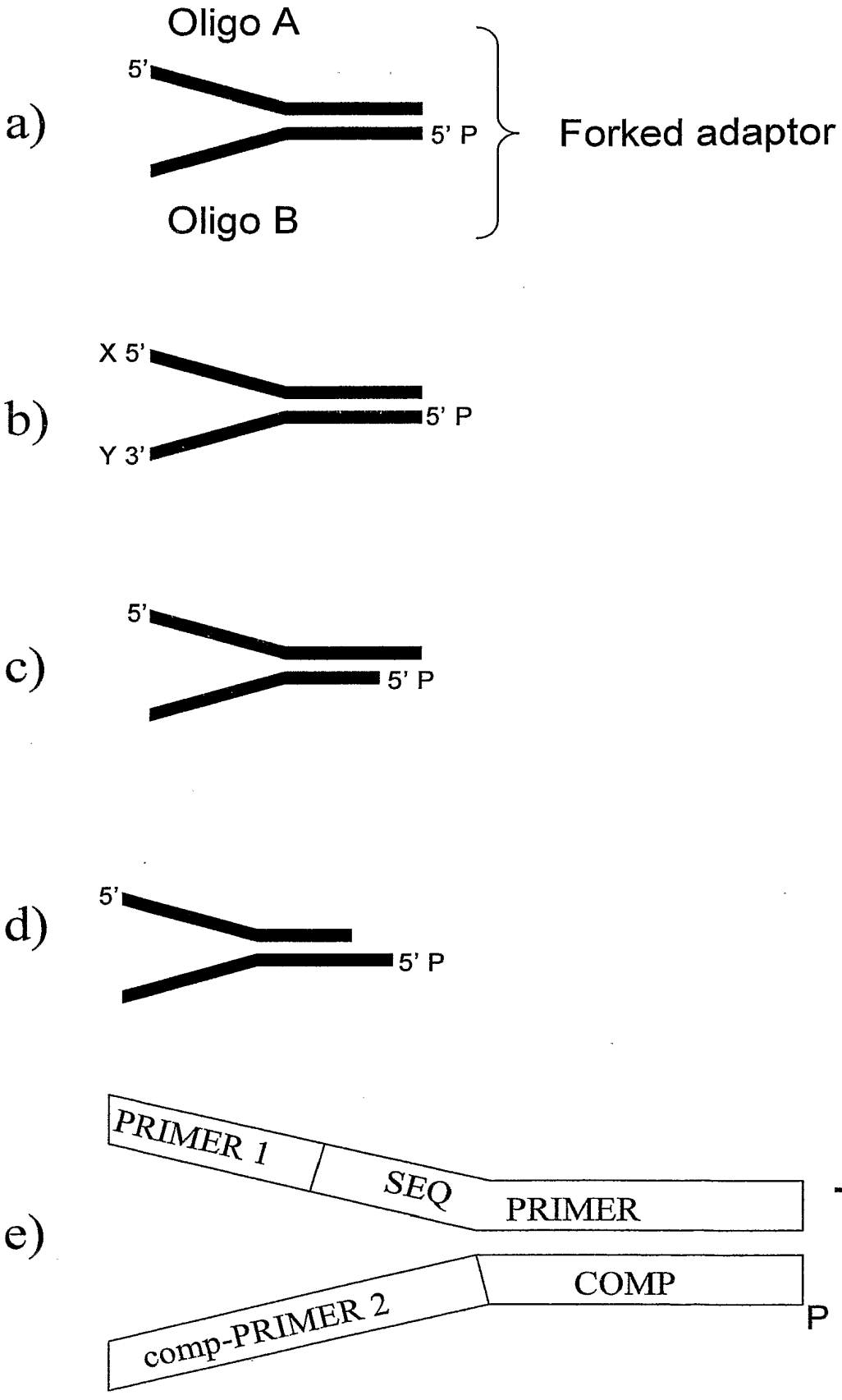

a)
Oligo A
5'
5' P
Oligo B
Forked adaptor
b)
X 5'
5' P
Y 3'
c)
5'
5' P
d)
5'
5' P
e)
PRIMER 1
SEQ
PRIMER
T
COMP
comp-PRIMER 2
P

Fig. 2a

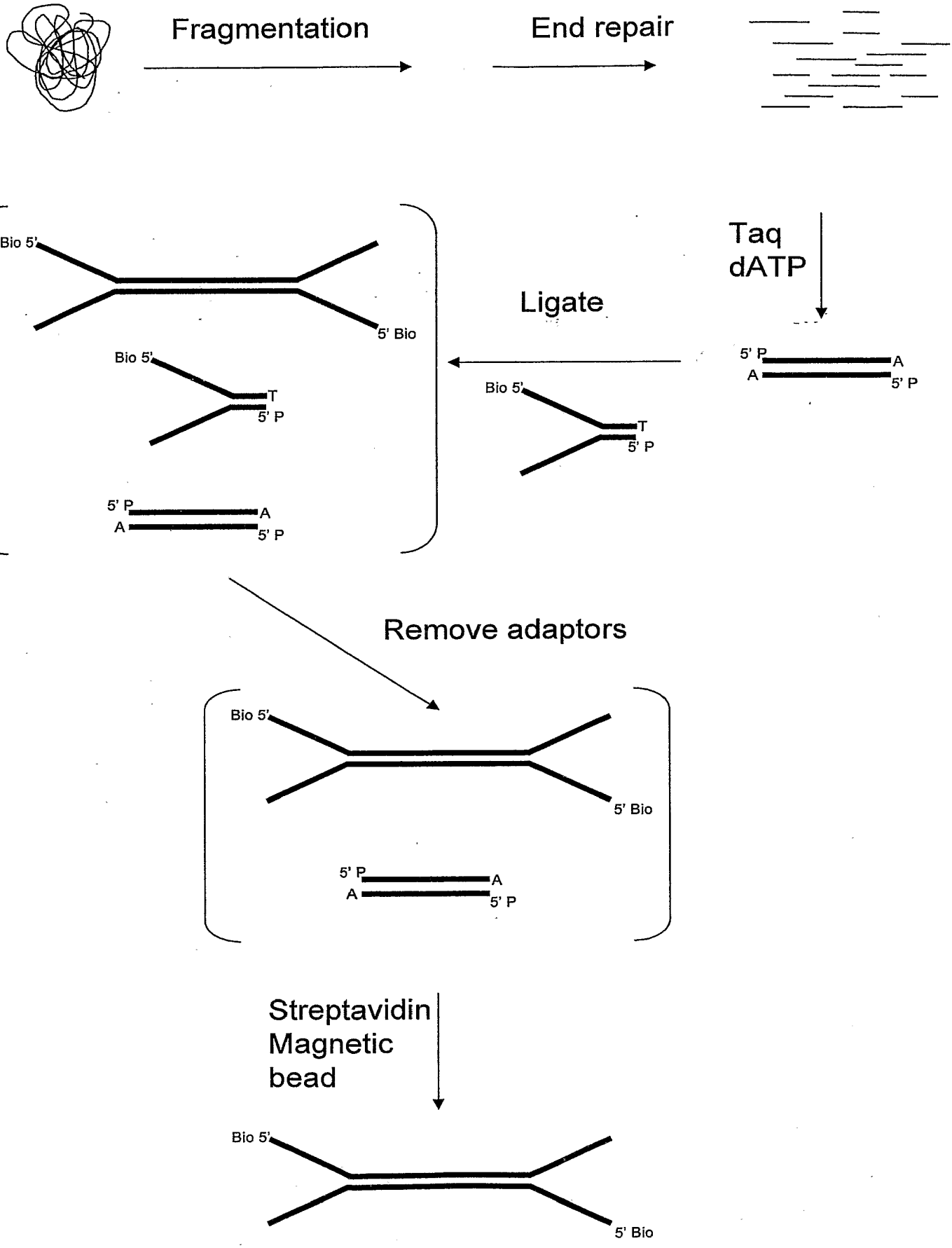

Fragmentation
End repair
Taq
dATP
Ligate
Bio 5'
5' Bio
T
5' P
A
Remove adaptors
Streptavidin
Magnetic
bead

Fig. 2b

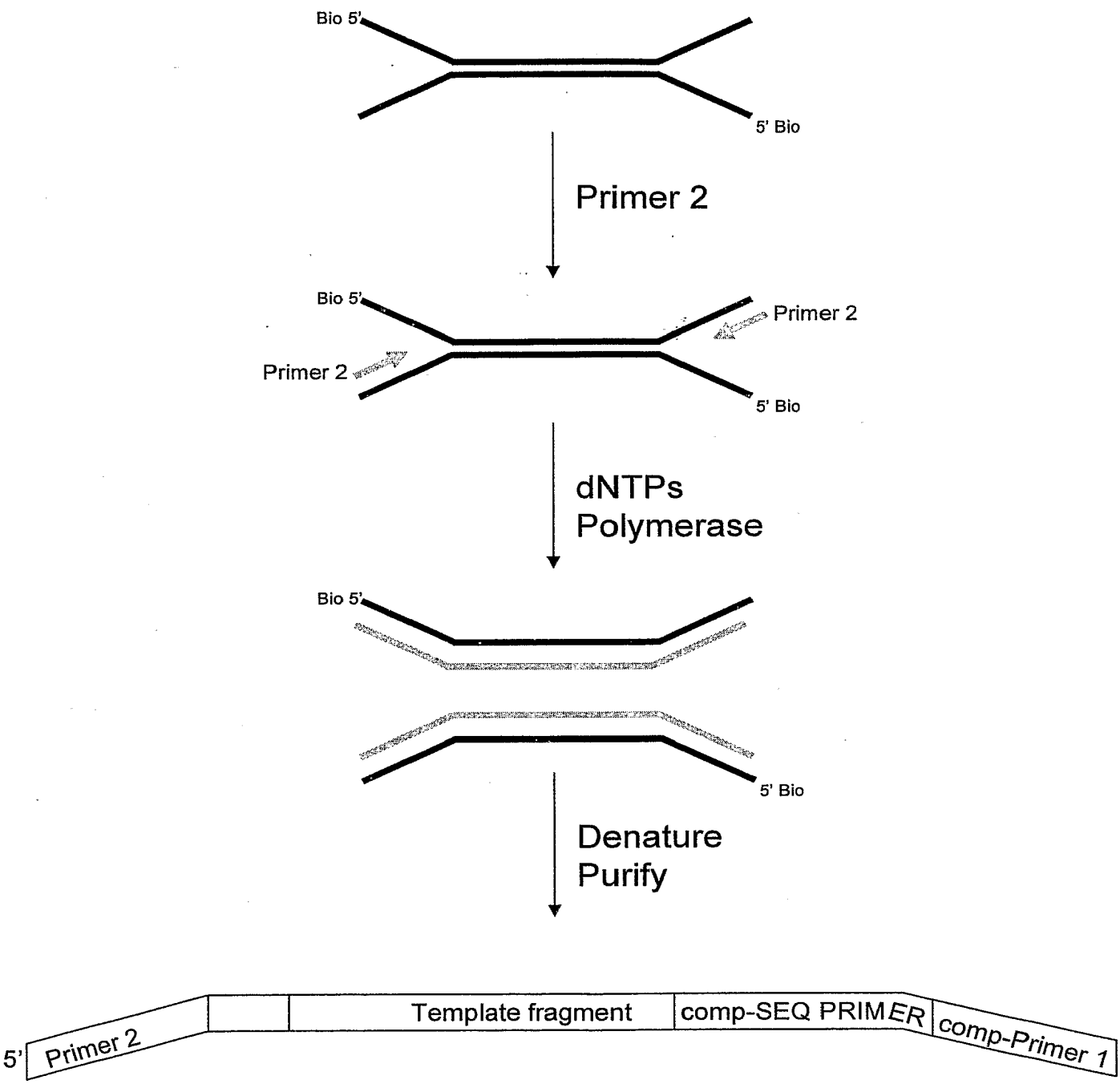

Bio 5'
5' Bio
Primer 2
Bio 5'
Primer 2
Primer 2
5' Bio
dNTPs
Polymerase
Bio 5'
5' Bio
Denature
Purify
5'
Primer 2
Template fragment
comp-SEQ PRIMER
comp-Primer 1

Fig. 3

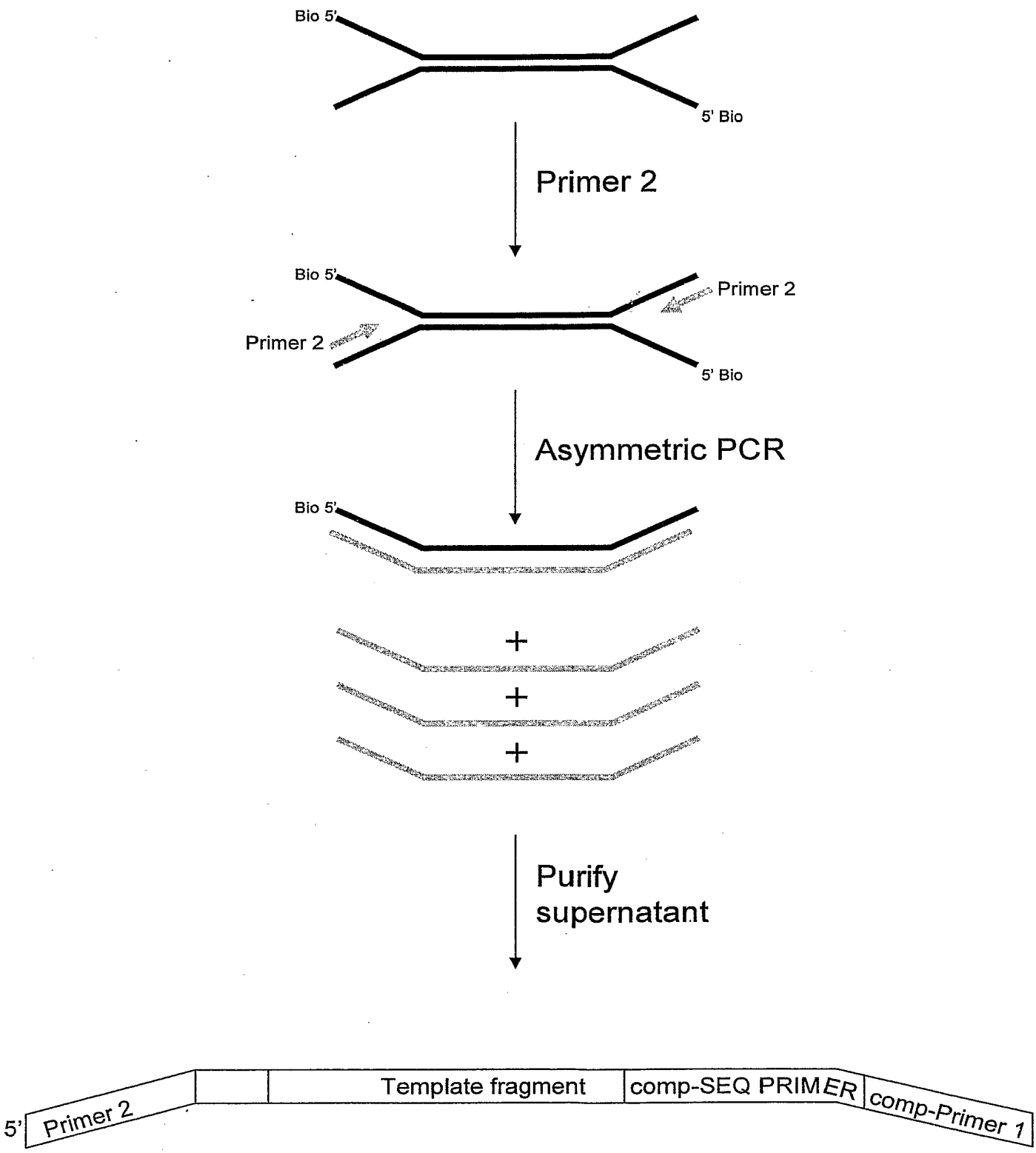
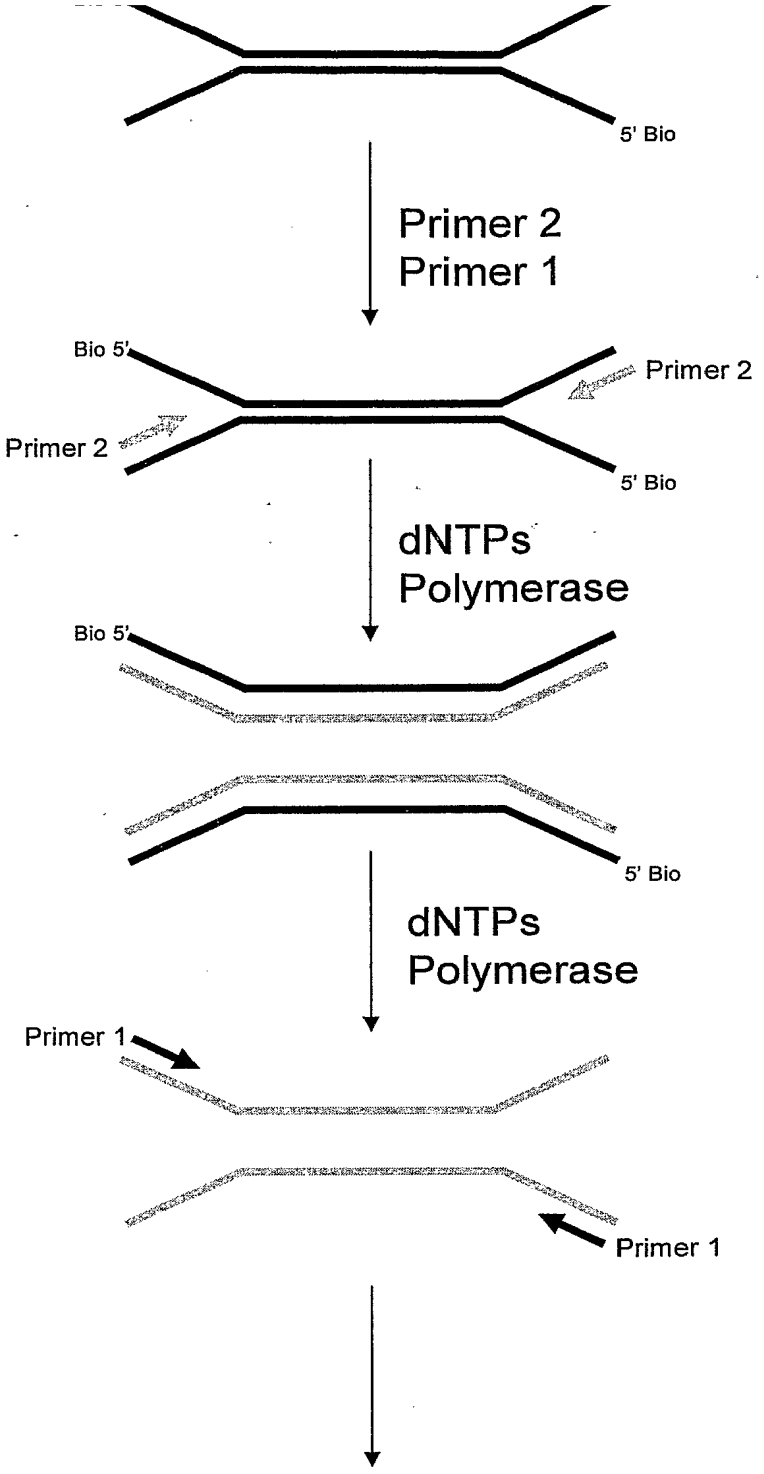
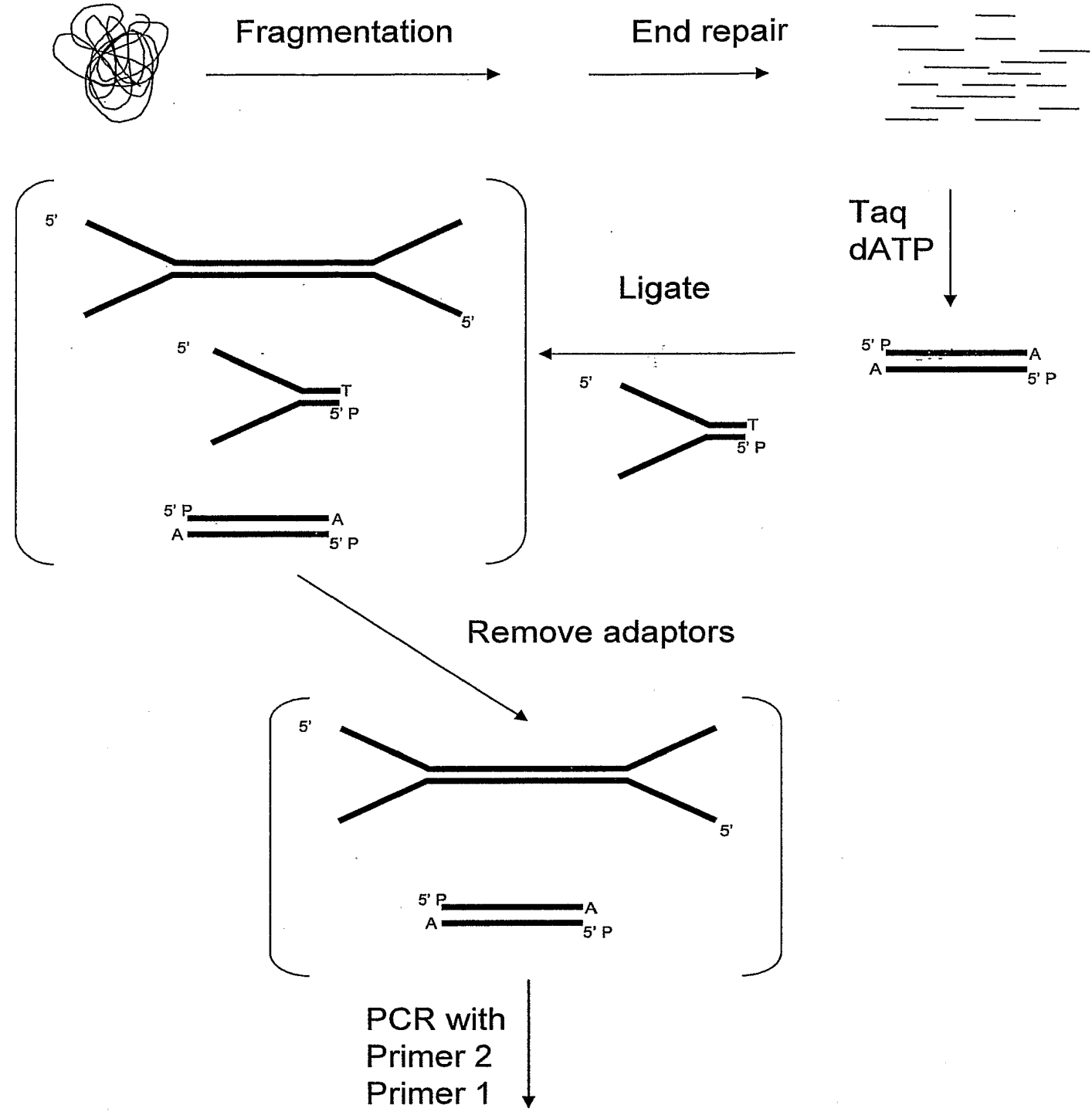

Bio 5'
5' Bio
Primer 2
Bio 5'
Primer 2
Primer 2
5' Bio
Asymmetric PCR
Bio 5'
+
+
+
Purify
supernatant
5'
Primer 2
Template fragment
comp-SEQ PRIMER
comp-Primer 1

Fig. 4



5'					
	Primer 1	SEQ Primer	Template fragment		comp-Primer 2
	comp-Primer 1	comp-SEQ Primer	comp-Template fragment		Primer 2

Fig. 5



5'					
	Primer 1	SEQ Primer	Template fragment		comp-Primer 2
	comp-Primer 1	comp-SEQ Primer	comp-Template fragment		Primer 2

Fig. 6

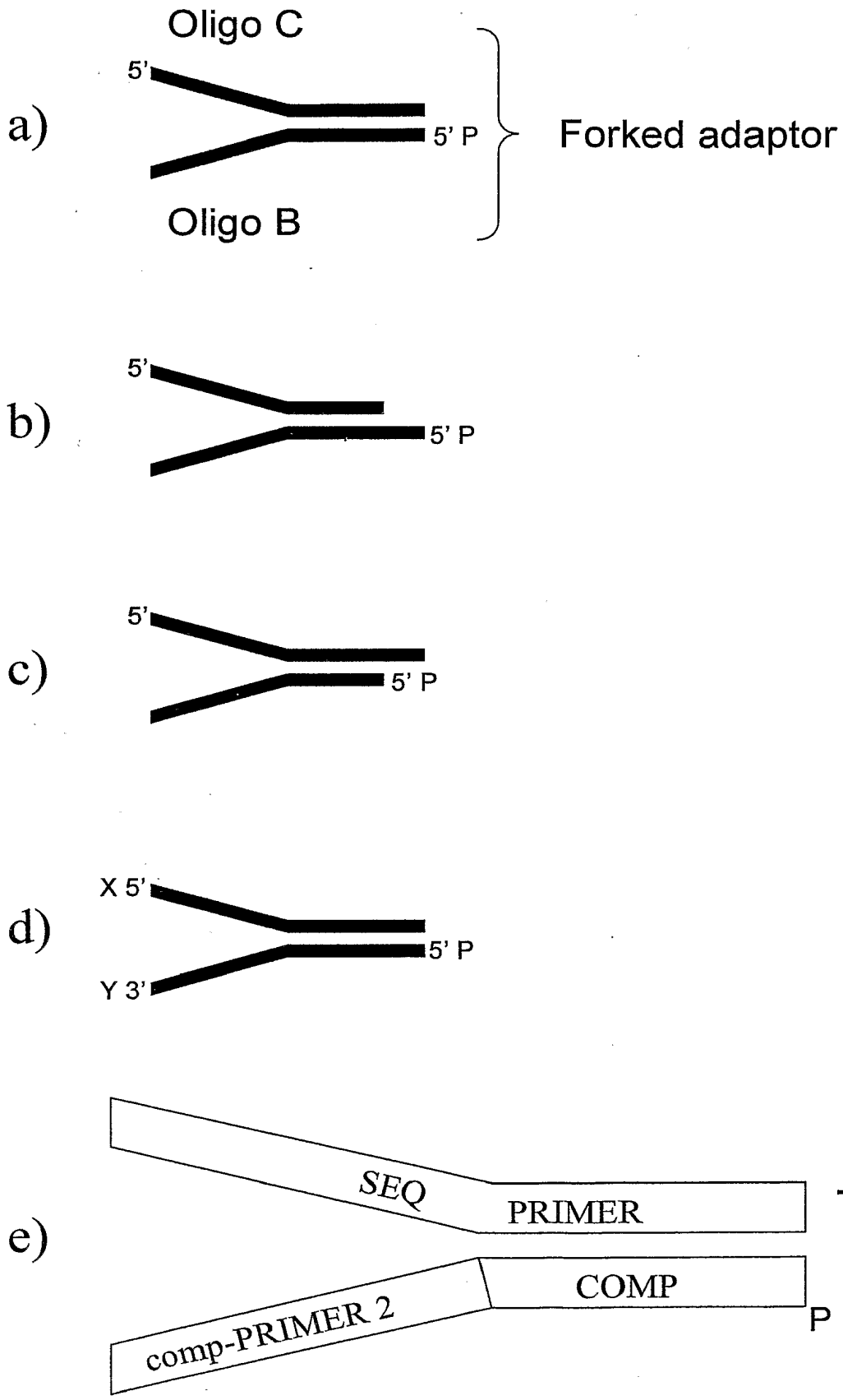

a)
Oligo C
5'
5' P
Oligo B
Forked adaptor
b)
5'
5' P
c)
5'
5' P
d)
X 5'
5' P
Y 3'
e)
SEQ
PRIMER
T
COMP
comp-PRIMER 2
P

Fig. 7a

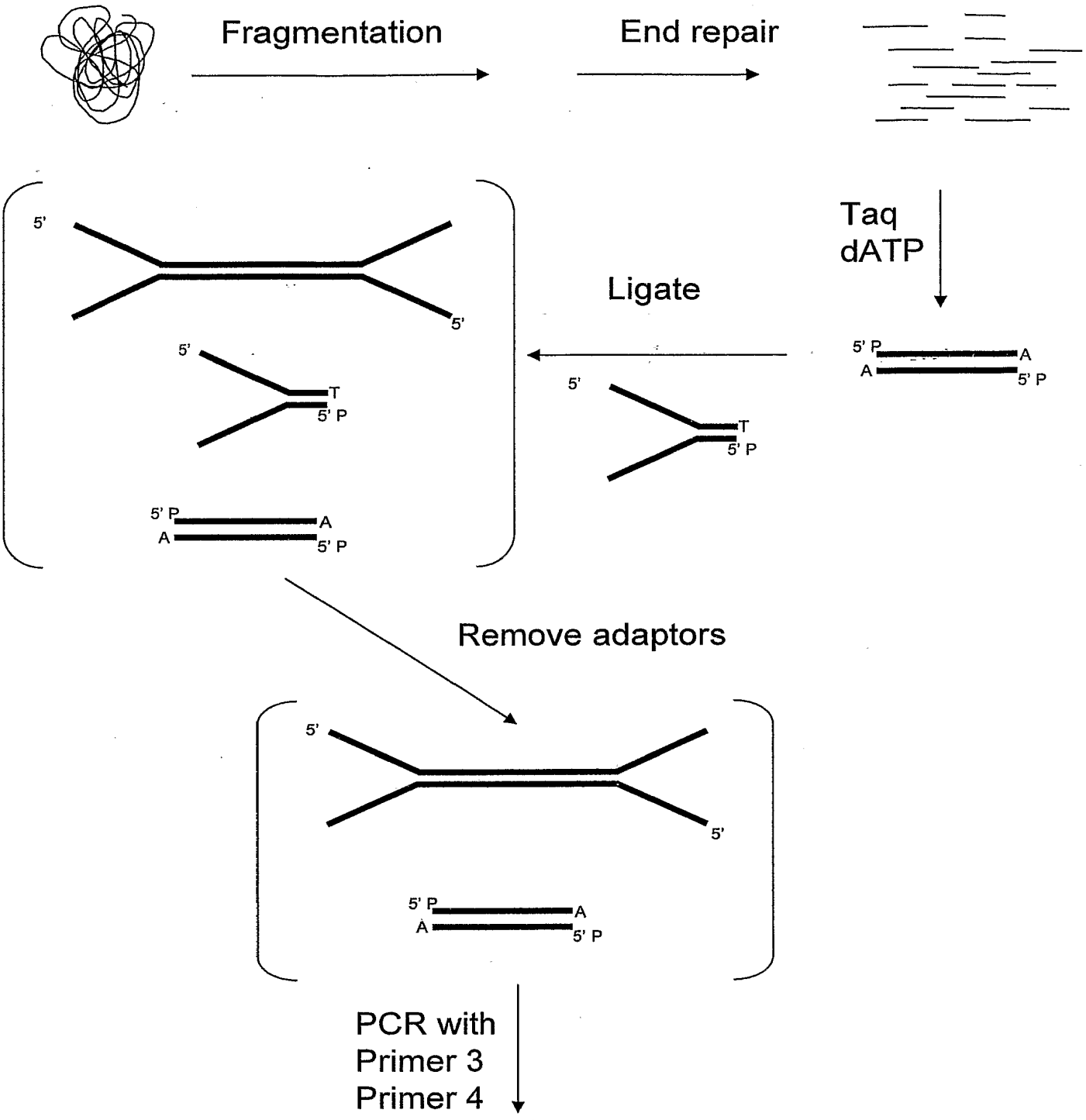
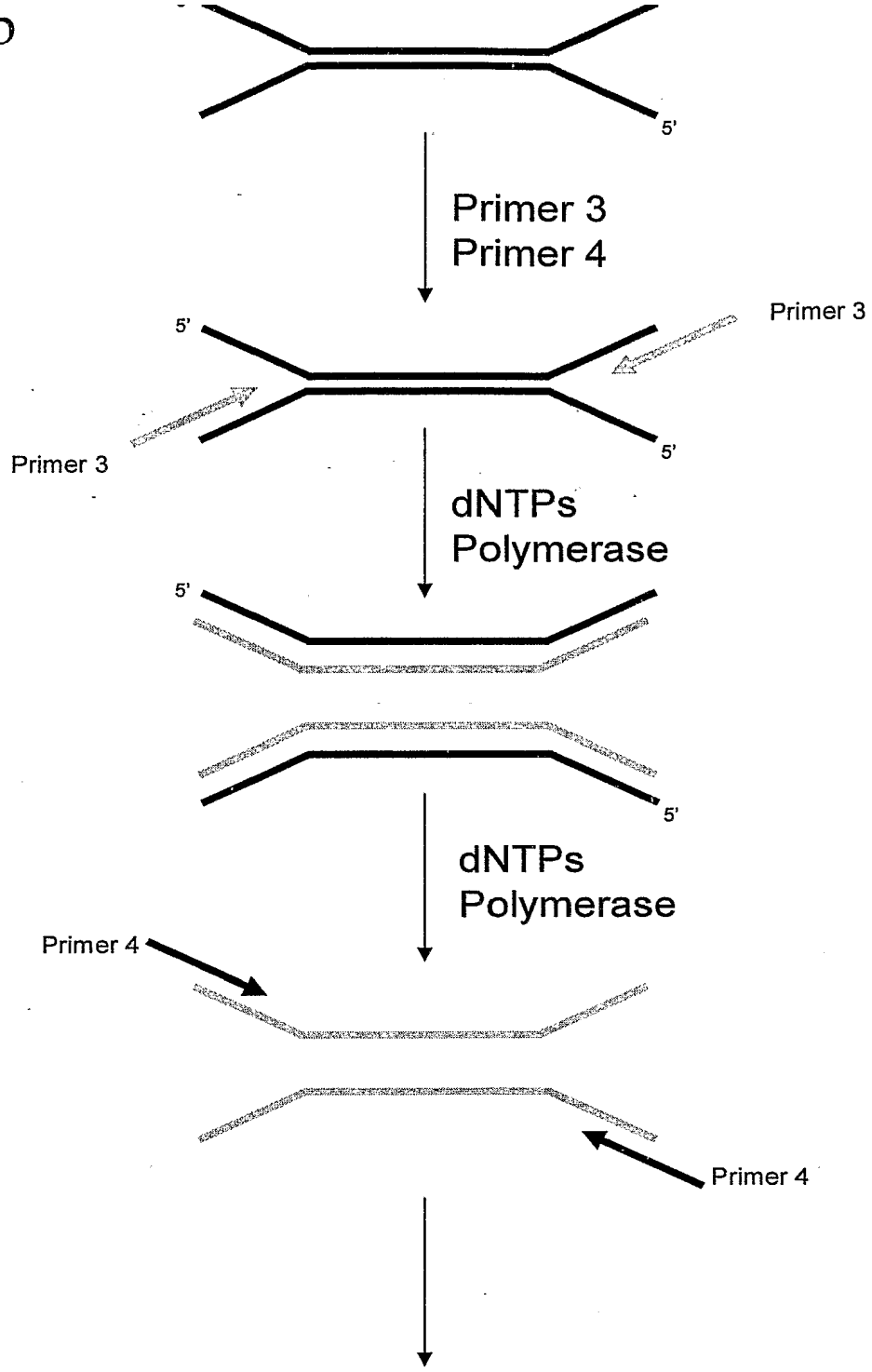

Fragmentation
End repair
Taq
dATP
Ligate
5'
5'
5'
T
5' P
5' P
A
A
5' P
5'
T
5' P
Remove adaptors
5'
5'
5' P
A
A
5' P
PCR with
Primer 3
Primer 4

Fig. 7b



5'				
	Primer 4	SEQ Primer	Template fragment	comp-Primer 3
	comp-Primer 4	comp-SEQ Primer	comp-Template fragment	Primer 3

Fig. 8

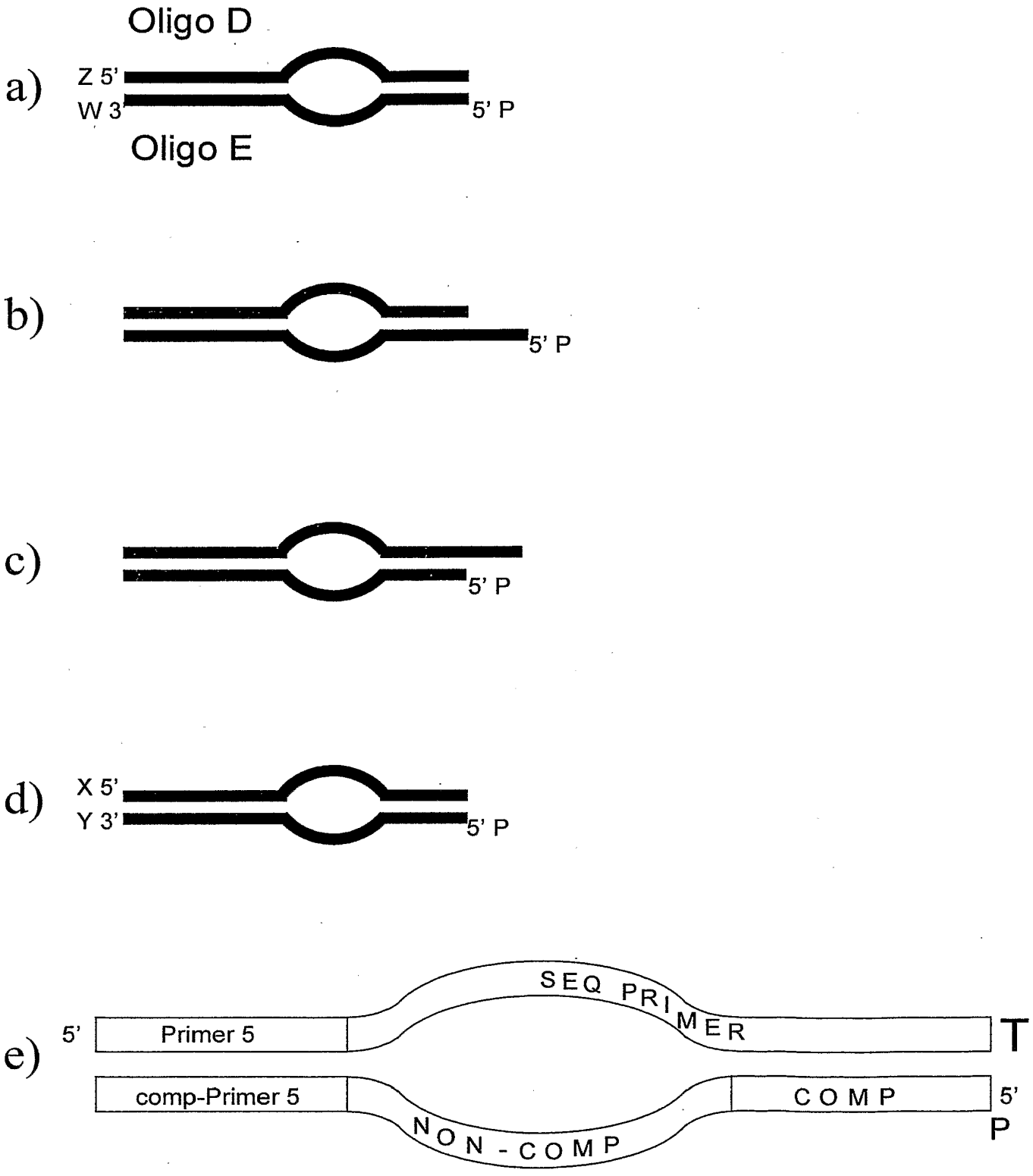

Oligo D
a)
Z 5'
W 3
5' P
Oligo E
b)
5' P
c)
5' P
d)
X 5'
Y 3'
5' P
e)
5'
Primer 5
SEQ PRIMER
T
comp-Primer 5
COMP
5'
P
NON-COMP

Fig. 9a

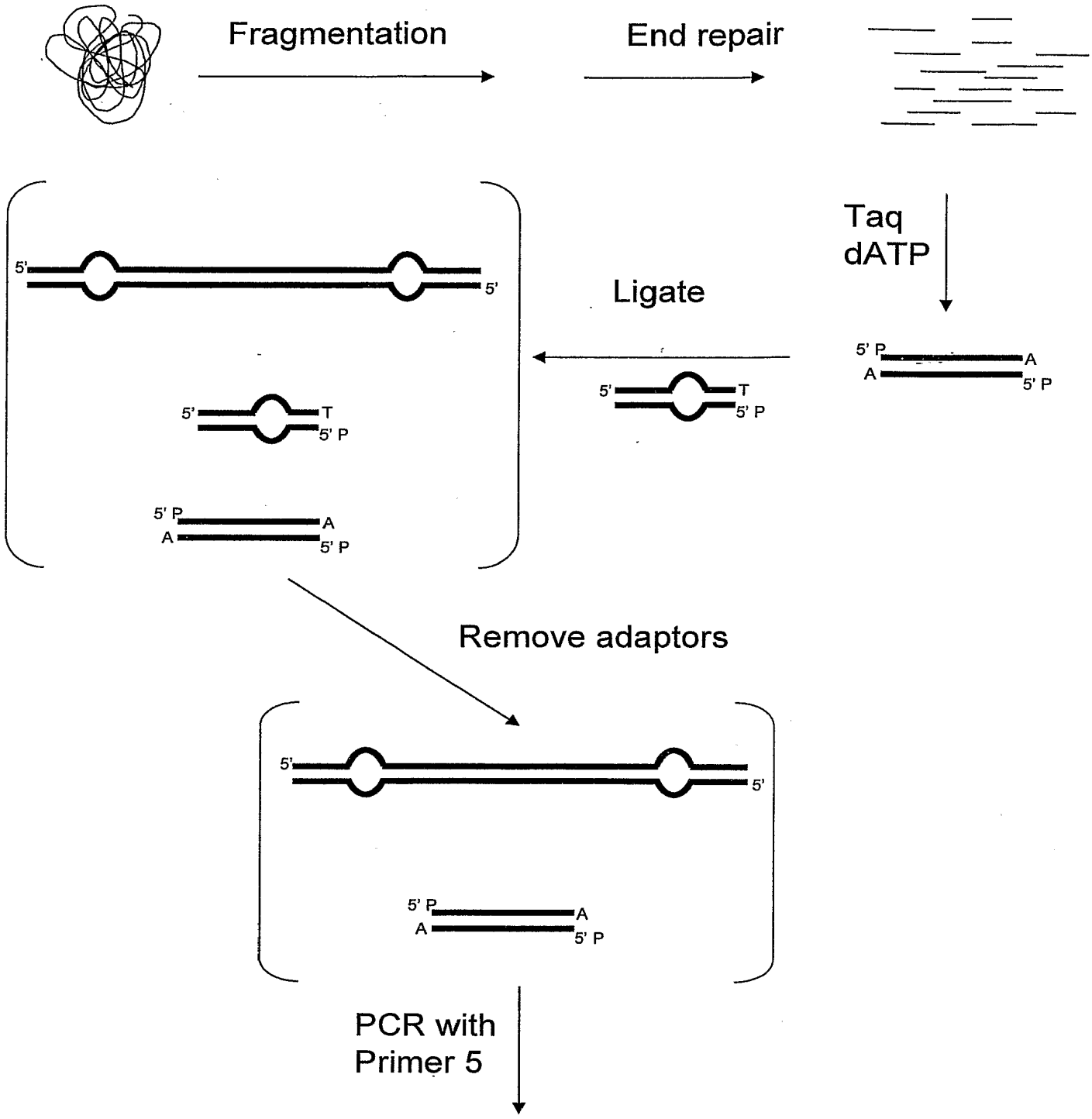
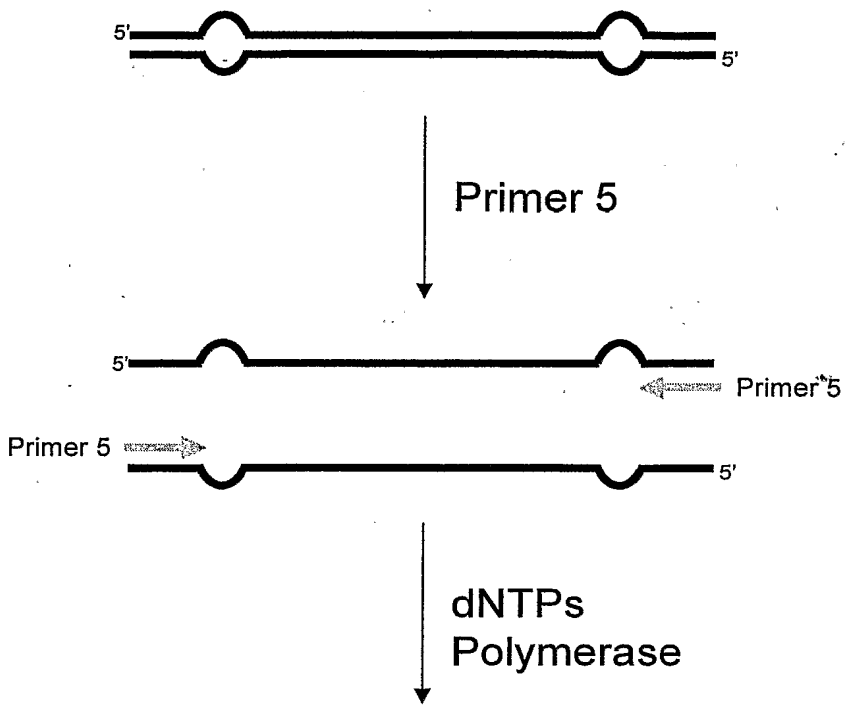

Fragmentation
End repair
Taq
dATP
Ligate
5'
5' P
T
A
Remove adaptors
PCR with
Primer 5

Fig. 9b



5' Primer 5	SEQ Primer	Template fragment		comp-Primer 5
comp-Primer 5	comp-SEQ Primer	comp-Template fragment		Primer 5 5'

SEQUENCE LISTING

<110> Solexa Limited

<120> Method of Preparing Libraries of Template Polynucleotides

<130> P81137WO00

<160> 23

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> synthetic oligonucleotide

<400> 1
acactctttc cctacacgac gctcttccga tct 33

<210> 2
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> synthetic oligonucleotide

<400> 2
gatcggaaga gctcgtatgc cgtcttctgc ttg 33

<210> 3
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> synthetic oligonucleotide

<400> 3
aatgatacgg cgaccaccga gatctacact ctttccctac acga 44

<210> 4
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> synthetic oligonucleotide

<400> 4
caagcagaag acggcatacg a 21

<210> 5
<211> 296
<212> DNA
<213> Homo sapiens

<400> 5
tgatacggcg accaccgaga tctacactct ttccctacac gacgctcttc cgatctgtgg 60
ggaccgtcct gtgcattgta gggtgttcaa cagcatccct gacctccacc tacaagatgc 120
cagtagcgaa tcccctcagc cctcatctcc ttgccatagt tgtgtcaacc aaaatcatct 180
ccacacattg ttagatgttt actgggaggc agactcactc ccacttgaga accactgtac 240
tagaaatatc accaagagaa tgagatcgga agagctcgta tgccgtcttc tgcttg 296

<210> 6
<211> 431
<212> DNA
<213> Homo sapiens

<400> 6
aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatctgt 60
gggctttgtt ctttgagagg ttgcagtcaa catgattctt taagaccaga accctgcaca 120
cttcttgggc tgtatttctt acattccttt tctattttaa ccatatccca tcttacctac 180
ttccagcata gtggtcatat ttaattttta caaaaccatt ttgccacttg ctgccaacta 240
tgttctttat aaagcagact ttgagatgga ggctagtgtt cagaggggat gcttaggaga 300
actttggaga ttaatactta tggcaggtaa gggaaggaag caggattaga cagaaatatt 360
gaactgtgat acaaagtcag caaagacttt agtcaataga tcggaagagc tcgtatgccg 420
tcttctgctt g 431

<210> 7
<211> 518
<212> DNA
<213> Homo sapiens

<400> 7
aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatcttt 60
cgattccctt caatgattat tccattcgag tacattcgat gattccattc gattctatat 120
gatgatgatt gcattcgagt ccgtgtatta ttccattcca ttccattaga tgattccatt 180
cgagtccatt cgatgattct cttcgattcc gttcgataat tacgcttgat tccgtttgat 240
gttgattcca ttcgagtcca ttcaatgtta attccattcg attctaagcg atgattccat 300
tcctttccat tagaagatga ttccattcga gaccattcga tgattgcatt caactcattc 360
gatgacgatt ccattcaatt ctgttcaatg attccattag attccatttg atgatgagtc 420
cattcgattc catttgatga tgattccatg cgattccatt agatgatgac ccctttcatt 480
tccaagatcg gaagagctcg tatgccgtct tctgcttg 518

<210> 8
<211> 387
<212> DNA
<213> Homo sapiens

<400> 8
aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatcttt 60
aaatgctagg catattgtgt accccacatt ggtttgtagc cagctctatg tcatagggcc 120
cttacccttt acctatttat tgttagtata atgtccataa acaagccaat ggctcagcat 180
gaactgatgc taaagaaagc tcatgcctga gtgataaatt aagtgacctc agctatttct 240
cttcagtgtt gtgaaagtta tttttaacag taggtttcct ggtagattct ctaaccactc 300
ggtatttcac atggcccaac ttggttaact cgactggtta cggcaaatgc tgaagatcgg 360
aagagctcgt atgccgtctt ctgcttg 387

<210> 9

<211> 536
<212> DNA
<213> Homo sapiens

<400> 9
aatgatacgg cgaccaccga gatctacact ctttccctac acgaagctct tccgatctta 60
aggaagttag gtagataatt tttgtttagg ccatatagct ttgattttct gataacaatt 120
ttataaactt agaaattttc atgtaagata caggaatact ggaagcaaaa aaaagaaggt 180
gctttaacct tagggattga aaaaatagta atttaggttg aaaatgctgc ttgaaagtta 240
atgctgatag cattactaca catgatgatt ttttctggaa ggaaagcttt atctgggcct 300
tcaatttagg aatttttctc tttggttttt aaaagctgcc atattcactt gagcttcatg 360
ggaaagatgc aaataactaa aacaaatgaa caaaaaccat gttgaggtca ggaacttatt 420
tcaagaaagc aagttctagg ttttctttta aagtgacagt agagccttag gcctcaaacc 480
atctacaacc atgttaacag taagatcgga agagctcgta tgccgtcttc tgcttg 536

<210> 10
<211> 364
<212> DNA
<213> Homo sapiens

<400> 10
aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tcagatcttc 60
ctgcctcagc ctcccgagta gttgggattg caggcatgtg ccaccatgcc ctgctaattt 120
ttgtattttt aactaaagga ggggttttgc catgttggcc aggctggtct tgaatgcctg 180
acctcaggtg atccgcccac ctcagcctcc cagtgctggg attacaggtg tgagccactg 240
cgcccagctg agggtaacta tttttaatgt ggctgatgaa tgtaactatc ctgtcccatg 300
tctctgtccc cagctgcaga gccctcgtcg agatcggaag agctcgtatg ccgtcttctg 360
cttg 364

<210> 11
<211> 374
<212> DNA
<213> Homo sapiens

<400> 11
aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatctaa 60
ggtttaaatt ttctatatag cctgaaaagt ttgaatgttt aattcaaact aatttattga 120
gcaatggcat ttaagaaaat ggaaagatac aaagggactt tcatcagatg ataagtggat 180
aagagagaaa aatgcagaca gatgagccag agttgtgtaa aagctgggag gctagcaggg 240
ccttgtagat agccaagctg actggggaac agagataaat gggagcagag atcagaaagt 300
tcatccttac cctgctgccc gtggtgaaag gagacttgca agatcggaag agctcgtatg 360
ccgtcttctg cttg 374

<210> 12
<211> 379
<212> DNA
<213> Homo sapiens

<400> 12
aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatctaa 60
ggaatcttta ttttctacat ttgagtttgg aaaactgagc tagcacatct aaatccatct 120
aatttttggtc attggtttta acaagttcat cttatttttt taaacatctg atctttattt 180
tatagaatag actacacaaa gtcttttgga aaattaaaat attttaactt ccaacaattt 240
tcagatttta cttataaaaa aatttaaaat cctctacttt actcgcatct ttattatttc 300
tgactttcta gctacttaaa gttaaggagg aaattaacct ctctaagatc ggaagagctc 360
gtatgccgtc ttctgcttg 379

<210> 13
<211> 455
<212> DNA
<213> Homo sapiens

<400> 13
aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatctcg 60
ccataaccac agcccaggcc tccgtagcca cagcacccat agccatggcc accattgtag 120
tttctgtagt agctgccaca catggtgttg gttgttgagg tcatccttgg gtaggaagga 180
gtgtaggtga cttcagtatg gacacttctc cgcagagggc cttttatatg cctcagtgaa 240
tcagaacata gcgtgcccct gcaaaaatat ctctaaaggc ctttcattgt gctgagaagt 300
tctggccctt acgtatctct ctgatttcat atcctgctac tctcctccat ttatctataa 360
tgctcaaact ctgctggctt tttgtctttt aaaatgcagc aggtttcttc tcacaataag 420
gagatcggaa gagctcgtat gccgtcttct gcttg 455

<210> 14
<211> 495
<212> DNA
<213> Homo sapiens

<400> 14
aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatctag 60
agagaagtta tttaagacaa gtaaggtatc aggttgtgac tcaaatacca cagaatctag 120
ctattgttag caatattaag tatattttct taaatgaagg attctccatt taccatatgc 180
cccttggata atttccagag aatttaattt tttaaaagga attttcacca attaaattat 240
tgttttgatc aaaagaggac ccactgaaca ccttattcat tattaaaatg tatcataaaa 300
cttaattatg gagctgggta caggggctca tgcctgtaat cccagcactt tgggaggctg 360
aggcaggagg actgcttgag tccaggagtt tgagaccagc ctgggtaaca tggtgaaacc 420
ctgtctctac aaaaaataca aaaaattagc caggtgtggt gagatcggaa gagctcgtat 480
gccgtcttct gcttg 495

<210> 15
<211> 491
<212> DNA
<213> Homo sapiens

<400> 15
aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatctcc 60
cagggaagcc aaaagattgg acacccctct ttcaaactat aaattcctcc catagttagt 120
ttggcctatg ctgggaaatg aacaagggtg gctttgaggt tagaagcaaa atggagtcag 180
ttaggtcaga ctttttttca ctatcatact ttttctatgt cagatttatc tcacttgtaa 240
tttttgcaag ggtggtttca gagccactaa gcttgtggta atttttttact gcaatagaaa 300
actaatgcat taggtaaccc tctttttttc cctctgattg cttgctctgg ggtaagccag 360
ctgccatgtt gaatttttca ttttgttact gagcaatcaa tgggctcact gcccgacgtg 420
catagaggcc aatactgtgg cactagtttt tgagaaaaga tcggaagagc tcgtatgccg 480
tcttctgctt g 491

<210> 16
<211> 603
<212> DNA
<213> Homo sapiens

<400> 16
aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatcttt 60
tcagatttta tatgtattaa ctcagaacac acacctctta tcacacatat tttttcatgt 120
aatttatcta aatcttatag aaaagggtcc atttgcattt tctcttatta gactcctgat 180
ttcaaataat atattactta tgagtatttt tctgtgctgt agttattcat tcttatagat 240
atgtaacata attccttttg caaaggtaaa aattgagcta tctcttgttg aggatttgtt 300
gatctctgtc taaagtttca aaaataaaga actttaaaag caaaatgtaa attcctttca 360
agttttagta aaattacttc aaacttagta gcttaaacaa tacagattta ttatgttaca 420
gttctgtaag acagaaatct gacttgatca caccatggta aaaccaagat actgccaggg 480
ttggtttttt cttgggggggg gtctgtggga agagtttgtt tcctttggtt ttccacagcc 540
cagaggctgc ttgcattcct ttgatcacta gatcggaaga gctcgtatgc cgtcttctgc 600
ttg 603

<210> 17

<211> 442
<212> DNA
<213> Homo sapiens

<400> 17
aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatctta 60
ttgtttaggg aataatgaca aggaaaaaaa gtctgtagat attcagtaca gaggcaccca 120
tctttttaaa tttctgaaga ttttttactc atgcttggtt gaatccacag atgcagaacc 180
cataggttca gagggccagc tgtgctttga aaatattagc ttgtgttttt attagaaaga 240
aaactctgag gccaggcacg gtggctcacg cctgtaatcc cagcactttg ggaggctgag 300
gtgggcggat cacaaggtga ggagatcgag accattctgg ctaacatggt gaaaccctgt 360
ctctactaaa aatacaaaaa aattagccgg gcgtggtagt gagcacctag atcggaagag 420
ctcgtatgcc gtcttctgct tg 442

<210> 18
<211> 537
<212> DNA
<213> Homo sapiens

<400> 18
aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatctgc 60
aattggtaaa acagtaagca atgaaacaga cacttctcaa atattccaag atggtacacg 120
cttttcagtg tgtatgatcc aataaagcca ttggaagtag gctttaatag tcaaaaaaga 180
ctattcagtt agataggaac tatttgccta taactattgg ccaaaaatag gttaaaaaat 240
tgttttaaat ttgtgcttta caaaacatgt ggacttttct agaaaatgtg tcaaatttca 300
aaagaaatat agacattatg gaaaggtcag ttaagcacag ccctaatcct gaaaacataa 360
ctatgaaaga tactagctgt tacttgtaac caaaaggaaa aaaaagatat tagtaaccaa 420
taattagcaa acaatgccca tatatttcct ttttttttct tttttttga gacaggggct 480
tactcaggct ggatgtgatc atgagatcgg aagagctcgt atgccgtctt ctgcttg 537

<210> 19
<211> 381
<212> DNA
<213> Homo sapiens

<400> 19
aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatcttt 60
agaagctcta ttatactgga aaagagatat gagacccttc ctactttaag aatcaatgaa 120
gccgggtgtg gtggctcacg cctgtactcc cagcattttg agaggccaag ctgggcagat 180
cacctgaggt cgggagtttg agaccagcct agccaacatg atgaaatcct gtttctacta 240
ataacacaaa aattagccgg gtgtggtggc gcacatctgg aatcccagct actccagagg 300
ctgaggcagg aaaattgctt gaagctggga agcagaagtt gcagtgaaga tcggaagagc 360
tcgtatgccg tcttctgctt g 381

<210> 20
<211> 650
<212> DNA
<213> Homo sapiens

<400> 20
aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatctgg 60
ctggactgaa taggatagcc ttagctgtaa aattgggctg atctttcaaa tggactcatg 120
cttgccgaat gactcacgct cctgtttaca aatcagctct gtgaagaaat gcagagtggg 180
aggctctgct tgccagacgg agaccttaga cctccagggg cggagaacgg agtacttcct 240
ctggtgctcg gcttcccttc ctgggggcag atctctcagc ttctggttgg tggctctcaa 300
aatccagaca caaggtcagc tgcagccagc gtgggccctg gagtagctcc agttatgggg 360
cagcaatggc cccctctcat tttgagagct cactttgcct gtggatggtt ttaatccatc 420
tggataaact tgaggcccat gggaatacca tatactatgg taaccatgta cactgctcta 480
aagatgtggc tgctgttgta taacttttc ctttattttt gtcaatttcc tattttccag 540
agtcttgcat acccactatg tctactgtga tagtgaacgt aaaaacatac aagatgttgg 600
tgttatcctc aatctcagat cggaagagct cgtatgccgt cttctgcttg 650

<210> 21
<211> 605
<212> DNA
<213> Homo sapiens

<400> 21
aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatctct 60
gtgagaaatc aatgtctgct gtttataagc cgccgggctg tgatatcctg tgagagtggc 120
cccagtggat gaagacagat gctctcaagg agcgcagatg acgcgggttc cgaaggactc 180
ggcacccagc ccggaggccg gcaacatggg caaggggcct ctcacggctg acctgtttcc 240
tcatcagcac atcaggacaa taagagctcc cacttcacag gtggtgaaga gccaacgtgg 300
tgaagaatga ataaagcagc tcgtggaaag tgctgtgcat gaggcctggc aaccggtccc 360
tgctctgagg tcacctgcca cggagctgct gacaggacca ttaaaaacac aattgtgcaa 420
gtgctcaccc acattcacag cagcagaatc tccaccagcc aagcattgga gacgatcctt 480
gcatccatag acatgaacag atgagcaaaa cgtggtctat acggacgatg aaatagcact 540
cagccctaag aagaaataaa atcccgacag agagatcgga agagctcgta tgccgtcttc 600
tgctt 605

<210> 22
<211> 439
<212> DNA
<213> Homo sapiens

<400> 22
aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatctgc 60
tgagaggaga ggagaggctg gcaggggget gatgcaggag gtgatagggc tcccggtgat 120
aagaggtgag aagaacagtc tctgtgtgcc tagagaagag attaccagaa gtctgctatc 180
tgtttgttcg cggatgtcgg acaggcagga tcggtgatgg caggtcttgg gggaaggatt 240
atcaggagct aaaagctgtc ttcaccttgg ctgctaagaa ctcatctcgg atcttcttag 300
aattccaaat cggacttttc tcctagcagt ggctacatcc ttaacctcaa aaatacccgt 360
attagcagat ctacctccat gaaatagaca attcttgaca aactaagatc ggaagagctc 420
gtatgccgtc ttctgcttg 439

<210> 23
<211> 698
<212> DNA
<213> Homo sapiens

<400> 23
aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatctgt 60
tatttttcat actgatgatt ctagggcggt tccctggact tggctgggac cagaactgtc 120
tcacagctgc cagcgctccc cacccccatg agaaccatcg tgctgaagat gaagagccaa 180
agccaaactg accagcctct gcaaccccaa gcatcaccgt gtgaggacag cgtgatgtgg 240
agccgcggag accccacgca gtgctgcgag gggcaaccac atgcaggggg gcagaggtgt 300
gggggagcaa gcagatgcca ccggcatgcc taggagccac agggagcatg cgggctgggc 360
agtgatggga atgaacgtga actccagtcc tgccccaaga agtcctccgg ccccctccttc 420
tcaattcagg gcacaaagtg gtaactgcag catgcagagg agtcgaggag tcttccctcg 480
tcccaggagc agcacctcgg gcacagtctc ggtcccacag aacagccaag tgtgggttgg 540
tggtctagag acctccgaag atccagtggg ggaaggatgg gcagcagagg gtctactctc 600
tgaaaataag gggaagggat tttccctccc cactgccaag gtcccagcta ctggacgtgg 660
gtggagatcg gaagagctcg tatgccgtct tctgcctt 698

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2006/004055

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68

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)
C12Q

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)
EPO-Internal, WPI Data, BIOSIS, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2005/090599 A2 (PLANT BIOSCIENCE LTD [GB]; FU GUOLIANG [GB]) 29 September 2005 (2005-09-29) page 6, line 29 - page 7, line 3; figures 1a,2 page 7, lines 15-21	1-29
X	US 2004/067493 A1 (MATSUZAKI HAJIME [US] ET AL) 8 April 2004 (2004-04-08) paragraphs [0022], [0038] - [0041], [0097]; figure 5	1-18, 25-29
X	US 6 287 825 B1 (WEISSMAN SHERMAN [US] ET AL) 11 September 2001 (2001-09-11) claims 1,5; figures 1,2a; example 1	1-18, 25-29

-/--

[X] Further documents are listed in the continuation of Box C.

[X] 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.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
7 February 2007	01/03/2007

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Lanzrein, Markus

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

IDT Exhibit 1020.3410

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2006/004055

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2004/081183 A2 (RUBICON GENOMICS INC [US]; PINTER JON [US]; KURIHARA TAKAO [US]; SLEPT) 23 September 2004 (2004-09-23) figure 18	1-29

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

IDT Exhibit 1020.3411

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/GB2006/004055

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2005090599 A2	29-09-2005	GB 2412170 A	21-09-2005
US 2004067493 A1	08-04-2004	NONE	
US 6287825 B1	11-09-2001	US 6346399 B1 US 6372434 B1	12-02-2002 16-04-2002
WO 2004081183 A2	23-09-2004	EP 1606417 A2	21-12-2005

IDT Exhibit 1020.3412

Espacenet

Bibliographic data: WO2009102896 (A2) — 2009-08-20

ISOTHERMAL NUCLEIC ACID AMPLIFICATION METHODS AND COMPOSITIONS

Inventor(s): KURN NURITH [US]; WANG SHENGLONG [US] ± (KURN, NURITH, ; WANG, SHENGLONG)

Applicant(s): NUGEN TECHNOLOGIES INC [US]; KURN NURITH [US]; WANG SHENGLONG [US] ± (NUGEN TECHNOLOGIES, INC, ; KURN, NURITH, ; WANG, SHENGLONG)

Classification: **- international:** ***C12N15/11; C12Q1/68***
- cooperative: **C12Q1/6844** more

Application number: WO2009US33964 20090212 Global Dossier

Priority number (s): US20080028146P 20080212 ; US20080038696P 20080321 ; US20080074991P 20080623 ; US20080085811P 20080801

Also published as: WO2009102896 (A3) EP2247727 (A2) EP2247727 (A4) EP2250283 (A2) WO2009102878 (A2) more

Abstract of WO2009102896 (A2)

Methods and compositions are provided related to the amplification of target polynucleotide sequences as well as total RNA and total DNA amplification. In some embodiments, the methods and compositions also allow for the immobilization and capture of target polynucleotides with defined 3' and or 5' sequences to solid surfaces. The polynucleotides attached to the solid surfaces can be amplified or eluted for downstream processing. In some cases, nucleotides attached to solid surfaces can be used for high throughput sequencing of nucleotide sequences related to target DNA or target RNA.

IDT Exhibit 1020.3413

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 August 2009 (20.08.2009)

PCT

(10) International Publication Number
WO 2009/102896 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01) *C12N 15/11* (2006.01)

(21) International Application Number:
PCT/US2009/033964

(22) International Filing Date:
12 February 2009 (12.02.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

61/028,146	12 February 2008 (12.02.2008)	US
61/038,696	21 March 2008 (21.03.2008)	US
61/074,991	23 June 2008 (23.06.2008)	US
61/085,811	1 August 2008 (01.08.2008)	US

(71) Applicant *(for all designated States except US)*: **NUGEN TECHNOLOGIES, INC.** [US/US]; 821 Industrial Road, Unit A, San Carlos, CA 94070 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **KURN, Nurith** [IL/US]; 2876 Ramona Street, Palo Alto, CA 94306 (US). **WANG, Shenglong** [CA/US]; 8248 Saturn Park Drive, San Ramon, CA 94582 (US).

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

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, 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, LT, 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, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

WO 2009/102896 A2

(54) **Title**: ISOTHERMAL NUCLEIC ACID AMPLIFICATION METHODS AND COMPOSITIONS

(57) **Abstract**: Methods and compositions are provided related to the amplification of target polynucleotide sequences as well as total RNA and total DNA amplification. In some embodiments, the methods and compositions also allow for the immobilization and capture of target polynucleotides with defined 3' and or 5' sequences to solid surfaces. The polynucleotides attached to the solid surfaces can be amplified or eluted for downstream processing. In some cases, nucleotides attached to solid surfaces can be used for high throughput sequencing of nucleotide sequences related to target DNA or target RNA.

IDT Exhibit 1020.3414

ISOTHERMAL NUCLEIC ACID AMPLIFICATION METHODS AND COMPOSITIONS

CROSS REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application Nos. 61/028,146, filed February 12, 2008, 61/074,991, filed June 23, 2008, and 61/085,811, filed August 1, 2008, which applications are incorporated herein by reference in their entirety. This application is also related to the co-pending patent application [Attorney Docket No 25115-726.201] filed February 12, 2009, which is incorporated herein by reference in its entirety.

BACKGROUND OF THE INVENTION

[0002] The quality and quantity of nucleic acid sample is important for many studies. High-throughput genomic analysis requires large amounts of template for testing, yet typically the yield of nucleic acids from individual patient samples is limited. Forensic and paleoarcheology work also can be severely limited by nucleic acid sample size. The limitation of starting material impacts the ability to carry out large scale analysis of multiple parameters, as is required for, for example, the genotyping of multiple loci in the study of complex diseases, detecting the presence or absence of specific nucleic acid species in a sample, large scale sequencing and the like. Moreover, it is well accepted that molecular analysis determination of genomic instability in various pathological condition such as cancer, is most precisely carried out in well defined cell populations, such as that obtained by laser capture micro-dissection or cell sorting. Nucleic acid amplification technologies that provide global amplification of very small polynucleotide samples, for example, from one or a very few cells, may provide a solution to the limited starting materials generally available for analysis.

[0003] Likewise, the ability to amplify ribonucleic acid (RNA) is an important aspect of efforts to elucidate biological processes. Total cellular mRNA represents gene expression activity at a defined time. Gene expression is affected by cell cycle progression, developmental regulation, response to internal and external stimuli and the like. The profile of expressed genes for any cell type in an organism reflects normal or disease states, response to various stimuli, developmental stages, cell differentiation, and the like. Non-coding RNAs have been shown to be of great importance in regulation of various cellular functions and in certain disease pathologies. Such RNAs are often present in very low levels. Thus, amplification methods capable of amplifying low abundance RNAs, are of great importance.

[0004] Various methods for global amplification of DNA target molecules (e.g., whole genome amplification) have been described, including methods based on the polymerase chain reaction (PCR). See, e.g., U.S. Pat. Nos. 5,731,171; 6,365,375; Daigo et al., (2001) Am. J. Pathol. 158 (5):1623-1631; Wang et al, (2001); Cancer Res. 61:4169-4174; Zheng et al, (2001) Cancer Epidemiol. 10:697-700; Dietmaier et al (1999) Am. J. Pathol. 154 (1) 83-95; Stoecklein et al (2002) Am. J. Pathol. 161 (1):43-51; U.S. Pat. Nos. 6,124,120; 6,280,949; Dean et al (2002) PNAS 99 (8):5261-5266. However, PCR-based global amplification methods, such as whole genome amplification (WGA), may generate non-specific amplification artifacts, give incomplete coverage of loci, or generate DNA of insufficient length that cannot be used in many applications. PCR-based methods also suffer from the propensity of the PCR reaction to generate products that are preferentially amplified, and thus resulting in biased representation of genomic sequences in the products of the amplification reaction. Methods of global amplification of DNA using composite primers have been described. See e.g. U. S. Patent application 10/824,829.

IDT Exhibit 1020.3415

[0005] Additionally, a number of methods for the analysis of gene expression have been developed in recent years. See, for example, U.S. Pat. Nos. 6,251,639, 6,692,918, 6,686,156, 5,744,308; 6,143,495; 5,824,517; 5,829,547; 5,888,779; 5,545,522; 5,716,785; 5,409,818; EP 0971039A2; EP0878553A2; and U.S. published patent applications nos. 2002/0115088, 2003/0186234, 2003/0087251, and 2004/0023271. These include quantification of specific mRNAs, and the simultaneous quantification of a large number of mRNAs, as well as the detection and quantification of patterns of expression of known and unknown genes. RNA amplification is most commonly performed using the reverse transcriptase-polymerase chain reaction (RT-PCR) method and variations thereof. These methods are based on replication of RNA by reverse transcriptase to form single stranded DNA complementary to the RNA (cDNA), which is followed by polymerase chain reaction (PCR) amplification to produce multiple copies of double stranded DNA. However, the total amount of sample RNA that is available is frequently limited by the amount of biological sample from which it is derived. Biological samples are often limited in amount and precious. Moreover, the amount of the various RNA species is not equal; some species are more abundant than others are, and these are more likely and easier, to analyze. The ability to amplify RNA sequences enables the analysis of less abundant, rare RNA species. The ability to analyze small samples, by means of nucleic acid amplification, is also advantageous for design parameters of large scale screening of effector molecule libraries, for which reduction in sample volume is a major concern both for the ability to perform very large scale screening or ultra high throughput screening, and in view of the limiting amounts of library components. Methods of amplification from RNA templates have been described, for example in U. S. Patent 6,946,251.

[0006] Sequencing of nucleic acids continues to be one of the most important and useful ways to analyze DNA and RNA samples. Recent developments have made possible highly parallel high throughput sequencing. Many of these approaches use an in vitro cloning step to generate many copies of each individual molecule. Emulsion PCR is one method, isolating individual DNA molecules along with primer-coated beads in aqueous bubbles within an oil phase. A polymerase chain reaction (PCR) then coats each bead with clonal copies of the isolated library molecule and these beads are subsequently immobilized for later sequencing. See, e.g. WO04069849A2, WO05010145A2. In other cases, surface methods of clonal amplification have been developed, for example, by the use of bridge PCR where fragments are amplified upon primers attached to a solid surface. These methods produce many physically isolated locations which each contain many copies of a single fragment. While these methods have provided improvements in sequencing throughput, there is a continuing need to improve the methods of obtaining samples appropriate for sequencing, and of handling, storing, and amplifying such samples. In particular, there is a need to improve methods for obtaining high throughput sequencing data for a specific set of genes or gene products from whole genome or transcriptome samples.

[0007] Therefore, there is a need for improved methods of obtaining, storing, amplifying, and analyzing DNA and RNA samples, including methods which can globally or specifically amplify DNA or RNA polynucleotide targets. The invention described herein fulfills this need.

SUMMARY OF THE INVENTION

[0008] In one aspect the present invention relates to a method for creating a partial RNA-DNA heteroduplex comprising: providing a template nucleic acid in a reaction mixture; adding an RNA-DNA chimeric primer;

IDT Exhibit 1020.3416

adding a DNA polymerase; and adding an exonuclease; wherein said DNA polymerase comprises RNA-dependent polymerase activity.

[0009] The present invention also describes a kit comprising a container containing reagents for forming a water in oil emulsion and a DNA polymerase with substantial strand-displacement activity; the kit optionally further comprising one or more all DNA primers, an RNA-DNA chimeric oligonucleotide, an exonuclease, RNA-DNA chimeric primer, RNase H, and instructions for use of said reagents. The DNA polymerase may have RNA-dependent polymerase activity. The invention also includes a kit comprising an all DNA first primer, an RNA-DNA chimeric oligonucleotide; the kit optionally further comprising a second primer, a polymerase having substantial strand-displacement activity, and RNase H.

[0010] In another aspect the invention relates to a method for producing amplified nucleic acid comprising: providing a template nucleic acid; annealing a first primer or set of first DNA primers comprising a 3' end whereby a portion of the 3' end comprises a template annealing sequence and a 5' tail sequence whereby a portion of the tail sequence comprises a sequence (A); extending said first primer or set of first primers with a DNA polymerase to create a first primer extension product or first set of DNA polymerase products; dissociating the primer extension product or products of said DNA polymerase from the template nucleic acid; annealing a second primer comprising a 3' DNA sequence and a 5' RNA sequence, whereby a portion of the 3' end comprises a randomized annealing sequence and a portion of the 5' end comprises sequence (A) or sequence (B), to said primer extension product or products; and extending said second primer with a DNA polymerase to create a second primer extension product or products to produce a double stranded DNA product or products comprising an A-A' DNA duplex at one end.

[0011] The template nucleic acid may be selected from the group consisting of DNA and RNA. The template nucleic acid may be RNA and the DNA polymerase of step (c) is an RNA dependant DNA polymerase. The dissociation step of step (d) is performed by degrading the RNA. The template nucleic acid is DNA and the DNA polymerase of step (c) can be a DNA dependant DNA polymerase. The dissociation step of step (d) may be performed by heat or chemical denaturation. The DNA polymerase may be one that does not exhibit substantial strand displacement activity. The methods herein may have each template annealing sequence of the set of primers is specific for a target or region of template nucleic acid.

[0012] The methods herein can have the template annealing sequence of the first primer comprises a random sequence. In some instances, the members of the set of first primers can each comprise a distinct 3' DNA annealing sequence, each specific for a target or region of template nucleic acid. The first primer tail sequence may not be complementary to the template nucleic acid. The 5' tail sequence of the second primer can comprise sequence (A). The 5' tail sequence of the second primer may be one that does not comprise tail sequence (A).

[0013] The method may further comprise step (g), degradation of single stranded 3' end of said first primer extension product or products in the complex of first and second primer extension products with a single stranded DNA specific 3' exonuclease. The method may further comprise step (h), the addition of a DNA polymerase with reverse transcriptase activity to create a double stranded nucleic acid with an A-A' DNA duplex at one end and an B-B' DNA-RNA heteroduplex at the other end.

[0014] The method may further comprise performing SPIA with a reaction mixture comprising RNAse H, a composite amplification primer comprising a 3' DNA portion and a 5' end whereby a portion of the 5' end comprises RNA wherein the RNase H reacts with the heteroduplex of step (h) to produce a single stranded

IDT Exhibit 1020.3417

end, and the amplification primer hybridizes to the single stranded end produced by RNaseH, and a DNA polymerase with strand displacement activity.

[0015] The first primer or set of primers described in any of the inventions herein can comprise a ligand portion at the 5'end.

[0016] Any of the methods herein can further comprise cleaving of the RNA from the DNA-RNA heteroduplex with RNase H; annealing a chimeric oligonucleotide comprising a 5' end whereby the 5' end comprises sequence (C) and a 3' end whereby the 3' end comprises the DNA sequence (B) to the double stranded nucleic acid product; extending the double stranded nucleic acid product with a DNA polymerase to produce a C-C' DNA-RNA heteroduplex; and amplifying the second primer extension product with a reaction mixture comprising RNase H, an amplification primer comprising a 3' DNA portion and a 5' end whereby a portion of the 5' end comprises RNA that hybridizes to a portion of the single stranded end produced by RNase H, and a DNA polymerase with strand displacement activity to produce an amplification product comprising sequence (B) near its 3' end and sequence (A') at the 5' end.

[0017] In another aspect the present invention relates to a method for producing amplified nucleic acid comprising: providing a template nucleic acid; annealing a first DNA primer or set of first DNA primers comprising a 3' end whereby a portion of the 3' end comprises a template annealing sequence and a 5' end whereby a portion of the 5' end comprises a DNA tail sequence (A); extending said first primer or set of first primers with a DNA polymerase to create a primer extension product or set of primer extension products; dissociation of the primer extension product or products from the template nucleic acid; annealing a second primer comprising a 3' DNA end, whereby a portion of the 3' end comprises a random template annealing sequence, and a 5' RNA end, whereby a portion of the 5' end comprises the tail sequence (A), to said primer extension product or products or to said template nucleic acid; extending said second primer with a DNA polymerase to create a second primer extension product or set of second primer extension products; and denaturing and annealing the second primer extension product or set of second primer extension products to form a stem loop structure comprising an A-A' RNA-DNA heteroduplex. The DNA polymerase may not exhibit substantial strand displacement activity. Each DNA annealing sequence in the set of first DNA primers can be specific for a target or region of the template nucleic acid. A portion of the 5' tail sequence of the first primer can be sequence (A) and a portion of the 5' tail sequence of the second primer can be sequence (A). The tail sequence may not be complimentary to the template nucleic acid.

[0018] The method can comprise the addition of a 3' single stranded DNA specific exonuclease and single strand specific RNase to degrade unincorporated primers and template nucleic acid. The methods can further comprising amplification with a reaction mixture comprising RNase H, an amplification primer comprising sequence (A'), and a DNA polymerase with strand displacement activity, wherein the amplification primer comprises a 3'-DNA sequence and a 5'-RNA sequence.

[0019] The template nucleic acid can be selected from the group consisting of RNA and DNA. In certain embodiments the template nucleic acid can be RNA and the DNA polymerase can be an RNA dependent DNA polymerase. In other embodiments the template nucleic acid can be DNA and the DNA polymerase can be a DNA dependent DNA polymerase.

[0020] In the methods described herein that include a dissociation step, such a step can be performed by degradation of the RNA. Dissociation can be carrier out or performed by chemical or heat denaturation.

[0021] In one aspect the present invention relates to a method of capturing and amplifying a target sequence or sequences comprising: providing a surface wherein a first primer or set of first primers comprising a 5' tail

IDT Exhibit 1020.3418

sequence comprising sequence (A) and a 3' template annealing sequence comprising sequence (P) are immobilized on said surface; annealing a target nucleic acid to said first primer or set of primers; extending the first primer or set of primers with a DNA polymerase to produce an immobilized primer extension product or set of immobilized primer extension products complimentary to and hybridized with the target nucleic acid to form a complex or set of complexes; dissociating the target nucleic acid-primer extension product complex or set of primer extension product complexes; annealing a second primer comprising a 3' DNA template annealing sequence (P) or set of template annealing sequences and a 5' RNA tail sequence comprising sequence (A) or sequence (B) to the immobilized DNA polymerase product or DNA polymerase products; and extending said second primer with a DNA polymerase to create a second primer extension product or set of primer extension products comprising a double stranded nucleic acid comprising a 3' DNA-DNA duplex A-A'.

[0022] The method can further comprise the step of eluting of the second primer extension product by heat or chemical denaturation. The method can also further comprise annealing an amplification primer comprising a 5' RNA sequence and a 3'DNA sequence; extending said annealed primer with a DNA polymerase; and amplifying the eluted DNA polymerase product with a reaction mixture comprising RNAse H, an amplification primer comprising a DNA portion and a 5' RNA portion, and a DNA polymerase with strand displacement activity to produce an amplified DNA product or products having a defined 5' sequence (A') or (B').

[0023] The DNA polymerase does not need to exhibit strand displacement activity.

[0024] The template nucleic acid can be selected from the group consisting of RNA and DNA. The first primer can include a spacer element between the immobilizing surface and the remaining portion of the primer, a 3' DNA end comprising a 3' DNA annealing sequence (P), and a 5' comprising a common 5' DNA sequence tail (A) or sequence tail (B) between said spacer element and annealing sequence.

[0025] The set of first primers can include a spacer element between the immobilizing surface and the remaining portion of the primer; a 3' DNA end wherein members of the set of first primers comprise unique target specific 3' DNA annealing sequences (P); and a 5' DNA end comprising a common 5' DNA sequence tail (A) or DNA sequence tail (B) between said spacer element and annealing sequence.

[0026] The annealing sequences (P) of set of first primers can each be specific to a target or region of the template nucleic acid. The first primers can be specific to one or more of the following gene targets: kinases, protein kinases, lipid kinases, phosphatases, G-protein coupled receptors, proteases, serine proteases, metalloproteases, cysteine proteases, aspartyl proteases, or cytokines.

[0027] The 5' tail sequence of the first primer can be sequence (A) and the 5' tail sequence of the second primer can be sequence (B). Alternatively, the 5' tail sequence of the first primer can be sequence (A) and the 5' tail sequence of the second primer can be sequence (A). The alternative method can further comprise separating the second primer extension product or products from the first primer extension product or products by heat or chemical denaturation; annealing the second primer extension product in solution to create a stem loop structure with a DNA sequence (A')-RNA sequence (A) heteroduplex end; degrading RNA in the heteroduplex with RNase H; annealing an amplification primer to the single stranded portion of the second primer extension product wherein the amplification primer has a DNA portion and a 5' RNA portion, wherein the RNA portion of the amplification primer forms a heteroduplex with a DNA portion of the second primer extension product; extending the amplification primer with a DNA polymerase having strand displacement activity to produce an amplified product hybridized to the second primer extension

IDT Exhibit 1020.3419

product to form a double stranded nucleic acid, said amplified product forming an RNA-DNA heteroduplex at one end of the double stranded nucleic acid; and repeating the steps to produce amplified DNA. A single stranded DNA specific 3' exonuclease can be added to degrade unincorporated primers.

[0028] The surface can be selected from the group consisting of a bead, a magnetic particle, a microarray, a gene chip, and an array.

[0029] In some embodiments the method can further include degradation of the single stranded 3' end of said first primer extension product and unextended immobilized first primers with a single stranded DNA specific 3' exonuclease; cleavage of spacer element with light or chemical cleavage to release the immobilized double stranded nucleic acid from the surface; and amplifying the released double stranded nucleic acid in solution with a reaction mixture comprising RNase H, an amplification primer comprising a DNA portion and a 5' RNA portion, and a DNA polymerase with strand displacement activity to produce an amplified DNA product or products having a defined 3' sequence comprising sequence (A').

[0030] In other embodiments the method can further include degrading the single stranded 3' end of said first primer extension product or products and unextended immobilized first primers with a single stranded DNA specific 3' exonuclease; extending said first primer extension product with a DNA polymerase; and degrading RNA in the heteroduplex from the first primer extension product and the RNA portion of the second primer with RNAse H; annealing an amplification primer to the single stranded portion of the second primer extension product wherein the amplification primer has a DNA portion and a 5' RNA portion; extending the amplification primer with a DNA polymerase having strand displacement activity to produce an amplified product hybridized to the second primer extension product; and repeating these steps to produce multiple copies of amplified product or products having a defined 3' sequence comprising sequence (A').

[0031] The present invention also describes a kit comprising a container or containers comprising a first primer comprising a 3' DNA portion and a 5' DNA portion, wherein the 5' DNA portion further comprises sequence (A) and the 3' portion further comprises an annealing sequence (P); a second primer comprising a 3' DNA portion and a 5' RNA portion, wherein the 5' RNA portion further comprises sequence (A) or (B) and the 3' portion further comprises an annealing sequence (P); a DNA polymerase; a DNA dependent DNA polymerase with strand displacing activity; RNase H; and a chimeric amplification primer comprising a 3' DNA portion and a 5' RNA portion wherein the sequence of the amplification primer is substantially the same as the (A) sequence.

[0032] The present invention also describes a kit comprising a container or containers comprising a first primer comprising a 3' DNA portion and a 5' DNA portion, wherein the 5' DNA portion further comprises sequence (A) and the 3' DNA portion further comprises annealing sequence (P); a second primer comprising a 3' DNA portion and a 5' RNA portion, wherein the 5' RNA portion further comprises sequence (A) or (B) and the 3' DNA portion further comprises annealing sequence (P); a DNA polymerase; a DNA dependent DNA polymerase with strand displacing activity; a chimeric oligonucleotide comprising a 3' DNA portion substantially comprising sequence (A) or (B) and a 5' RNA portion comprising sequence (C)' RNase H; and a chimeric amplification primer comprising a 3' DNA portion and a 5' RNA portion wherein the sequence of the amplification primer is substantially the same as the (A) sequence.

[0033] The present invention also describes a kit comprising a container or containers comprising a first primer comprising a DNA portion wherein the DNA portion further comprises an annealing sequence (P); a second primer comprising a 3' DNA portion and a 5' RNA portion, wherein the 5' RNA portion further

IDT Exhibit 1020.3420

comprises sequence (A) or (B) and the 3' portion further comprises an annealing sequence (P); a DNA polymerase; a DNA dependent DNA polymerase with strand displacing activity; RNase H; and a chimeric amplification primer comprising a 3' DNA portion and a 5' RNA portion wherein the sequence of the amplification primer is substantially the same as the (A) sequence.

[0034] The present invention also describes a kit comprising an immobilized first primer comprising a 3' DNA portion and a 5' DNA portion, wherein the 5' DNA portion further comprises sequence (A), and the 3' DNA portion further comprises annealing sequence (P); a second primer comprising a 3' DNA portion and a 5' RNA portion, wherein the 5' RNA portion further comprises sequence (A) or (B); a DNA polymerase; a DNA dependent DNA polymerase with strand displacing activity; RNase H; and a chimeric amplification primer comprising a 3' DNA portion and a 5' RNA portion wherein the sequence of the amplification primer is substantially the same as the (A) sequence.

[0035] The kits of the invention can have the DNA polymerase that is an RNA dependant DNA polymerase. The kits can have a DNA dependant DNA polymerase. The kits can have a DNA polymerase that does or does not possess substantial strand displacing activity.

[0036] The kits can comprise a first primer where the 3' DNA portion of the first primer comprises a random sequence. The kits can comprise a first primer where the 3' DNA portion of the first primer comprises a specific sequence. The kits can comprise a first primer where the 3' DNA portion of the first primer comprises a set of specific sequences. The kits can comprise a first primer which further comprises a ligand.

[0037] The kits can comprise a second primer where the 3' DNA portion of the second primer comprises a random sequence. The kits can comprise a second primer where the 3' DNA portion of the second primer comprises a specific sequence. The kits can comprise a second primer where the 3' DNA portion of the second primer comprises a set of specific sequences.

[0038] The kits can further comprise a single stranded DNA specific 3' exonuclease. The kits can further comprise an RNase specific for single and double stranded RNA.

[0039] In another aspect an amplification method comprises forming a polynucleotide strand comprising a sequence (A) at its 5' end and complementary sequence (A') at its 3' end wherein sequence (A) comprises RNA within a mixture comprising other polynucleotides that are single stranded; treating said polynucleotide so as to intramolecularly anneal the (A) sequence to the (A') sequence; purifying the polynucleotide by degrading the other polynucleotides by treating the mixture with one or more enzymes that degrade single stranded polynucleotides; and performing SPIA amplification with the purified polynucleotide from step (c) using sequence (A') as a priming site for the SPIA amplification.

[0040] In yet another one aspect an amplification method comprises: obtaining a polynucleotide strand comprising sequence (B') at its 3' end; annealing a chimeric oligonucleotide comprising a 3' DNA portion and a 5' RNA portion, wherein the 5'portion comprises sequence (C), and the 3' portion comprises sequence (B); extending the polynucleotide strand with a RNA dependent DNA polymerase to add a DNA sequence (C') to the polynucleotide strand; and performing SPIA amplification using the product of step (c) as the SPIA substrate using sequence (C') as a priming site for SPIA amplification.

IDT Exhibit 1020.3421

INCORPORATION BY REFERENCE

[0041] All publications and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0042] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention may be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0043] **Figure 1** illustrates the initial steps of a method for producing amplified DNA by way of generating a single primer isothermal linear amplification (SPIA) substrate using an all DNA first primer with a sequence tail (A) and an annealing sequence (P), a DNA-RNA chimeric second primer with a sequence tail (A) or (B), and a DNA template.

[0044] **Figure 2** illustrates methods for producing a SPIA substrate for producing amplified DNA with a defined sequence tail (A') from the product of **Figure 1**.

[0045] **Figure 3** illustrates an alternative method of producing amplified DNA from the product of **Figure 1.**

[0046] **Figure 4** illustrates methods for producing captured and immobilized DNA using immobilized all DNA first primers with sequence tails (A) and annealing sequences (P), a template nucleic acid, and a DNA-RNA chimeric second primer with a sequence tail (A) or (B).

[0047] **Figure 5** illustrates methods for generating amplified DNA from immobilized DNA of **Figure 4** having a defined 3' sequence (A').

[0048] **Figure 6** illustrates an alternative method for generating amplified DNA from immobilized DNA of **Figure 4** having a defined 3' sequence (A') in which the SPIA substrate is cleaved from the surface prior to amplification.

[0049] **Figure 7** illustrates the initial steps of a method for producing a SPIA substrate using an immobilized all DNA first primer with sequence tail (A), a DNA-RNA chimeric second primer with the same sequence tail (A), and a template nucleic acid.

[0050] **Figure 8** illustrates a method for producing a stem loop SPIA substrate from the product of **Figure 7**.

[0051] **Figure 9** illustrates a method of producing a SPIA substrate with a defined sequence (A) at its 5' end using an all DNA first primer with a sequence tail (A), a DNA-RNA chimeric second primer with a sequence tail (B), an RNA template, and a chimeric DNA/RNA amplification primer.

[0052] **Figure 10** illustrates a method of producing amplified DNA with a defined sequence (A') at its 3' end and a defined sequence (B) at its 5' end using a chimeric oligonucleotide.

[0053] **Figure 11** illustrates a method of producing a SPIA substrate with a 5' ligand for immobilization, functionalization, or capture and a defined 5'sequence (A) using a ligand attached all DNA first primer with a sequence tail (A), a DNA-RNA chimeric second primer with sequence tail (B) and an RNA template.

[0054] **Figure 12** illustrates a method of producing a SPIA substrate with a 5' ligand, a defined sequence (A) at its 5' end, and a defined sequence (B') at its 3' end using a ligand attached all DNA first primer with tail sequence (A), a DNA-RNA chimeric second primer with a sequence tail (B), a DNA-RNA chimeric

IDT Exhibit 1020.3422

oligonucleotide with a DNA sequence (B) at its 3' end and an RNA sequence (C) at its 5' end, and an RNA template.

[0055] **Figure 13** illustrates a method of producing a SPIA substrate from a DNA template using an all DNA first primer with a randomized annealing sequence (P) and without a sequence tail, and a DNA-RNA chimeric second primer with a sequence tail (B).

[0056] **Figure 14** illustrates a method of producing a SPIA substrate from a DNA template using an all DNA first primer with a target specific annealing sequence (P) and lacking a sequence tail, and a DNA-RNA chimeric second primer with a sequence tail (B).

[0057] **Figure 15** illustrates a method of producing a SPIA substrate with a 5' ligand for immobilization, functionalization, or capture and a defined 5' sequence (A) using a ligand attached all DNA first primer with a sequence tail (A) and a randomized annealing sequence (P), a DNA-RNA chimeric second primer with sequence tail (B) and an DNA template.

[0058] **Figure 16** illustrates a method of producing a SPIA substrate from a target nucleic acid for generation of amplified DNA having a defined 5' sequence (A) using an all DNA first primer with a sequence tail (A) and a randomized or target specific annealing sequence (P), an all DNA second primer, and an RNA-DNA chimeric oligonucleotide with DNA sequence (A) at its 3' end and RNA sequence (C) at its 5' end.

[0059] **Figure 17** illustrates a method of producing a SPIA substrate from a target nucleic acid for generation of amplified DNA having a defined 5' sequence (A) and a defined 3' sequence (B) using an all DNA first primer with a sequence tail (A) and a randomized or target specific annealing sequence (P), an all DNA second primer having a tail sequence (B), and an RNA-DNA chimeric oligonucleotide with DNA sequence (A) at its 3' end and RNA sequence (C) at its 5' end.

[0060] **Figure 18** illustrates a method of producing amplified DNA from the SPIA substrate of **Figure 17.**

[0061] **Figure 19** illustrates the results of SPIA-based target specific amplification of DNA using a tailed all-DNA target specific first primer.

[0062] **Figure 20** illustrates normalized results of SPIA-based target specific amplification of DNA using a tailed all-DNA target specific first primer.

[0063] **Figure 21** illustrates results of sequence-specific single and multiplex amplification of genes from yeast genomic DNA using a tailed all-DNA target specific first primer.

[0064] **Figure 22** illustrates results the results of using the Taq polymerase enzyme for extension of the tailed all-DNA sequence specific first primer and an RNA-DNA chimeric second primer.

DETAILED DESCRIPTION OF THE INVENTION

Introduction

[0065] This invention provides new methods of amplifying nucleic acids. The invention also provides for systems reaction mixtures, and kits relating to these methods. The methods of the invention generally incorporate single primer isothermal amplification (SPIA). SPIA is an isothermal amplification method described, for example, in U. S. Patents 6,692,918, 6,251,639, 6,946,251 and 7,354,717. In some instances SPIA amplification generally involves a chimeric RNA-DNA amplification primer, RNase H, and a DNA polymerase enzyme to produce amplified DNA product. The SPIA amplification generally amplifies the sequences downstream of the hybridization site of the SPIA primer on the nucleic acid to which the SPIA primer binds. SPIA is a rapid and effective method of isothermal, linear amplification that can be used to

IDT Exhibit 1020.3423

produce amplified DNA that is representative of DNA or RNA from a variety of sources, for example, genomic DNA or the messenger RNA transcripts.

[0066] The methods of the present invention generally utilize all-DNA first primers that are designed to hybridize to a nucleic acid target, often referred to as a template, that has sequences to be amplified. Prior methods have described how to generate a substrate for SPIA from a target nucleic acid by using chimeric RNA-DNA first primers. The methods of the present invention can be useful, for example, for amplifying specific sequences within the target nucleic acid. The methods of the invention are also useful for multiplex amplification, performing the simultaneous amplification of multiple specific sequences. The number of specific sequences can be, for example, 2 – 10, 10 – 100, 100 – 1,000, 1,000 -10,000, 10,000 – 100,000 or more than 100,000. In some cases, the number of specific sequences can be for example at least 2, 20, 200, 2,000, 20,000, 50,000, 100,000, 200,000 or more.

[0067] In some embodiments of the invention the all-DNA first primer is a tailed primer, for example, having a tailed sequence (A) that generally does not hybridize to the target nucleic acid. The use of a tailed all-DNA first primer having, for example, sequence (A) allows, for example, the preparation of amplified product having either sequence (A) or its complement (A') at one end.

[0068] In some embodiments of the invention, the all-DNA first primer is attached to a solid surface, or solid substrate such as a microarray or a bead. The all-DNA first primer can be attached, for example, to a microarray, such as an oligonucleotide microarray, such as a microarray described by Affymetrix, Agilent, or NimbleGen. The different regions of the microarray can each have different all-DNA primer sequences that can hybridize to different regions of a target nucleic acid sample. These primer sequences can then be extended by the methods of the invention to produce substrates for SPIA amplification. The use of solid phase-bound all-DNA primers to produce SPIA substrates is useful for multiplex amplification of a subset of sequences within a target DNA population. For example, the solid phase or microarray may comprise primers that can hybridize to a subset of genes in a sample that represents the genome of an organism or a population of organisms. This method results in the SPIA amplification of only that subset of genes or factions thereof. The SPIA amplified subset of genes or fraction thereof, can then be analyzed, for example, by sequencing, for example by the newly developed massively parallel high throughput next generation sequencing methods and instruments such as the sequencing methods described by Illumina, 454 (Roche), and Applied Biosystems (ABI) and in US Patent Nos. 7,211,390; 6,355,431; 5,750,341; 5,969,119; 6,274,320; 6,258, 568; 6,210, 891; and 6,306,597 herein incorporated by reference in their entirety. These same methods can be applied to RNA, for example to SPIA amplification of a subset of the messenger RNA transcripts from the transcriptome of an organism. The same type of parallel multiplex amplification described for microarrays can be carried out on beads by having a plurality of beads containing all-DNA primers. The beads can be used in the same manner as above, and in some cases may be useful for clonal amplification and/or archiving of sequences in the sample.

[0069] When the all-DNA primers are bound to a solid surface, in some embodiments, the primers have a tail sequences, for example, sequence (A), that are 5' to the primer sequences. By incorporating a tailed sequence, substrates for SPIA amplification and SPIA amplification products having either an (A) sequence, or the complementary (A') sequence can be produced. The generation of amplified product comprising as sequence (A) or (A'), e.g. a universal sequence, can be useful for the further processing and/or analysis of the sample. For example, the universal sequence can be used for sequencing, e.g. as the priming site for sequencing by polymerization and pyrosequencing.

IDT Exhibit 1020.3424

[0070] In some cases, the primer bound to the solid surface is bound covalently. In some cases, the primer is bound via a spacer or a linker. In some cases, the primer is bound by a spacer having a cleavable group, where the group can be cleaved, for example chemically, by heat, or by light.

[0071] One aspect of the invention comprises using an all-DNA first DNA primer having a 5' tail sequence (A), and using a second chimeric RNA-DNA primer having a 5' RNA tail sequence (A). Extension of the second primer along the first primer extension product produces a second primer extension product having a 5' RNA portion with sequence (A), and a 3' DNA end comprising sequence (A') complementary to (A). Denaturing and annealing can then result in the formation of stem-loop structures by intramolecular hybridization of the (A) and (A') sequences. The reaction mixture thus produced can then be treated with an exonuclease specific for single stranded DNA, and with an RNase that is specific for single and or double stranded RNA such as, for example, RNase 1. This process may result in the degradation of the primers and other products that do not form the stem-loop structure. The products that do form stem-loop structures may not be degraded by the enzymes. The stem-loop products can act as SPIA substrates by using a chimeric SPIA amplification primer that anneals to the (A') DNA sequence of the stem-loop product. Cleavage of the RNA portion of the RNA-DNA heteroduplex stem structure of the stem-loop products are cleaved by RNase H thus freeing the priming site for binding of the SPIA amplification chimeric primer and enabling isothermal linear SPIA amplification. Thus, this procedure results in the clean-up of the reaction mixture from the undesired products as well as target nucleic acid and unincorporated primers, for subsequent amplification, which can result in a purer amplification products and or a better yield of amplification products. The method of using the stem-loop can also be used with embodiments that employ a solid substrate as described herein.

Exemplary Methods of the Invention

Initial steps for generation of SPIA substrate using all DNA first primer

[0072] In one aspect of the present invention, target nucleic acids are amplified by SPIA using an all DNA first primer. The all-DNA first primer may comprise a 5'-tail sequence, e.g. sequence (A), and a 3'-annealing sequence (P) as described previously. In other cases, the primer comprises an annealing sequence and no tail sequence. The annealing sequence may comprise a random sequence suitable for binding to any or substantially any sequence portion of the target nucleic acid such as for example random hexamers or random decamers. In other cases, the annealing sequence may be designed to hybridize to a specific region of the target nucleic acid.

[0073] In some embodiments the target nucleic acid comprises DNA. In some cases, the DNA comprises genomic DNA such as for example the human genomic DNA from a subject or set of subjects. In other cases, the DNA comprises a subset of a genome. In still other cases, the target DNA comprises cDNA generated from an RNA source. In some cases, the target nucleic acid comprises RNA. In some cases, the RNA comprises messenger RNA or total RNA. In some cases, the RNA comprises a transcriptome or a subset of a transcriptome.

[0074] In some embodiments, a set or sets of more than one unique non-random all DNA first primer is used such that multiple regions of the target nucleic acid are amplified. This is often referred to in the art as multiplex amplification because multiple different amplification products are produced in a single reaction. The set or sets of target-specific primers of the invention may comprise a set of related sequences for hybridizing to a set of related genes, introns, exons, splice variants and the like. For example, the annealing sequences may

IDT Exhibit 1020.3425

hybridize to nucleic acid encoding serine/threonine kinases or their complement. In some cases, the annealing sequences may be randomized at certain positions and specific at other positions, such that they may hybridize to more than one member of a set of homologous target nucleic acids.

[0075] The all-DNA first primer may be annealed to a target nucleic acid and extended using a polymerase to generate a first primer extension product. In some cases, the polymerase may comprise no, or substantially no, strand displacement activity, for example to avoid whole genome or transcriptome amplification. In other cases, a polymerase may be chosen that possesses a substantial amount of strand displacement activity, for example to enable whole genome or transcriptome amplification. In some cases, a substantially thermostable polymerase that is maximally active above about 60°C such as Taq polymerase for example may be used, for example to enable higher processivity or avoid problems with secondary structure within target nucleic acid. In other cases, a polymerase that is slightly thermostable that is maximally active between 40°C and 60°C may be used. In still other cases, a mesophilic polymerase that is maximally active below 40°C may be used to extend the first and or second primer. In some cases, an RNA-dependant DNA polymerase may be used, while in other cases a DNA dependant DNA polymerase may be used. In some cases, it may be advantageous to use a polymerase that does not exhibit either 3'-5' or 5'-3' exonuclease activity. In other cases 3'-5' or 5'-3' exonuclease activity may be advantageous. In some cases, the polymerase used to generate the first primer extension product may be a reverse transcriptase. In some cases, the target nucleic acid may be removed after generation of the first primer extension product by heat or chemical denaturation and or degradation. In other cases, the target nucleic acid may be removed by enzymatic methods. In some cases the DNA polymerase comprises both DNA-dependent DNA polymerase activity and RNA-dependent RNA polymerase.

[0076] In some embodiments of the present invention, target nucleic acids are amplified using a chimeric DNA-RNA second primer. The chimeric second primer may comprise an RNA segment and a DNA segment. In some cases, the RNA segment is generally at the 5' end and the DNA segment is generally at the 3' end of the chimeric primer. In some cases, a portion of the RNA segment comprises a tail sequence that may comprise the same or substantially the same sequence (A) as the tail sequence of the all DNA first primer. In some cases, the tail sequence comprises a different tail sequence (B) than the tail sequence (A) of the all DNA first primer.

[0077] In some cases, a portion of the DNA segment of the chimeric second primer comprises a target nucleic acid annealing sequence. In some cases, the target nucleic acid annealing sequence is random such that it may hybridize with any, or substantially any, region of the target nucleic acid. In other embodiments, the annealing sequence is specific for a region of the target nucleic acid. In still other embodiments, a set of chimeric DNA-RNA second primers comprising a set of specific annealing sequences may be utilized to amplify a set of target nucleic acid regions in a multiplex fashion.

[0078] In some cases the chimeric DNA-RNA second primer may be annealed to a target nucleic acid or to the first primer extension product and extended using a polymerase to generate a second primer extension product. The polymerase may be a thermostable or mesophilic DNA dependant DNA polymerase as described above. Additionally, the polymerase may or may not exhibit substantial strand displacement or exonuclease activity.

[0079] In one aspect of the present invention, methods are provided for creating an RNA-DNA heteroduplex, for example on a nucleic acid, such as a target nucleic acid. The RNA-DNA heteroduplex may be useful, for example in providing a substrate for SPIA amplification, or for providing a defined 3', 5' or 3' and 5' end

IDT Exhibit 1020.3426

to the nucleic acid. The method may comprise providing a template nucleic acid in a suitable reaction mixture as provided herein, adding an RNA-DNA chimeric primer, adding a DNA polymerase or allowing a DNA polymerase previously present in the reaction mixture to extend the RNA-DNA chimeric primer, and adding an exonuclease. In some cases, the method may further comprise purifying the resulting product or products by for example standard nucleic acid purification techniques known in the art such as Agencourt magnetic beads. In some cases, the method may further comprise amplifying the product by adding an RNA-DNA chimeric amplification primer, RNase H, and a DNA-polymerase exhibiting substantial strand displacement activity.

[0080] A schematic example of an embodiment of the invention relating to the initial steps of a method for generating amplified DNA by way of producing a substrate for single primer isothermal linear amplification (SPIA) is shown in **Figure 1**. Step I shows the annealing of an all DNA first primer comprising a 5'-tail of sequence (A) and a 3'-target specific, or randomized, sequence for annealing to a target nucleic acid (P). In some embodiments it may be advantageous for tail sequence (A) to be designed to minimize hybridization with the target nucleic acid. Further, in some embodiments, the tail sequence of the first primer comprises only a portion of the 5' end of the primer. Likewise, in some embodiments, the sequence (P) comprises only a portion of the 3' end of the first primer.

[0081] In one aspect, a single first primer with a specific annealing sequence (P) may be used to amplify a target sequence. In other embodiments, a set of first primers each with a different specific annealing sequence may be used to amplify multiple target sequences in the same reaction. It is anticipated that the methods of the present invention may be use to amplify many specific target sequences in the same reaction in a multiplexed fashion. In some embodiments the target sequences to be amplified represent a set of genes or gene expression products (e.g. RNA, or mRNA), or fragments thereof, of interest. Such sets of genes or gene expression products may include but are not limited to kinase, phosphatases, G-protein coupled receptors, proteases, cytokines, genes or gene expression products related to a specific metabolic or signaling pathway, genes or gene expression products related to multiple metabolic or signaling pathways, genes or gene expression products related to cancer, hereditary disease, metabolic disorder, ageing and age related disorders, normal or pathogenic specific developmental stages, specific differentiation stages, tissue specific, or any gene, gene expression product, or fragment thereof; or combination of genes, gene expression products, or fragments thereof of interest. It is anticipated that the number of target sequences amplified in this fashion can be as few as 1 to as many as tens of thousands including any number between one and 99,999, or more than tens of thousands of target sequences. In some embodiments 2, 3, 4, 5, 6, 7, 8, 9, 10 – 20, 20-30, 30-50, 50-80, 80-100, 100-200, 200-400, 400-500, 500 – 1,000, 1,000 – 2,000, 2,000-4,000, 4,000-5,000, 5,000-10,000, 10,000-20,000, 20,000-50,000 or more than 50,000 target sequences can be amplified in multiplex fashion. In some cases, at least 2; 20; 200; 2,000; 20,000; 200,000; 2,000,000 or more sequences may be amplified in multiplex fashion using the methods of the present invention.

[0082] The annealing sequence (P) can act as a primer for a DNA polymerase to extend the first primer as also depicted in step I to create a first primer extension product. In some embodiments, the target nucleic acid comprises DNA and the DNA polymerase is a DNA dependant DNA polymerase. In other embodiments, the target nucleic acid comprises RNA and the DNA polymerase is an RNA dependant DNA polymerase or a reverse transcriptase. In some embodiments, especially wherein the annealing sequence (P) of the first primer is randomized, it may be advantageous to use a DNA polymerase that does not exhibit strand displacement activity to avoid whole or substantially whole genome or transcriptome amplification. In

IDT Exhibit 1020.3427

other embodiments, especially wherein randomized annealing sequences are used, it may be advantageous to use a DNA polymerase that exhibits strand displacement activity to provide whole or substantially whole genome or transcriptome amplification. In some cases, it may be advantageous to use a thermostable polymerase, in other cases it may be advantageous to use a thermolabile polymerase. In some cases it is advantageous to use polymerases devoid of 5'-exonuclease activity. In some cases, a polymerase devoid of 5'-exonuclease activity may be useful to avoid cleaving or degrading the 5'-tail of a first or second primer or the displaced primer extension product or products.

[0083] Step II shows the separation of the target nucleic acid from the newly synthesized primer extension product followed by annealing of a second primer. In the case of a DNA target, target nucleic acid may be separated from the first primer extension product by heat or chemical denaturation. In the case of an RNA target, target nucleic acid may be degraded by heat, an enzyme, or an enzyme combination such as RNaseH and RNase 1. Alternatively, RNA target nucleic acid may be removed by heat or chemical denaturation.

[0084] The second primer as shown in step II may contain a 5' tail sequence (A) that is the same or substantially the same as the tail sequence of the first primer, or may contain a tail sequence (B) that is substantially different than the tail sequence of the second primer. The term substantially the same and substantially different as used herein describes two or more nucleic acid sequences that are substantially the same if each are capable of hybridizing to the other's complimentary sequence, or different if each do not show significant hybridization to the other's complimentary sequence. For example, a sequence substantially similar to sequence (A) would be capable of hybridizing to the complement of sequence (A) also known as (A'). In some cases, it may be advantageous to design the tail sequence such that it does not hybridize to the target nucleic acid or to the first primer extension product.

[0085] The second primer may also contain an annealing sequence (P) that is either random or target specific. In some embodiments the use of a set of second primers with multiple target specific annealing sequences is anticipated. As described previously for sequence (P) of the first primer, the simultaneous amplification of multiple genes or gene expression products or fragments thereof is anticipated by this method using multiple target specific annealing sequences or a random annealing sequence.

[0086] In the embodiment shown in **Figure 1**, the second primer is a DNA-RNA chimeric primer in which the 5'-end is RNA and the 3'-end is DNA. In some cases, the portion of the 5'- end of the second primer that is RNA corresponds to the portion of the 5'-end that comprises the tail sequence (A) or (B). In other cases, a portion of the tail sequence of the second primer is RNA and a portion of the tail sequence is not RNA, for example a portion of the tail sequence may be DNA. Likewise, in some cases the 3'-end of the second primer that is DNA corresponds to the annealing sequence (P). In other cases, a portion of the annealing sequence may be RNA.

[0087] The annealed second primer may then be extended, as depicted in step III, by the use of a DNA polymerase to create a second primer extension product. In some cases it may be advantageous to use a polymerase that does not exhibit strand displacement activity to avoid whole or substantially whole genome or transcriptome amplification. In other cases, it may be advantageous to use a polymerase that does exhibit strand displacement activity to enable whole or substantially whole genome or transcriptome amplification. The double stranded nucleic acid product such as that depicted in **Figure 1** can be a useful product or, in the case of multiplexed primers or random primers, can be a useful set of products. The products have defined double stranded region (A)/(A') at one end, when the all-DNA first primer comprises a tail sequence (A), and the second primer extension product strand has a tailed region comprising sequence (A)

IDT Exhibit 1020.3428

or (B), and comprises an RNA tailed sequence. The double stranded nucleic acid product or products can, for example, be amplified as described below. The products can be analyzed before or after amplification, for example, to determine the sequence of portions of the template nucleic acid.

Amplification of target nucleic acid using all-DNA first primer

[0088] The double stranded nucleic acid product of the method depicted in **Figure 1** may be further manipulated and amplified by single primer isothermal linear amplification. In some embodiments, a 3'-exonuclease specific for single stranded DNA such as exonuclease 1 or a polymerase that exhibits substantial single stranded DNA 3'-exonuclease activity such as T4 DNA polymerase can be used to degrade any remaining non hybridized 3'-DNA of the first primer extension product. The first primer extension product may then be extended with a polymerase that exhibits substantial reverse transcriptase activity such as for example Bst polymerase such that the first primer extension product is extended along the tail of the second primer and the first primer extension product is substantially hybridized to the tail sequence of the chimeric DNA-RNA second primer. The first and second primer extension products may form a double stranded nucleic acid with a chimeric DNA-RNA heteroduplex at one end. The double stranded nucleic acid with a chimeric DNA-RNA heteroduplex may comprise a SPIA substrate. The SPIA substrate may be amplified using the SPIA method. By way of example, the double stranded nucleic acid may be contacted with a chimeric amplification primer, an RNAse such as RNAse H that can degrade the RNA portion of a DNA-RNA heteroduplex, and a polymerase with strand displacement activity. The chimeric amplification primer may comprise a DNA portion and a 5'-RNA portion wherein the chimeric amplification primer comprises a sequence which is substantially the same as sequence (A) or (B) or a portion of sequence (A) or (B).

[0089] In some embodiments a double stranded nucleic acid product comprising a hybridized first and second primer extension product as depicted in **Figures 1 and or 2**, may be amplified, or a region of the first and second primer extension product may be amplified, by a method known to the art that is not SPIA. The amplification methods include but are not limited to PCR, linear PCR, SPIA, rolling circle amplification, loop-mediated isothermal amplification, ligation mediated rolling circle amplification and the like.

[0090] The formation of a substrate for single primer isothermal linear amplification (SPIA) from the double stranded nucleic acid product of the method depicted in **Figure 1** is depicted in **Figure 2**. In **Figure 2**, step I shows the use of a single stranded DNA specific 3'-exonuclease such as, but not limited to, exonuclease 1 or the 3'-exonuclease activity of a DNA polymerase such as, but not limited to, T4 DNA polymerase. The exonuclease or exonuclease activity may remove the single stranded 3'-nucleotides from the 3'-end of the first primer extension product which are not hybridized to the second primer extension product.

[0091] The method depicted in **Figure 2** further comprises step II whereby a DNA polymerase with reverse transcriptase activity (e.g. Bst polymerase) further extends the first primer extension product to produce a sequence (A') or (B') complimentary to the tail sequence of the second primer (A) or (B). This extension produces a double stranded DNA product with two defined ends, referred to as the first primer end in which the first primer is incorporated and the second primer end in which the second primer is incorporated.

[0092] The method further comprises step III, whereby SPIA is used to amplify the further extended first primer extension product of step II. In some embodiments of the SPIA method, an RNA-DNA heteroduplex degrading enzyme such as, but not limited to, RNase H is used to remove the RNA portion of the second primer in the RNA-DNA heteroduplex. Then a third primer, referred to as the amplification primer, is annealed to the 3' single stranded portion of the first primer extension product. The third primer, the

IDT Exhibit 1020.3429

amplification primer, comprises a 5'-RNA portion and a 3'-DNA portion. A DNA polymerase with strand displacement activity such as has been described for example in U.S. Pat. Nos. 6,251,639 and 6,692,918 may then extend the primer to create a third primer extension product while displacing the second primer extension product. The enzyme additions to carry out step III may be done sequentially and repeatedly to produce SPIA products, or preferably the RNase H, amplification primer, and strand displacing polymerase are added as a mixture and the reaction allowed to proceed under conditions that allow amplification as described in US Patent Nos. 6,251,639 and 6,692,918. The amplification products of the SPIA method as described in **Figure 2** all have a 3' end sequence (A') that is the compliment of the first primer tail sequence (A), when the all DNA first primer is a tailed primer comprising sequence (A).

Stem loop nucleic acid structures for SPIA amplification

[0093] Alternatively, in some embodiments, the double stranded nucleic acid product of **Figure 1** may be used to generate amplified nucleic acid by way of a different method than depicted in **Figure 2**. In this method, the chimeric DNA-RNA second primer may comprise a tail sequence (A) that is the same or substantially the same as the tail sequence (A) of the all DNA first primer. The primer annealing sequences may be specific, or random sequence. The double stranded nucleic acid product of **Figure 1** in which the tail sequences of the first and second primer are substantially the same comprises a second primer extension product that may be separated from the first primer extension product by heat, chemical or enzymatic means and annealed to form a stem-loop structure.

[0094] The stem loop structure may comprise a DNA-RNA heteroduplex of RNA sequence (A), and complementary DNA sequence (A'). The annealing reaction may then be treated with a single strand DNA specific 3'-exonuclease such as for example exonuclease 1, and a single stranded nucleic acid specific RNase such as for example RNase 1. The exonuclease and RNase treatment may provide a clean-up step in that a substantial portion of extraneous nucleic acids such as the target nucleic acid, and the first and second primers are degraded. The degradation products may then be removed by a number of methods known in the art for removing nucleotide degradation products from polynucleic acids such as size exclusion spin columns, ion-exchange resin, ultrafiltration, dialysis, or any method known to the art. In some cases, removal of degradation products may not be required. In other cases, additional steps may be used to remove other undesirable compounds from the stem loop structure such as pyrophosphate and the like. In still other cases, the exonuclease and RNase may be inactivated by heat or chemical denaturation.

[0095] The resulting stem loop structure, for example, the purified stem loop structure may then be amplified by SPIA by first degrading the RNA portion of the DNA-RNA heteroduplex with for example RNase H, thus yielding a substantially linear nucleic acid comprising a sequence A' at the 3'-end. Amplification further comprises providing a chimeric amplification primer and a DNA polymerase with strand displacement activity. The chimeric amplification primer may comprise a DNA portion and a 5'-RNA portion wherein the chimeric amplification primer comprises a sequence which is substantially the same as sequence (A) or a portion of sequence (A).

[0096] Alternatively, the stem loop structure may be amplified by any other method of nucleic acid amplification known to the art such as but not limited to rolling circle template amplification, polymerase chain reaction (PCR), and linear polymerase chain reaction.

[0097] **Figure 3** depicts an embodiment for producing amplification products by the use of the SPIA method by way of a stem-loop intermediate product. Here, the tailed regions of both the first primer and the second

IDT Exhibit 1020.3430

primer comprise the sequence (A). In this case the hybridized first and second primer extension products of **Figure 1** are denatured, for example, by heat or chemical denaturation and subsequently are allowed to anneal. The result of having the same tailed sequence for the first and second primers is that the 3'-end and 5'-end of the second primer extension product comprises complimentary or substantially complementary sequences such that they are capable of hybridization to one another to form a stem-loop structure. Hybridization to form the stem –loop structures is carried out under conditions that are readily determined by one skilled in the art.

[0098] The annealing step can be carried out in such manner that the unimolecular nature of the self annealing of the second primer extension product is preferred over renaturation of the double stranded product of the method of **Figure 1**. The self annealed product thus forms a stem loop type structure. In some cases the length of time or rate of the annealing process, the addition of DNA binding proteins or enzymes, the use of chaotropic or cosmotropic solutes, or the concentration of first and primer extension products may be varied in order to promote stem-loop structure formation. Self annealing may be favored for certain sizes of the second primer extension product. In some cases, self annealing is favored by having a second primer extension product size that is smaller than about 100 kilobases or about 10 kilobases. In some cases, self annealing is favored for second primer extension product having sizes that are smaller than about 1 kilobase. In other embodiments, self annealing is favored for second primer extension product sizes are smaller than about 500 bases or between 10 and 500 bases. Where the stem-loop formation does occur for a given size range, this property can be used in order to select for that size range at the exclusion of the others, as the products which do not form the stem-loop structure may be susceptible to degradation by the enzymes described above. In some cases, the process selects for second primer extension products less than about 50, 80, 100, 200, 400, 500, 800, 1,000, 2,000, 4,000, 5,000, 8,000, 10,000, or 20,000 bases.

[0099] Step II shows the process of degradation of extraneous polynucleotides such as the target nucleic acid, the first primer extension product, and unincorporated primers. Degradation of the reaction components described above, is accomplished by using one or more enzymes that degrade single stranded nucleic acids, and to not degrade or do not substantially degrade double-stranded nucleic acids. Extraneous DNA polynucleotides may be removed by digestion with 3'-single stranded DNA specific exonucleases. Since sequences (A) and (A') of the second primer extension product are hybridized to form a double stranded portion, the stem-loop products are not degraded by these enzymes. Extraneous RNA polynucleotides may be removed by digestion with RNase 1. The combination may be used to remove extraneous chimeric primers such as the second primer depicted in **Figure 1.**

[00100] Step III shows a method for removing the RNA from the stem loop structure created in step I. RNase H is used to degrade the RNA portion of the DNA-RNA heteroduplex formed by the stem loop structure. This leads to an essentially linear polynucleic acid with a defined 3'- sequence (A'). The step of degradation of the RNA in the RNA-DNA heteroduplex by RNase H may be carried out in the reaction mixture for carrying SPIA amplification of the primer extension product as described below.

[00101] In step IV, SPIA amplification is performed in a manner as described in US Patent Nos. 6,251,639 and 6,692,918 using a mixture comprising a chimeric amplification primer that hybridizes to the A' sequence of step III, a strand displacing DNA polymerase, and RNase H.

[00102] In an alternative step not depicted, a chimeric oligonucleotide may be employed to allow the generation of amplified DNA with a defined 5' end. The suitable chimeric oligonucleotide may comprise a 3'-DNA sequence (A) capable of hybridizing with the (A') sequence of the linear polynucleic acid product of step

IDT Exhibit 1020.3431

III. The chimeric oligonucleotide may also comprise a 5'-RNA tail sequence (C). The linear polynucleic acid is extended with an RNA dependent DNA polymerase creating an RNA DNA heteroduplex of DNA sequence (C') and RNA sequence (C) which may be amplified by the SPIA method using a mixture comprising RNase H which cleaves RNA tail sequence C, an amplification primer that hybridizes to DNA tail sequence C', and a suitable polymerase. The amplified DNA products may then comprise a defined sequence (A) at or near its 5'-end.

Capture of target sequence on immobilized first primers

[00103] In some embodiments of the present invention the first primer is immobilized on a surface. The surface may be a chip or a slide such as for example a gene chip or a microarray. In other cases, the surface may comprise a bead such as for example an agarose bead, a polymeric bead, polystyrene beads, or a magnetic bead. In still other cases, the surface may be any surface known to the art capable of being derivatized with oligonucleotides. In some cases, the first primers are immobilized to discrete sites. In some cases, the discrete sites comprise immobilized primers with a specific known annealing sequence or set of sequences such that target nucleic acids of known sequence are immobilized at known discrete sites. The immobilized first primer, or primers, may be useful for capturing a targeted nucleic acid sequence (e.g. RNA or DNA) for subsequent amplification or analysis. The subsequent amplification may comprise SPIA, PCR, rolling circle template amplification or linear PCR. The immobilization may be performed by a number of methods known to the art. For example, the first primer, or primers, may be covalently linked to the surface by a photolabile, heat labile, reduction labile, oxidation labile, acid labile, or base labile linker. Alternatively, the immobilized first primer, or primers, may be hybridized to another nucleic acid or nucleic acid analog such as a locked nucleic acid, pyroanosyl nucleic acid, or peptide nucleic acid that is covalently linked to the surface. The immobilized first primer, or primers, may comprise a linker, as described in the following.

[00104] In some embodiments, because the captured target sequences that are produced are isolated as discrete populations, subsequent clonal amplification products can be generated and analyzed, for example sequenced, in a parallel manner. Therefore, in some embodiments, at least at least 100, 500, 1000, 10000, 50000, 100000, 300000, 500000, or 1000000 populations of clonal amplification products can be analyzed in parallel. The skilled artisan will appreciate that various methods can be suitable for parallel analysis of clonal amplicons. Generally, such methods can produce a discrete detectable signal that can be associated or linked to individual populations of clonal amplicons.

[00105] In some embodiments the target nucleic acid comprises DNA. In some cases, the DNA comprises genomic DNA such as for example the human genomic DNA from a subject or set of subjects. In other cases, the DNA comprises a subset of a genome. In still other cases, the DNA comprises cDNA generated from an RNA source. In some cases, the target nucleic acid comprises RNA. In some cases, the RNA comprises messenger RNA or total RNA. In some cases, the RNA comprises a transcriptome or a subset of a transcriptome.

[00106] In one aspect of the present invention the first primer comprises an immobilized all-DNA first primer. The first primer may comprise an annealing sequence (P), and a tail sequence (A). The annealing sequence may be random such that it is capable of hybridizing to substantially any portion of the target nucleic acid. The random annealing sequences may comprise for example random hexamers or random decamers. The annealing sequences alternatively may comprise specific sequences for hybridizing to a specific region of

IDT Exhibit 1020.3432

the target nucleic acid. In some cases, the annealing sequence of first primers comprise a set of specific target annealing sequences, wherein each first primer of the method comprises a specific target annealing sequence, such that a specific set of target nucleic acid regions are captured and or subsequently amplified. The sets may comprise a set of related sequences for hybridizing to a set of related genes, introns, exons, splice variants and the like. For example, the annealing sequences may hybridize to nucleic acid encoding serine/threonine kinases or their complement. In some cases, the annealing sequences may be randomized at certain positions and specific at other positions, such that they may hybridize to more than one member of a set of homologous target nucleic acids.

[00107] The immobilized all DNA first primer may be annealed to a target nucleic acid and extended using a polymerase to generate a first primer extension product. In some cases, the polymerase may comprise substantially no strand displacement activity. In other cases, a polymerase may be chosen that possesses a substantial amount of strand displacement activity. In some cases, a substantially thermostable polymerase that is maximally active above about 60°C such as Taq polymerase for example may be used, for example to enable higher processivity or avoid problems with secondary structure within primers and or target nucleic acid. In other cases, a polymerase that is slightly thermostable that is maximally active between 40°C and 60°C may be used. In still other cases, a mesophilic polymerase that is maximally active below 40°C may be used to extend the first and or second primer. In some cases, an RNA-dependant DNA polymerase may be used, while in other cases a DNA dependant DNA polymerase may be used. In some cases, it may be advantageous to use a polymerase that does not exhibit either 3'-5' or 5'-3'-exonuclease activity. In other cases 3'-5' or 5'-3'-exonuclease activity may be advantageous. In some cases, the polymerase used to generate the first primer extension product may be a reverse transcriptase. In some cases, the use of a polymerase with high processivity may be advantageous. In some cases, the target nucleic acid may be removed after generation of the first primer extension product by heat or chemical denaturation and or degradation. In other cases, the target nucleic acid may be removed by enzymatic methods.

[00108] In some embodiments of the present invention, target nucleic acids are amplified using a chimeric DNA-RNA second primer. The chimeric second primer may comprise an RNA segment and a DNA segment. In some cases, the RNA segment is generally at the 5' end and the DNA segment is generally at the 3' end of the chimeric primer. In some cases, a portion of the RNA segment comprises a tail sequence that may comprise the same or substantially the same sequence (A) as the tail sequence of the all DNA first primer. In some cases, the tail sequence comprises a different tail sequence (B) than the tail sequence (A) of the all DNA first primer.

[00109] In some cases, a portion of the DNA segment of the chimeric second primer comprises a target nucleic acid annealing sequence. In some cases, the target nucleic acid annealing sequence is random such that it may hybridize with any, or substantially any, region of the target nucleic acid. In other embodiments, the annealing sequence is specific for a region of the target nucleic acid. In still other embodiments, a set of chimeric DNA-RNA second primers comprising a set of non-random annealing sequences may be utilized to amplify a set of target nucleic acid regions in a multiplex fashion as described previously.

[00110] In some cases the chimeric DNA-RNA second primer may be annealed to a target nucleic acid or to the immobilized first primer extension product and extended using a polymerase to generate an immobilized second primer extension product. The polymerase may be a thermostable or mesophilic DNA dependant

IDT Exhibit 1020.3433

DNA polymerase as described above. Additionally, the polymerase may or may not exhibit substantial strand displacement or error correcting activity.

[00111] **Figure 4** depicts a method for using DNA primers immobilized on a solid substrate that can capture of target nucleic acid from solution. Step I shows a surface to which all DNA first primers are immobilized. The first primers comprise a DNA tail sequence (A), and a DNA target nucleic acid annealing sequence (P) as described above. While the solid substrate is depicted in **Figure 4** as a spherical particle, it is to be understood that the method depicted can be carried out on a solid substrate of any suitable geometry, including wells, planar surfaces, and other geometries described herein. The surface may comprise, for example a particle or bead, a nanoparticle, a microarray, a glass slide, a chip or cassette, a polynucleic acid chip, a fused fiber optic array, a polymer, or any type of surface known to the art to which nucleic acid, such as an oligonucleotide may be attached.

[00112] The method of attachment may or may not require a linker element between the first primer and the surface. The linker element may be a spacer element such as by way of example ethylene glycol, polyethylene glycol, polyvinyl, polystyrene, polyether, a saturated carbon chain, an unsaturated carbon chain, a nucleic acid sequence, an amino acid sequence or any other spacer element capable of linking a nucleic acid to a surface known to the art. The spacer element may also comprise a cleavable element for removal of immobilized first primers from the surface. The cleavable spacer element may be cleaved through the action of a strong base, a strong acid, heat, light, reduction, oxidation, enzymatic reaction, or any of a variety of other methods known to those skilled in the art.

[00113] Step I also shows the annealing of a target nucleic acid to the immobilized first primer. The target nucleic acid may comprise RNA or DNA. The target nucleic acid may comprise fragments from an RNA or DNA sample. The annealing may occur between the sequence (P) of the first primer and a sequence on the target nucleic acid. The sequence on the primer that anneals to the target nucleic acid may be a random sequence, a target specific sequence, or a set of target specific sequences. In some embodiments a plurality of specific primer sequences are used, wherein, for example, different primer sequences are immobilized on different regions of a microarray or different sequences are immobilized on different beads.

[00114] In step II, the first primer is extended by the action of a polymerase to create a first primer extension product. In the case of a DNA target nucleic acid, it may be advantageous to use a DNA dependent DNA polymerase. In the case of an RNA target nucleic acid, it may be advantageous to use an RNA dependent DNA polymerase or reverse transcriptase. In some embodiments it may be advantageous to utilize a polymerase that does or does not exhibit substantial strand displacement activity. It may also be advantageous to extend the first primer with a thermostable or a thermolabile polymerase. The thermostable polymerase may exhibit higher rates of polymerization, fewer difficulties with secondary structure of target nucleic acid or first primer, or greater fidelity. The thermolabile polymerase may exhibit greater fidelity, greater rates of polymerization, or enable facile removal of polymerase activity by heat denaturation. In some cases, the use of other nucleic acid binding proteins, or reaction additives, during the extension step is anticipated. The proteins or additives include but are not limited to DMSO, DMF, cations, anions, glycerol, single-stranded nucleic acid binding proteins including but not limited to T4 gene 32 protein (gp32), buffers, and salts.

[00115] The target nucleic acid and the first primer extension product are then separated by heat, chemical, or enzymatic treatment, and a second DNA-RNA chimeric primer is then annealed to the first primer extension product in step III. The primer may comprise a 5' tail sequence (A) or (B) and an annealing

IDT Exhibit 1020.3434

sequence (P) as described previously. The annealed second primer then may be used to extend the primer to create a second primer extension product complementary to the entire, or a portion of the entire, first primer extension product as depicted in step IV. As described for extension of the first primer, it is anticipated that for certain aspects of the present invention, the DNA polymerase used to create the second primer extension product may be thermostable, thermolabile, strand displacing, non strand displacing, high fidelity, highly processive, or any combination thereof.

Generation of amplified DNA from captured target sequences

[00116] In some embodiments of the present invention, the immobilized double stranded nucleic acid product depicted in **Figure 4** may be used to generate amplified DNA. Amplified DNA may be generated by a number of methods known to the art. Preferred amplification methods include but are not limited to PCR, SPIA, rolling circle amplification, loop-mediated isothermal amplification, ligation mediated rolling circle amplification and the like. In the case of SPIA amplification, the method of amplification may further comprise providing a single stranded DNA specific 3'-exonuclease such as exonuclease 1, or a polymerase with 3'-exonuclease activity such as for example T4 DNA polymerase. The exonuclease or exonuclease activity may degrade the 3'-portion of the first primer extension product that is not hybridized to the second primer extension product. The first primer extension product may then be extended with a polymerase such as Bst that exhibits reverse transcriptase activity. The extension may generate a DNA-RNA heteroduplex at one end. The immobilized first and second primer extension products comprising a DNA-RNA heteroduplex at one end may be amplified by SPIA by providing an RNase such as RNaseH to cleave the RNA portion of the heteroduplex, a chimeric amplification primer, and a polymerase with strand displacing activity. The chimeric amplification primer may comprise a DNA portion and a 5'-RNA portion wherein the chimeric amplification primer comprises a sequence which is substantially the same as sequence (A) or (B) or a portion of sequence (A) or (B). The amplified nucleic acid may comprise a sequence tail (A') at the 3'-end.

[00117] A method of generating amplified DNA using SPIA from the immobilized first and second primer extension products depicted in **Figure 4** is shown in **Figure 5**. This method generates amplified product directly from an extension product of the primer immobilized to the solid substrate. Step I of **Figure 5** depicts the use of a single stranded DNA specific 3'-exonuclease such as exonuclease 1 or the 3'-exonuclease activity of a DNA polymerase such as T4 DNA polymerase to remove the portion of the 3'-end of the first primer extension product that is not hybridized to the second primer extension product. In step II, the immobilized first primer extension product or first primer extension products are then extended by a DNA polymerase that exhibits reverse transcriptase activity. A DNA polymerase that exhibits reverse transcriptase activity includes but is not limited to Bst polymerase. Step III depicts the use of RNase H to cleave the RNA portion of the DNA-RNA heteroduplex. This provides a site for annealing an amplification primer to the priming site A' or B', or a portion of site A' or B', as depicted in step IV. Thus amplified DNA products with a defined 3'-end sequence (A') may be generated by SPIA methods using a reaction mixture comprising RNase H, a DNA polymerase with strand displacement activity, and a chimeric amplification primer.

IDT Exhibit 1020.3435

Generation of amplified DNA from immobilized target sequence using solution phase SPIA

[00118] In one aspect of the present invention, the immobilized double stranded first and second primer extension product depicted in **Figure 4** may be amplified by solution phase SPIA. In some cases, the double stranded nucleic acid product depicted in **Figure 4** may be treated with a single stranded DNA specific 3'-exonuclease to remove the 3'-portion of the first primer extension product that is not hybridized to the second primer extension product. The immobilized first and second primer extension product may then be cleaved from the surface releasing the first and second primer extension product into the solution phase. The cleavage may be performed by application of heat, base, acid, reducing agent, oxidation or any method known to the art. The cleavage may occur at or within the linker between the nucleic acid portion of the all DNA first primer and the surface. Alternatively, the cleavage may occur before exonuclease treatment.

[00119] The first primer extension product may then be extended using a polymerase such as Bst that exhibits substantial reverse transcriptase activity. The extension may produce a double stranded nucleic acid a portion of which comprises a DNA-RNA heteroduplex. The double stranded nucleic acid may be amplified using any method of nucleic acid amplification known to the art such as but not limited to PCR, linear PCR, SPIA, rolling circle amplification, loop-mediated isothermal amplification, ligation mediated rolling circle amplification and the like. Alternatively, the extension of the first primer extension product in the complex with the second primer extension product may be extended with a DNA polymerase comprising an RNA-dependent DNA polymerase activity, such as for example Bst polymerase, prior to the release from the surface, such that the released complex comprises an RNA-DNA heteroduplex at one end.

[00120] **Figure 6** depicts an alternative method for producing amplified DNA using SPIA from the immobilized first and second primer extension products depicted in **Figure 4.** Here, the products of the first steps of the method are cleaved from the solid surface such that the SPIA amplification occurs in solution rather than occurring on the solid surface. While the figure depicts cleavage occurring after the exonuclease step of the method, it is understood that the cleavage can occur at any suitable step before or after this step. Step I depicts a 3'-exonuclease step as described previously for degrading extraneous nucleic acids including but not limited to target nucleic acid and unincorporated primers, as well as trimming the unhybridized 3'-end of the first primer extension product. Step II depicts the use of a cleaving step to release the hybridized first and second primer extension products from the surface or particle to which they are immobilized. The cleavage mechanisms anticipated include chemical, enzymatic, or light as described previously. Following cleavage, the remaining steps are carried out in solution. Step III depicts the use of a DNA polymerase such as but not limited to Bst polymerase to extend the first primer extension product. Step IV depicts the use of RNase H to produce an amplification primer hybridization site. Step V depicts the use of SPIA to generate amplified DNA products having a defined 3'-DNA sequence (A').

Capture of target sequences for formation of stem loop strucutres

[00121] In some embodiments of the present invention an all DNA first primer comprising a tail sequence (A) immobilized to a bead or surface may be annealed to capture target nucleic acid sequences and extended, and a chimeric second primer may be annealed comprising a tail sequence (A) which is the same or substantially the same as the tail sequence of the all DNA first primer and extended. The capture and first and second primer extension steps may be performed by the methods described previously. The second primer extension products may be used further to generate a stem loop nucleic acid structure. The stem loop nucleic acid structure may be amplified by a number of methods known the art as described previously

IDT Exhibit 1020.3436

such as but not limited to PCR, linear PCR, SPIA, rolling circle amplification, loop-mediated isothermal amplification, ligation mediated rolling circle amplification and the like.

[00122] **Figure 7** depicts an alternative method for amplification of captured of target nucleic acid using immobilized primers. This method utilizes the same sequence, e.g. sequence (A) in the 5'-tail of both the first and the second primers. As described above, this method produces a product which can form a stem-loop structure, which facilitates clean-up of the desired products prior to SPIA amplification. Step I depicts the use of immobilized all DNA first primers with a 5'-tail sequence (A) and a 3'-template annealing sequence (P) for annealing to a target nucleic acid as described previously. Step II depicts the use of an RNA or DNA dependant DNA polymerase to extend the first primer and produce a first primer extension product as described previously. In step III, the target nucleic acid is removed by heat, enzymatic, or chemical means, and a DNA-RNA chimeric second primer comprising a 5'-tail sequence (A) and 3' template annealing sequence (P) as described previously. For the method depicted in this figure, it may be advantageous to ensure that the 5'-tail sequence of the second primer is identical to or substantially the same as the 5'-tail sequence of the first primer. In step IV the annealed second primer is extended with a DNA polymerase as described previously to yield a second primer extension product.

Generation of stem loop amplification substrate from captured target sequences

[00123] In some embodiments of the present invention, the second primer extension product or products of a captured target sequence as described previously, may be utilized to generate amplified DNA. The amplified DNA may be separated from the first primer extension product by heat or chemical denaturation, and annealed to form a stem loop structure. In some cases, primers and other extraneous nucleic acids may be removed from the reaction by degradation before or after separation and annealing of the second primer extension product. For example, unextended first and second primers and target nucleic acid may be degraded with a single stranded DNA specific 3'-exonuclease and or a single and or double stranded RNA specific RNase as described previously. The stem loop structure may be resistant to degradation by the exonuclease and RNase because of a lack of a single stranded 3'–DNA or RNA end. The stem loop structure may then be amplified by any methods known in the art such as but not limited to PCR, linear PCR, SPIA, rolling circle amplification, loop-mediated isothermal amplification, ligation mediated rolling circle amplification and the like.

[00124] **Figure 8** depicts a method of utilizing the products depicted in **Figure 7** for the generation of a substrate suitable for DNA amplification using the SPIA method. Step I is an optional step. In step I, a 3'-exonuclease is used as described previously to remove extraneous nucleic acids, and trim the 3'-end of the first primer extension product. In step II, the second primer extension product is separated from first primer extension product by heat or chemical means. The second primer extension product is then annealed to create a stem loop structure as described previously. Exonucleases and RNase 1 can then be added to remove extraneous polynucleotides and unincorporated primers. The stem loop structure then acts as a suitable substrate for SPIA amplification to generate amplified DNA. Alternatively, a chimeric oligonucleotide may be employed to generate amplified DNA with a defined 5'-end as described in **Figure 3** alternative step.

Amplification of RNA using a chimeric second primer

IDT Exhibit 1020.3437

[00125] In some embodiments of the present invention, amplified DNA can be produced from an RNA template using an all DNA first primer comprising a tail sequence (A) or (B) and an annealing sequence (P) as described previously for amplification of a specific sequence, multiplex amplification, or whole transcriptome amplification. In some cases, the first primer extension product is formed by providing a RNA dependant DNA polymerase. In some cases, further steps may include the removal of RNA target nucleic acid by heat, chemical, heat and chemical, or enzymatic means. For example, RNA template may be removed by providing RNase H and RNase1 to degrade any single and or double stranded RNA and any RNA portion of an RNA-DNA heteroduplex. In one aspect of the present invention a chimeric DNA-RNA second primer as described previously comprising a tail sequence (B) that is substantially different from the tail sequence (A) in the all DNA first primer may be annealed to the first primer extension product and extended as described previously to create a second primer extension product. In some cases, a single stranded DNA specific 3'-exonuclease may be provided to degrade unincorporated first and second primers as well as any unhybridized 3'-end of the first primer extension product. The degraded 3'-end of the first primer extension product may then be extended using a polymerase such as for example Bst that exhibits reverse transcriptase activity to generate a double stranded nucleic acid product comprising the first and second primer extension product and a DNA-RNA heteroduplex at one end. The first and second primer extension product may then be used to generate amplified DNA using a number of methods known to the art such as but not limited to PCR, linear PCR, SPIA, rolling circle amplification, loop-mediated isothermal amplification, ligation mediated rolling circle amplification and the like. In some cases, SPIA is used to generate amplified products comprising a defined 3'-end sequence (A').

[00126] **Figure 9** depicts a method for producing amplified DNA from an RNA template using a SPIA method using an all DNA first primer of tail sequence (A), a DNA-RNA chimeric second primer of tail sequence (B), and an RNA target nucleic acid. It is also anticipated that in some cases, a second primer tail sequence (A) may be employed. In this method, the first primer is annealed to the target in a target specific or random fashion and is extended with an RNA dependent DNA polymerase or reverse transcriptase to generate a first primer extension product. Target nucleic acid is removed by RNase H and RNase 1 as depicted, or any other RNases or combination thereof known in the art, which could be easily inactivated so as not to interfere with the following steps of the methods of the invention. It is also anticipated that other methods may be employed to remove RNA target nucleic acid selectively such as a thermal or chemical method. Following removal of target nucleic acid, the chimeric second primer comprising RNA containing tail sequence (B) is annealed to the first primer extension product and extended. A single stranded DNA specific 3'-exonuclease is then employed to trim back the unhybridized 3'-end of the first primer extension product, and a third DNA polymerase step is then used to create a double stranded hybridized product with a defined DNA-RNA heteroduplex end of DNA sequence (B') and RNA sequence (B), and a DNA-DNA duplex end of sequence (A) and (A'). This product of the method depicted in **Figure 9** may then be amplified DNA of a defined 3'-sequence (A') is then produced by the SPIA method using a mixture of RNase, amplification primer, and a suitable polymerase.

Generation of amplified DNA having a defined 5' and 3' end

[00127] In some embodiments of the present invention, a double stranded nucleic acid comprising a first and second primer extension product and further comprising a DNA-RNA heteroduplex at one end may be used to generate amplified DNA comprising a defined 5' and 3'-end. In some cases, the RNA portion of the DNA-

IDT Exhibit 1020.3438

RNA heteroduplex may be cleaved or degraded by an enzyme such as RNase H. The method may further include providing a DNA-RNA chimeric oligonucleotide. The chimeric oligonucleotide may comprise a 5'-segment and a 3'-segment. The 5'-segment may comprise RNA and the 3'-segment may comprise DNA. The 3'-DNA segment may comprise a sequence (B) the same or substantially the same as the RNA portion of the DNA-RNA heteroduplex degraded by RNase H. The 5'-RNA segment may comprise a sequence (C) that is substantially different from previously described sequences (A) and (B). The chimeric oligonucleotide may be annealed to the first primer extension product such that sequence (B) of the chimeric oligonucleotide hybridizes with sequence (B') of the first primer extension product. The first primer extension product may then be extended by a polymerase as described previously. The resulting double stranded nucleic acid comprising a first and second primer extension product and a DNA-RNA heteroduplex may then be amplified by a number of methods known to the art such as but not limited to PCR, linear PCR, SPIA, rolling circle amplification, loop-mediated isothermal amplification, ligation mediated rolling circle amplification and the like. In some cases, SPIA may be used to generate amplified products comprising a defined 3' end sequence (A') and a defined sequence at or near its 5'-end comprising sequence (B).

[00128] **Figure 10** depicts an alternative method of generating amplified DNA using the SPIA method from the product the method depicted in **Figure 9**. In this method, RNase H is used to degrade the RNA portion of the DNA-RNA (B')-(B) heteroduplex and a chimeric oligonucleotide comprising 3'-DNA sequence (B) and a 5'-RNA sequence (C) is annealed to the DNA sequence (B'). The first primer extension product is then extended using a suitable polymerase to create a double stranded nucleic product with a DNA-RNA heteroduplex at one end of DNA sequence (C') and RNA sequence (C) and a DNA-DNA duplex at the other end of sequence (A) and (A'). Following the extension step, SPIA amplification is then employed using a mixture of RNaseH, an RNA-DNA chimeric amplification primer, and a suitable polymerase as described previously. In this case, the products of the SPIA amplification have a defined 3' sequence (A') and a defined 5' sequence (B).

Generating a SPIA substrate with a ligand

[00129] In some embodiments of the present invention a double stranded nucleic acid comprising a first and second primer extension product and further comprising a ligand may be generated. The ligand may be useful for capture of the double stranded nucleic acid, purification, or further downstream manipulation such as sequencing for example. The ligand may be but is not limited to a biotin molecule or other small organic molecule like a fluorophore, a peptide, a nucleic acid, a magnetic particle, a bead, a nanocrystal or quantum dot, a protein, or any other ligand capable of being linked to a nucleic acid covalently or non-covalently. In some cases, the ligand is linked to an all DNA first primer comprising a tail sequence (A) and an annealing sequence (P) as described previously. The all DNA first primer is annealed to a target nucleic acid (e.g. RNA or DNA) and extended using a polymerase as described previously to generate a first primer extension product linked to a ligand. The first primer extension primer may comprise random sequences, a specific sequence, or a set of multiple specific sequences as described previously.

[00130] In some cases, a DNA-RNA chimeric second primer comprising a tail sequence (A) or (B) and an annealing sequence as described previously may be annealed to the first primer extension product and extended using a DNA polymerase to generate a second primer extension product. The first and second primer extension product may further be treated with a single stranded DNA specific 3'-exonuclease to

IDT Exhibit 1020.3439

generate a double stranded nucleic acid comprising the first and second primer extension products, a ligand, and a DNA-RNA heteroduplex. In some embodiments, the double stranded nucleic acid product may be amplified using a number of methods known to the art such as but not limited to PCR, linear PCR, SPIA, rolling circle amplification, loop-mediated isothermal amplification, ligation mediated rolling circle amplification and the like. In some cases SPIA may be used to generate an amplified DNA product with a 3'-end comprising a defined sequence (A').

[00131] **Figure 11** depicts a method for using a ligand attached to an all DNA first primer, a DNA-RNA chimeric second primer and an RNA template nucleic acid to generate amplified DNA products using SPIA. The method follows the method outlined in **Figure 9** with exception of the ligand attached all DNA first primer of tail sequence (A). Various nucleic acid ligands anticipated by the present invention are known to the art and include but are not limited to biotin, proteins such as antibodies, antibody fragments, avidin, streptavidin, maltose binding protein, S-protein, enzymes, peptides such as hexahistidine, myc tag, flag tag, HA tag, or any affinity tag or member of any pair of binding agents known in the art, fluorophores, particles, nanoparticles, magnetic particles, or other nucleic acid sequences. Methods for attaching ligands to nucleic acid primers are known to the art. The ligand may be used for attachment to a solid substrate, for example for archiving and clonal expansion. The ligand may be useful for functionalizing the first primer extension product for further chemical or enzymatic manipulation, or for purification of the first and or second primer extension products. The SPIA amplified products of the method depicted in this figure have a 3' end comprising the defined sequence (A').

Generating a SPIA substrate with a ligand and amplified product with two defined ends

[00132] In some embodiments of the present invention a double stranded nucleic acid SPIA substrate may be generated having a ligand and a two defined ends comprising a first and second primer extension product. The ligand may be useful for capture of the double stranded nucleic acid, purification, or further downstream manipulation such as sequencing for example. The ligand may be but is not limited to a biotin molecule or other small organic molecule like a fluorophore, a peptide, a nucleic acid, a magnetic particle, a bead, a nanocrystal or quantum dot, a protein, or any other ligand capable of being linked to a nucleic acid covalently or non-covalently. In some cases, the ligand is linked to an all DNA first primer comprising a tail sequence (A) and an annealing sequence (P) as described previously. The all DNA first primer is annealed to a target nucleic acid (e.g. RNA or DNA) and extended using a polymerase as described previously to generate a first primer extension product linked to a ligand. The first primer extension primer may comprise random sequences, a specific sequence, or a set of multiple specific sequences as described previously.

[00133] In some cases, a DNA-RNA chimeric second primer comprising a tail sequence (A) or (B) and an annealing sequence as described previously may be annealed to the first primer extension product and extended using a DNA polymerase to generate a second primer extension product. The first and second primer extension product may further be treated with a single stranded DNA specific 3'-exonuclease to generate a double stranded nucleic acid comprising the first and second primer extension products, a ligand, and a DNA-RNA heteroduplex.

[00134] The double stranded nucleic acid product may be further treated with RNase H to degrade or cleave the RNA portion of the DNA-RNA heteroduplex. In some cases, a DNA-RNA chimeric oligonucleotide comprising a 5'-RNA sequence (C) and a 3'-DNA sequence (B) as described previously may be annealed

IDT Exhibit 1020.3440

to the first primer extension product. The first primer extension product may be further extended with a polymerase exhibiting substantial reverse transcriptase activity such as for example Bst polymerase. The further extension may generate a double stranded first and second primer extension product comprising a ligand and DNA-RNA heteroduplex. In some embodiments, the double stranded extension product may be amplified using a number of methods known to the art such as but not limited to PCR, linear PCR, SPIA, rolling circle amplification, loop-mediated isothermal amplification, ligation mediated rolling circle amplification and the like. In some cases SPIA may be used to generate an amplified DNA product with a 3'-end comprising a defined sequence (A') and a 5'-end comprising sequence (B).

[00135] **Figure 12** depicts a method as in the previous figure for producing amplified DNA using a ligand attached all DNA first primer. In this method a chimeric oligonucleotide is used to enable the generation of amplified DNA using the SPIA method comprising a defined 3'-end sequence (A') and a defined 5'-end sequence (B).

Amplification of a target nucleic acid using an non-tailed all DNA first primer

[00136] In some embodiments of the present invention, a method for amplification of a target nucleic acid (DNA or RNA) may comprise using an all DNA first primer that comprises an annealing sequence (P) as described previously but the all DNA first primer may not comprise a tail sequence. The all DNA first primer may be annealed to a specific, random or multiple regions of the target nucleic acid and extended with a polymerase as described previously to generate a first primer extension product. In some cases, a set of first primers may be used for amplification of a set of target sequences within the target nucleic acid in a multiplex fashion. A DNA-RNA chimeric second primer comprising a tail sequence (A) or (B) and an annealing sequence (P) as described previously may then be annealed to the first primer extension product and or the target nucleic acid and extended with a polymerase. The first and second primer extension products, primers, and target nucleic acid may then be treated with a single stranded DNA specific 3'-exonuclease such as exonuclease 1 and or a single and or double stranded RNA specific RNase such as RNase 1 to degrade unincorporated primers, unhybridized target nucleic acid, unhybridized 3'-ends of the first primer extension product, and unincorporated primers. The remaining target nucleic acid, and first and second primer extension products may then be extended with a polymerase exhibiting substantial reverse transcriptase activity such as for example Bst polymerase to generate a double stranded nucleic acid with a DNA-RNA heteroduplex. In some cases, the double stranded nucleic acid may be amplified using a number of methods known to the art such as but not limited to PCR, linear PCR, SPIA, rolling circle amplification, loop-mediated isothermal amplification, ligation mediated rolling circle amplification and the like. In some cases SPIA may be used to generate an amplified DNA product with a 3'-end comprising a defined sequence (A') and in some cases a 5'-end comprising sequence (B).

[00137] **Figure 13** depicts a method of producing amplified DNA using the SPIA method with an all DNA first primer where the first primer has no tail sequence, but comprises a template annealing sequence (P). The template annealing sequence may be target specific or randomized. The method further comprises a DNA-RNA second primer of tail sequence (B), an annealing sequence for annealing to the first primer extension product, and a DNA target nucleic acid. The use of randomized annealing sequences may require the use of a non-strand displacing polymerase in order to avoid whole genome amplification. In some cases, a non-strand-displacing polymerase may enable sequence specific amplification when using a sequence-specific annealing sequence with one primer (first or second) and a randomized annealing sequence with another

IDT Exhibit 1020.3441

primer. In some cases, a non-strand-displacing polymerase may enable sequence specific amplification when using a sequence-specific annealing sequence with an all DNA first primer and a random annealing sequence with an RNA-DNA chimeric second primer. Alternatively, whole genome amplification may be desired in which case it may be advantageous to employ a strand displacing polymerase to generate the first or second primer extension products or both. In some cases, the use of a strand displacing polymerase is advantageous to generate longer primer extension products or to enable the generation of second primer extension products.

[00138] **Figure 14** depicts a method of producing amplified DNA using the SPIA method with a sequence specific non-tailed all DNA first primer comprising an annealing sequence (P). In step I, the first primer is hybridized to a DNA target nucleic acid, and extended using a DNA polymerase. The method further comprises a DNA-RNA chimeric second primer comprising tail sequence (B) and an annealing sequence for annealing to the first primer extension product. The annealing sequence of the chimeric second primer may be a randomized sequence. In step II, the double stranded complex of the DNA target nucleic acid and the first primer extension product is denatured, and the second primer is annealed to the first primer extension product. Figure 14 also shows that some of the second primer may anneal to the other DNA strand, not containing the target sequence, or other portions of the DNA strand comprising the target sequence, but that most of the products from these reactions may result in SPIA amplified products. The second primer is extended in step III with a DNA polymerase. The DNA polymerase may lack strand displacement activity. In step IV a 3'-exonuclease is used to remove the non-hybridized portion of the first primer extension product. In step V, a DNA polymerase is used to extend the first primer extension product. The product of step V can then be used as a substrate for SPIA amplification using a chimeric amplification primer as shown in step VI. and a DNA target nucleic acid.

Additional amplification strategies using all DNA first primers

[00139] In **Figure 15**, a ligand attached all DNA first primer comprising a 5' tail sequence (A), a 3'-annealing sequence (P), and a DNA template nucleic acid. This figure depicts a method for amplifying two stands of the DNA template nucleic acid simultaneously. The ligand attached primer also allows functionalization and or purification of one of the strands to enable amplification of either strand selectively. In the first step, the ligand attached all DNA first primer is annealed to a target DNA and extended with a DNA dependent DNA polymerase to create a first primer extension product. The method further comprises denaturing the hybridization complex between the target nucleic acid and the first primer extension product, and annealing a second primer to the first primer extension product as well as the target DNA. The second primer comprises a 5'-segment a portion of which comprises RNA comprising a tail sequence (B) and a 3'-segment a portion of which comprises DNA comprising an annealing sequence (P) as described previously. The second primer may then be extended by a DNA dependent DNA polymerase to create a second primer extension product. In some cases, the first and second primer extension product form a double stranded nucleic acid product. In other cases, the second primer extension product and the target DNA form a double stranded nucleic acid product. The method further comprises treating the double stranded nucleic acid products with a single stranded DNA specific 3'-exonuclease and a DNA polymerase to create two double stranded nucleic acid products. In one case the double stranded nucleic acid product comprises a ligand at one end and a DNA-RNA heteroduplex comprising sequence (B) and (B') at the other end. In another case, the double stranded nucleic acid product does not comprise a ligand at one end. These double

IDT Exhibit 1020.3442

stranded nucleic acid products may act as substrates for SPIA amplification by providing a reaction mixture comprising RNase H, a chimeric amplification primer comprising a 5'-RNA segment and a 3'-DNA segment as described previously, and a DNA polymerase with substantial strand displacing activity.

[00140] **Figure 16** depicts a method of producing amplified DNA using the SPIA method without the use of a DNA-RNA chimeric second primer. In this method, an all DNA first primer of 5'-tail sequence (A) and a random or target specific template nucleic acid annealing sequence (P) is used to form the first primer extension product hybridized to a DNA or RNA template nucleic acid as previously described. In step II, the template nucleic acid is separated from the first primer extension product and an all DNA second primer comprising an annealing sequence that is annealed to the first primer extension product. The resulting double stranded first and second primer extension products are optionally treated with a 3'-exonuclease as described, the exonuclease is removed or inactivated and the first and second primer extension products are denatured. A chimeric oligonucleotide comprising a 5'-RNA sequence (C) and a 3'-DNA sequence (A) is then annealed to the second primer extension product, generating a substrate suitable for subsequent SPIA amplification. The SPIA amplifications products comprise a defined 5'-sequence (A).

[00141] **Figure 17** depicts an alternative method for producing amplified DNA using the SPIA method without the use of a DNA-RNA chimeric second primer. This method allows for the production of amplified product with defined sequences at or near both ends of the amplified product. The method comprises an all DNA second primer comprising a 5'-tail sequence (B) and a 3'-template annealing sequence (P). The method generates a substrate suitable for subsequent SPIA amplification. The SPIA amplification products comprise a defined 3'-sequence (B') and a defined 5'-end comprising sequence (A). In step I, an all DNA first primer comprising a 5'-segment comprising tail sequence (A) is annealed to a target nucleic acid and extended with a DNA polymerase to create a first primer extension product. In step II, the first primer extension product is denatured from the target nucleic acid, and an all DNA second primer as described hereinabove is annealed to the first primer extension product. In some cases, the target nucleic acid comprises RNA which is degraded in this step by heat, enzymes, chemical degradation etc. In step III, the second primer is extended with a DNA polymerase to create a double stranded nucleic acid product. Step IV depicts an optional step in which the double stranded nucleic acid product is treated with a single stranded DNA specific 3'-exonuclease. In step V, the double stranded nucleic acid product of step III or IV is denatured by heat or chemical means. In step VI a DNA-RNA chimeric olignucleotide (or "oligo") as described previously is annealed to the second primer extension product. In step VII, the chimeric oligonucleotide is extended to yield a third primer extension product comprising a 3'- portion comprising sequence (B') and a 5'-portion comprising RNA comprising sequence (C). Step VII further comprises extending the second primer extension product such that the 3'- segment comprises sequence (C'). Step VII yields a double stranded nucleic acid comprising sequence (B) and (B') at one end and a DNA-RNA heteroduplex comprising sequence (C) and (C') at the other end.

[00142] **Figure 18** depicts a method of amplifying a target sequence from the double stranded nucleic acid product of **Figure 17** by employing the double stranded nucleic acid product as a substrate for SPIA amplification. In the first step of **Figure 18**, RNase H cleaves the RNA portion of the (C')-(C) DNA-RNA heteroduplex of the double stranded nucleic acid product, and a chimeric amplification primer as described previously is annealed to the DNA sequence (C'). In the next step, SPIA amplification products comprising a 3' end comprising sequence (B') and a 5'-end comprising sequence (A) are generated by providing RNase H, a chimeric amplification primer, and a strand displacing DNA polymerase.

IDT Exhibit 1020.3443

[00143] .While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

General Techniques

[00144] The practice of the invention may employ, unless otherwise indicated, conventional techniques of molecular biology (including recombinant techniques), microbiology, cell biology, biochemistry, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature, such as, "Molecular Cloning: A Laboratory Manual", second edition (Sambrook et al., 1989); "Oligonucleotide Synthesis" (M. J. Gait, ed., 1984); "Animal Cell Culture" (R. I. Freshney, ed., 1987); "Methods in Enzymology" (Academic Press, Inc.); "Current Protocols in Molecular Biology" (F. M. Ausubel et al., eds., 1987, and periodic updates); "PCR: The Polymerase Chain Reaction", (Mullis et al., eds., 1994).

[00145] The terms "Polynucleotide," or "nucleic acid," as used interchangeably herein, refer to polymers of nucleotides of any length, and include DNA and RNA. The nucleotides can be deoxyribonucleotides, ribonucleotides, modified nucleotides or bases, and/or their analogs, or any substrate that can be incorporated into a polymer by DNA or RNA polymerase. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and their analogs. If present, modification to the nucleotide structure may be imparted before or after assembly of the polymer. The sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component. Other types of modifications include, for example, "caps", substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as, for example, those with uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoamidates, cabamates, etc.) and with charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), those containing pendant moieties, such as, for example, proteins (e.g., nucleases, toxins, antibodies, signal peptides, ply-L-lysine, etc.), those with intercalators (e.g., acridine, psoralen, etc.), those containing chelators (e.g., metals, radioactive metals, boron, oxidative metals, etc.), those containing alkylators, those with modified linkages (e.g., alpha anomeric nucleic acids, etc.), as well as unmodified forms of the polynucleotide(s). Further, any of the hydroxyl groups ordinarily present in the sugars may be replaced, for example, by phosphonate groups, phosphate groups, protected by standard protecting groups, or activated to prepare additional linkages to additional nucleotides, or may be conjugated to solid supports. The 5' and 3'-terminal OH can be phosphorylated or substituted with amines or organic capping groups moieties of from 1 to 20 carbon atoms. Other hydroxyls may also be derivatized to standard protecting groups. Polynucleotides can also contain analogous forms of ribose or deoxyribose sugars that are generally known in the art, including, for example, 2'-O-methyl-, 2'-O-allyl, 2'-fluoro- or 2'-azido-ribose, carbocyclic sugar analogs, .alpha.-anomeric sugars, epimeric sugars such as arabinose, xyloses or lyxoses, pyranose sugars, furanose sugars, sedoheptuloses, acyclic analogs and abasic nucleoside analogs such as methyl riboside. One or more phosphodiester linkages may be replaced by alternative linking groups. These alternative linking groups include, but are not limited to, embodiments wherein

IDT Exhibit 1020.3444

phosphate is replaced by P(O)S("thioate"), P(S)S ("dithioate"), "(O)NR_2 ("amidate"), P(O)R, P(O)OR', CO or CH_2 ("formacetal"), in which each R or R' is independently H or substituted or unsubstituted alkyl (1-20 C) optionally containing an ether (--O--) linkage, aryl, alkenyl, cycloalkyl, cycloalkenyl or araldyl. Not all linkages in a polynucleotide need be identical. The preceding description applies to all polynucleotides referred to herein, including RNA and DNA.

[00146] "Oligonucleotide," as used herein, generally refers to short, generally single stranded, generally synthetic polynucleotides that are generally, but not necessarily, less than about 200 nucleotides in length. The oligonucleotide(s) are generally comprised of a sequence of at least 5 nucleotides, generally from about 10 to about 100 nucleotides, about 20 to about 50 nucleotides, and often about 10 to about 30 nucleotides in length. The oligonucleotides of the invention can be DNA, RNA, DNA-RNA, or other polynucleotide. The terms oligonucleotide or sequence may be used interchangeably herein.

[00147] Various techniques can be employed for preparing an oligonucleotide utilized in the present invention. Such oligonucleotide can be obtained by biological synthesis or by chemical synthesis. For short sequences (up to about 100 nucleotides) chemical synthesis may frequently be more economical as compared to the biological synthesis. In addition to economy, chemical synthesis provides a convenient way of incorporating low molecular weight compounds and/or modified bases during the synthesis step. Furthermore, chemical synthesis is very flexible in the choice of length and region of the target polynucleotide binding sequence. The oligonucleotide can be synthesized by standard methods such as those used in commercial automated nucleic acid synthesizers. Chemical synthesis of DNA on a suitably modified glass or resin can result in DNA covalently attached to the surface. This may offer advantages in washing and sample handling. For longer sequences standard replication methods employed in molecular biology can be used such as the use of M13 for single stranded DNA as described by J. Messing (1983) Methods Enzymol, 101, 20-78.

[00148] In the present invention, nucleoside triphosphates are incorporated by a polymerase enzyme in the extension of the primer to produce an extension product. Nucleoside triphosphates are generally nucleosides having a 5'-triphosphate substituent. The nucleosides are pentose sugar derivatives of nitrogenous bases of either purine or pyrimidine derivation, covalently bonded to the 1'-carbon of the pentose sugar, which is usually a deoxyribose or a ribose. The purine bases include adenine (A), guanine (G), inosine (I), and derivatives and analogs thereof. The pyrimidine bases include cytosine (C), thymine (T), uracil (U), and derivatives and analogs thereof. Nucleoside triphosphates include deoxyribonucleoside triphosphates such as the four common triphosphates dATP, dCTP, dGTP and dTTP and ribonucleoside triphosphates such as the four common triphosphates rATP, rCTP, rGTP and rUTP. The term "nucleoside triphosphates" also includes derivatives and analogs thereof, which are exemplified by those derivatives that are recognized in a similar manner to the underivatized nucleoside triphosphates. Examples of such derivatives or analogs, by way of illustration and not limitation, are those which are biotinylated, amine modified, alkylated, and the like and also include phosphorothioate, phosphite, ring atom modified derivatives, and the like.

[00149] As used herein, the term “nucleotide” generally refers to a base-sugar-phosphate combination that is the monomeric unit of nucleic acid polymers, i.e., DNA and RNA. In some aspects of the invention modified nucleotides are used, for example, where a nucleotide is connected to a ligand. A modified nucleotide is generally the unit in a nucleic acid polymer that results from the incorporation of a modified nucleoside triphosphate during an amplification reaction and therefore becomes part of the nucleic acid polymer.

IDT Exhibit 1020.3445

[00150] As used herein, a "nucleoside" is generally a base-sugar combination or a nucleotide lacking a phosphate moiety.

[00151] The term "double stranded product" is used herein to refer to products that are produced by the extension of a primer. It is understood that the products are at least partially double stranded, for example, in the region comprising the primer extension product and its complement. The double stranded product need not be completely double-stranded, and may have single stranded regions. It is also understood that the double stranded product can have heteroduplex regions in which one strand comprises RNA and the complementary strand comprised DNA in that region.

Reaction Conditions and Detection

[00152] Appropriate reaction media and conditions for carrying out the methods of the invention include those that permit nucleic acid extension, copying, and amplification according to the methods of the invention. Such media and conditions are known to persons of skill in the art, and are described in various publications, such as U.S. Pat. Nos. 5,554,516; 5,716,785; 5,130,238; 5,194,370; 6,090,591; 5,409,818; 5,554,517; 5,169,766; 5,480,784; 5,399,491; 5,679,512; and PCT Pub. No. WO99/42618. For example, a buffer may be Tris buffer, although other buffers can also be used as long as the buffer components are non-inhibitory to enzyme components of the methods of the invention. The pH is preferably from about 5 to about 11, more preferably from about 6 to about 10, even more preferably from about 7 to about 9, and most preferably from about 7.5 to about 8.5. The reaction medium can also include bivalent metal ions such as Mg.sup.2+, or Mn.sup.2+, at a final concentration of free ions that is within the range of from about 0.01 to about 15 mM, and most preferably from about 1 to 10 mM. The reaction medium can also include other salts, such as KCl or NaCl, that contribute to the total ionic strength of the medium. For example, the range of a salt such as KCl is preferably from about 0 to about 125 mM, more preferably from about 0 to about 100 mM, and most preferably from about 0 to about 75 mM. The reaction medium can further include additives that could affect performance of the amplification reactions, but that are not integral to the activity of the enzyme components of the methods. Such additives include proteins such as BSA, single stranded binding protein (for example, T4 gene 32 protein), and non-ionic detergents such as NP40 or Triton. Reagents, such as DTT, that are capable of maintaining enzyme activities can also be included. Such reagents are known in the art. Where appropriate, an RNase inhibitor (such as Rnasin) that does not inhibit the activity of the RNase employed in the method can also be included. Any aspect of the methods of the invention can occur at the same or varying temperatures. In some embodiments, the amplification reactions (particularly, primer extension and transcription; and generally not the step of denaturing) are performed isothermally, which substantially avoids the thermocycling process. The isothermal amplification reaction is carried out at a temperature that permits hybridization of the oligonucleotides (primer) of the invention to the template polynucleotide and that does not substantially inhibit the activity of the enzymes employed. The temperature can be in the range of preferably about 25 °C. to about 85 °C., more preferably about 30 °C. to about 80 °C., and most preferably about 37 °C. to about 75 °C. The temperature for the transcription steps can be lower than the temperature(s) for the preceding steps. The temperature of the transcription steps can be in the range of preferably about 25 °C. to about 85 °C., more preferably about 30 °C. to about 75 °C., and most preferably about 37 °C. to about 70 °C.

IDT Exhibit 1020.3446

Extension of primers

[00153] The methods of the present invention involve the extension of primers. In general, primers hybridize to, and are extended along (chain extended), a sequence within the target polynucleotide and, thus, the target polynucleotide acts as a template. The extended primers are chain "extension products." The sequence over which the primer is extended may lie between two defined sequences but need not be. In general, the primers hybridize with a sequence within the target polynucleotide. The target sequence usually contains from about 10 to 30 to 50 to 100 or up to 5,000 or more nucleotides, often 10 to 1,000 nucleotides, or 10-10,000 nucleotides. The target polynucleotide may be a fraction of a larger molecule or it may be substantially the entire molecule (polynucleotide analyte).

Single Primer Isothermal Amplification Using a Complex Comprising an RNA/DNA Partial Heteroduplex as a Template

[00154] In some aspects of the invention, the amplification method that is used is a single primer isothermal amplification using a complex comprising an RNA/DNA partial heteroduplex as a template. In this method, termed single primer isothermal amplification, a complex comprising an RNA/DNA partial heteroduplex is a substrate for further amplification as follows: an enzyme which cleaves RNA sequence from an RNA/DNA heteroduplex (such as RNase H) cleaves RNA from the partial heteroduplex, leaving a partially double stranded polynucleotide complex comprising a 3' single stranded DNA sequence. The 3' single stranded sequence (formed by cleavage of RNA in the complex comprising an RNA/DNA partial heteroduplex) is generally the complement of the amplification composite primer, and thus forms a specific binding site for a composite primer. Extension of a bound composite primer by a DNA-dependent DNA polymerase with strand displacement activity produces a primer extension product, which displaces the previously bound cleaved primer extension product, whereby polynucleotide (generally, DNA) product accumulates. See, for example, U.S. Pat. Nos. 6,251,639 and 6,692,918.

[00155] Amplification using a complex comprising an RNA/DNA partial heteroduplex as a template for further amplification (also termed single primer isothermal amplification) generally occurs under conditions permitting composite primer hybridization, primer extension by a DNA polymerase with strand displacement activity, cleavage of RNA from an RNA/DNA hybrid and strand displacement. In so far as the composite primer hybridizes to the 3'-single stranded portion (of the partially double stranded polynucleotide which is formed by cleaving RNA in the complex comprising an RNA/DNA partial heteroduplex) comprising, generally, the complement of at least a portion of the composite primer sequence, composite primer hybridization may be under conditions permitting specific hybridization.

[00156] In some embodiments, the methods of the invention result in amplification of a multiplicity, a large multiplicity, or a very large multiplicity of template polynucleotide sequences. In some embodiments, essentially all of the template polynucleotide present in the initial sample (e.g., all of the mRNA or all of the genomic DNA) is amplified. In other embodiments, at least 1, at least 2, at least 3, at least 4, at least 5, at least 10, at least 20. at least 50, at least 100, at least 200, at least 300, at least 500, at least 1,000, at least 5,000, at least 10,000 or more than 10,0000 distinct sequences (such as a gene or other sub-segment of a polynucleotide, transcripts of a nucleic acid target, a marker (such as a SNP or other polymorphism) are amplified, as assessed, e.g., by analysis of marker sequences known to be present in the template sample under analysis, using methods known in the art. Template polynucleotide sequences that are amplified may be present on the same polynucleotide (e.g., a chromosome or portion of a chromosome for genomic DNA

IDT Exhibit 1020.3447

template or on the same RNA for RNA template) or on different template polynucleotides (e.g., different chromosome or portions of chromosomes for DNA template, or different RNAs for RNA template). In some case, amplification of genomic DNA is exemplified herein, it will be understood by those of skill in the art, however, that the global amplification methods of the invention are suitable for amplification of any pool or subset of polynucleotides.

[00157] In some embodiments, the methods of the invention are used to globally amplify double stranded DNA target. It is understood that in these cases, the amplified product generally is a mixture of sense and antisense copies of the template DNA. In some embodiments, the methods of the invention are used to globally amplify a single stranded DNA or RNA target. In these cases, the amplification product may generally be a copy of either the target polynucleotide (sense copy) or of the complement to the target nucleotide (antisense copy). Whether the sense or antisense copy is produced will depend on the method, as will be understood by one of ordinary skill in the art. In some embodiments, the amplification product of different senses can be annealed to form a double stranded (or partially double stranded) complex. In other embodiments, they can be prevented from annealing (or subsequently denatured) to produce a mixture of single stranded amplification products. The amplified products may be of differing lengths.

[00158] As illustrated in these embodiments, all steps are isothermal (in the sense that thermal cycling is not required), although the temperatures for each of the steps may or may not be the same. It is understood that various other embodiments may be practiced, given the general description provided above. For example, as described and exemplified herein, certain steps may be performed as temperature is changed (e.g., raised, or lowered).

[00159] For simplicity, the isothermal amplification methods of the invention are described as two distinct steps or phases, above. It is understood that the two phases may occur simultaneously in some embodiments (for example, if the enzyme that cleaves RNA from RNA/DNA hybrid is included in the first reaction mixture).

[00160] Although generally only one composite primer is described above, it is further understood that the amplification methods may be performed in the presence of two or more different composite primers that randomly prime template polynucleotide. In addition, the amplification polynucleotide products of two or more separate amplification reactions conducted using two or more different composite primers that randomly prime template polynucleotide can be combined.

Other amplification methods

[00161] Some aspects of the invention comprise the amplification of polynucleotide molecules or sequences within the polynucleotide molecules. Amplification generally refers to a method that results in the formation of one or more copies of a nucleic acid or polynucleotide molecule or in the formation of one or more copies of the complement of a nucleic acid or polynucleotide molecule. Amplifications can be used in the invention, for example, to amplify or analyze a polynucleotide bound to a solid surface. The amplifications can be performed, for example, after archiving the samples in order to analyze the archived polynucleotide.

[00162] In some aspects of the invention, exponential amplification of nucleic acids or polynucleotides is used. These methods often depend on the product catalyzed formation of multiple copies of a nucleic acid or polynucleotide molecule or its complement. The amplification products are sometimes referred to as "amplicons." One such method for the enzymatic amplification of specific double stranded sequences of DNA is known as the polymerase chain reaction (PCR). This in vitro amplification procedure is based on repeated cycles of denaturation, oligonucleotide primer annealing, and primer extension by thermophilic

IDT Exhibit 1020.3448

template dependent polynucleotide polymerase, resulting in the exponential increase in copies of the desired sequence of the polynucleotide analyte flanked by the primers. The two different PCR primers, which anneal to opposite strands of the DNA, are positioned so that the polymerase catalyzed extension product of one primer can serve as a template strand for the other, leading to the accumulation of a discrete double stranded fragment whose length is defined by the distance between the 5' ends of the oligonucleotide primers.

[00163] Another method for amplification involves amplification of a single stranded polynucleotide using a single oligonucleotide primer. The single stranded polynucleotide that is to be amplified contains two non-contiguous sequences that are complementary to one another and, thus, are capable of hybridizing together to form a stem-loop structure. This single stranded polynucleotide already may be part of a polynucleotide analyte or may be created as the result of the presence of a polynucleotide analyte.

[00164] Another method for achieving the result of an amplification of nucleic acids is known as the ligase chain reaction (LCR). This method uses a ligase enzyme to join pairs of preformed nucleic acid probes. The probes hybridize with each complementary strand of the nucleic acid analyte, if present, and ligase is employed to bind each pair of probes together resulting in two templates that can serve in the next cycle to reiterate the particular nucleic acid sequence.

[00165] Another method for achieving a nucleic acid amplification is the nucleic acid sequence based amplification (NASBA). This method is a promoter-directed, enzymatic process that induces in vitro continuous, homogeneous and isothermal amplification of a specific nucleic acid to provide RNA copies of the nucleic acid. The reagents for conducting NASBA include a first DNA primer with a 5'-tail comprising a promoter, a second DNA primer, reverse transcriptase, RNAse-H, T7 RNA polymerase, NTP's and dNTP's.

[00166] Another method for amplifying a specific group of nucleic acids is the Q-beta-replicase method, which relies on the ability of Q-beta-replicase to amplify its RNA substrate exponentially. The reagents for conducting such an amplification include "midi-variant RNA" (amplifiable hybridization probe), NTP's, and Q-beta-replicase.

[00167] Another method for amplifying nucleic acids is known as 3SR and is similar to NASBA except that the RNAse-H activity is present in the reverse transcriptase. Amplification by 3SR is an RNA specific target method whereby RNA is amplified in an isothermal process combining promoter directed RNA polymerase, reverse transcriptase and RNase H with target RNA. See for example Fahy et al. PCR Methods Appl. 1:25-33 (1991).

[00168] Another method for amplifying nucleic acids is the Transcription Mediated Amplification (TMA) used by Gen-Probe. The method is similar to NASBA in utilizing two enzymes in a self-sustained sequence replication. See U.S. Patent No. 5,299,491 herein incorporated by reference.

[00169] Another method for amplification of nucleic acids is Strand Displacement Amplification (SDA) (Westin et al 2000, Nature Biotechnology, 18, 199-202; Walker et al 1992, Nucleic Acids Research, 20, 7, 1691-1696), which is an isothermal amplification technique based upon the ability of a restriction endonuclease such as HincII or BsoBI to nick the unmodified strand of a hemiphosphorothioate form of its recognition site, and the ability of an exonuclease deficient DNA polymerase such as Klenow exo minus polymerase, or Bst polymerase, to extend the 3'-end at the nick and displace the downstream DNA strand. Exponential amplification results from coupling sense and antisense reactions in which strands displaced from a sense reaction serve as targets for an antisense reaction and vice versa.

IDT Exhibit 1020.3449

[00170] Another method for amplification of nucleic acids is Rolling Circle Amplification (RCA) (Lizardi et al. 1998, Nature Genetics, 19:225-232). RCA can be used to amplify single stranded molecules in the form of circles of nucleic acids. In its simplest form, RCA involves the hybridization of a single primer to a circular nucleic acid. Extension of the primer by a DNA polymerase with strand displacement activity results in the production of multiple copies of the circular nucleic acid concatenated into a single DNA strand.

[00171] In some embodiments of the invention, RCA is coupled with ligation. For example, a single oligonucleotide can be used both for ligation and as the circular template for RCA. This type of polynucleotide can be referred to as a "padlock probe" or a "RCA probe". For a padlock probe, both termini of the oligonucleotide contain sequences complementary to a domain within a nucleic acid sequence of interest. The first end of the padlock probe is substantially complementary to a first domain on the nucleic acid sequence of interest, and the second end of the padlock probe is substantially complementary to a second domain, adjacent to the first domain near the first domain. Hybridization of the oligonucleotide to the target nucleic acid results in the formation of a hybridization complex. Ligation of the ends of the padlock probe results in the formation of a modified hybridization complex containing a circular polynucleotide. In some cases, prior to ligation, a polymerase can fill in the gap by extending one end of the padlock probe. The circular polynucleotide thus formed can serve as a template for RCA that with the addition of a polymerase results in the formation of an amplified product nucleic acid. The methods of the invention described herein, can produce amplified products with defined sequences on both the 5'-and 3'-ends. Such amplified products can be used as padlock probes.

[00172] Some aspects of the invention utilize the linear amplification of nucleic acids or polynucleotides. Linear amplification generally refers to a method that involve the formation of one or more copies of the complement of only one strand of a nucleic acid or polynucleotide molecule, usually a nucleic acid or polynucleotide analyte. Thus, the primary difference between linear amplification and exponential amplification is that in the latter process, the product serves as substrate for the formation of more product, whereas in the former process the starting sequence is the substrate for the formation of product but the product of the reaction, i.e. the replication of the starting template, is not a substrate for generation of products.. In linear amplification the amount of product formed increases as a linear function of time as opposed to exponential amplification where the amount of product formed is an exponential function of time.

[00173] In some embodiments, amplification methods can be solid-phase amplification, polony amplification, colony amplification, emulsion PCR, bead RCA, surface RCA, surface SDA, etc., as will be recognized by one of skill in the art. In some embodiments, amplification methods that results in amplification of free DNA molecules in solution or tethered to a suitable matrix by only one end of the DNA molecule can be used. Methods that rely on bridge PCR, where both PCR primers are attached to a surface (see, e.g., WO 2000/018957 and Adessi et al., Nucleic Acids Research (2000): 28(20): E87) can be used. In some cases the methods of the invention can create a "polymerase colony technology", or "polony", referring to a multiplex amplification that maintains spatial clustering of identical amplicons (see Harvard Molecular Technology Group and Lipper Center for Computational Genetics website). These include, for example, in situ polonies (Mitra and Church, Nucleic Acid Research 27, e34, Dec. 15, 1999), in situ rolling circle amplification (RCA) (Lizardi et al., Nature Genetics 19, 225, July 1998), bridge PCR (U.S. Pat. No. 5,641,658), picotiter PCR (Leamon et al., Electrophoresis 24, 3769, November 2003), and emulsion PCR

IDT Exhibit 1020.3450

(Dressman et al., PNAS 100, 8817, Jul. 22, 2003). The methods of the invention provide new methods for generating and using polonies.

Compositions

Primers and Oligonucleotides

[00174] The invention provides compositions, kits, complexes, reaction mixtures and systems comprising various components and their combinations used in the amplification methods described herein. In one aspect, for example, the invention provides compositions comprising an all DNA primer. All DNA or RNA-DNA chimeric primers may comprise a nucleotide sequence (i.e. a polynucleotide), generally with a free 3'-OH group, that hybridizes with a template sequence (such as a target RNA or DNA, or a primer extension product) and is capable of promoting polymerization of a polynucleotide complementary to the template. An all DNA primer or an RNA-DNA chimeric primer maybe, for example, an oligonucleotide. It can also be, for example, a sequence of the template (such as a primer extension product or a fragment of the template created following RNase cleavage of a template-DNA complex) that is hybridized to a sequence in the template itself (for example, as a hairpin loop), and that is capable of promoting nucleotide polymerization. Thus, a primer can be an exogenous (e.g., added) primer or an endogenous (e.g., template fragment) primer.

[00175] The primers of the invention are usually oligonucleotide primers. A primer is generally an oligonucleotide that is employed in an extension on a polynucleotide template. The oligonucleotide primer is usually a synthetic nucleotide that is single stranded, containing a sequence at its 3'-end that is capable of hybridizing with a sequence of the target polynucleotide. Normally, the 3'-region of the primer that hybridizes with the target nucleic acid has at least 80%, preferably 90%, more preferably 95%, most preferably 100%, complementarity to a sequence or primer binding site. The number of nucleotides in the hybridizable sequence of a specific oligonucleotide primer should be such that stringency conditions used to hybridize the oligonucleotide primer may prevent excessive random non-specific hybridization. Usually, the number of nucleotides in the hybridizing portion of the oligonucleotide primer may be at least as great as the defined sequence of the target polynucleotide, namely, at least five nucleotides, at least about ten nucleotides, at least about 15 nucleotides, generally from about 5 to about 200, and usually about 10 to about 50 nucleotides.

[00176] A "random primer," or "random annealing sequence" as used herein, is a primer that comprises a sequence that is designed not necessarily based on a particular or specific sequence in a sample, but rather is based on a statistical expectation (or an empirical observation) that the sequence of the random primer is hybridizable (under a given set of conditions) to one or more sequences in the sample. The random primers used herein are generally tailed random primers comprising a 3'-segment that acts as a random primer to the target polynucleotide, and a 5'-sequence that generally does not hybridize to the target polynucleotide. The sequence of any one primer molecule from a sample of random primers (or their complement) may or may not be naturally-occurring, and may or may not be present in the sequences in a sample of interest. The amplification of a plurality of polynucleotides, e.g. DNA or RNA species in a single reaction mixture would generally, but not necessarily, employ a multiplicity, preferably a large multiplicity, of random primers. As is well understood in the art, a "random primer" can also refer to a primer that is a member of a population of primers (a plurality of random primers) which collectively are designed to hybridize to a desired and/or a significant number of target sequences. A random primer may

IDT Exhibit 1020.3451

hybridize at a plurality of sites on a nucleic acid sequence. The use of random primers provides a method for generating primer extension products complementary to a target polynucleotide which does not require prior knowledge of the exact sequence of the target. In some embodiments one portion of a primer is random, and another portion of the primer comprises a defined sequence. For example, in some embodiments, a 3'-portion of the primer may comprise a random sequence, while the 5'-portion of the primer comprises a defined sequence. In other embodiments, for example, a 3'-portion of the primer may comprise a defined sequence, while the 5' portion of the primer comprises a random sequence. In some embodiments a 3'-random portion of the primer may comprise DNA, and a 5'-portion defined portion of the primer may comprise RNA, in other embodiments, both the 3'-and 5'-portions may comprise DNA.

[00177] All DNA primers generally comprise polynucleotide annealing sequences. The length of the annealing sequence can be for example from about 2 to about 100 nucleotides. In some cases, the all DNA primers comprise polynucleotide annealing sequences of length about 3 nucleotides to about 50 nucleotides. In other cases, the all DNA primers comprise short polynucleotide annealing sequences of about 4 nucleotides to about 20 nucleotides in length. In still other embodiments, the all DNA primers comprise polynucleotide annealing sequences of about 4 to about 10 or 15 nucleotides in length. In some embodiments, the all DNA primer annealing sequence is about 8 nucleotides. In other embodiments, the all DNA primer annealing sequence is about 6 nucleotides. In some cases the primer may consist of only the annealing sequence, whereas in other cases the primer may also comprise other portions which are not annealing sequences such as a tail sequence or an element other than the annealing sequence and the tail sequence such as a linker for example. In other cases, the all DNA primers also comprise a ligand. The ligand may or may not be comprised of nucleic acid. In some cases, the ligand is a small organic molecule such as for example biotin or a fluorophore. The primers may be synthesized by a number of common methods known in the art. In some cases, the methods of synthesizing primers of the present invention include solid phase synthesis methods such as provided in US Patent 5,623,068.

[00178] Precise nucleotide lengths and compositions can be determined by one with ordinary skill in the art for designing DNA primers for hybridization to an RNA or DNA template nucleic acid and extension by a polymerase. The nucleotide length and sequence composition may take into account the temperature of the hybridization and polymerization reactions to be performed, the propensity for secondary structure inherent in the all DNA primer or the nucleic acid template, the likelihood of a given all DNA primer to hybridize to non-specific or closely related sequences, and the sequence of the nucleic acid template. Suitable primer lengths include from about 5 to about 200 or more nucleotides including about 5, 6, 7, 8, 9, 10, 12, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 70 , 80, 90, or about 100 or more nucleotides in length.

[00179] The nucleotide length and sequence composition may be designed to provide specific hybridization to a desired sequence, specific hybridization to a family of related sequences, or hybridization to any sequence or substantially any sequence of a genome or transcriptome. In the case of designing a primer that anneals to any sequence, randomized sequences may be used such as for example random heptamers, hexamers, septamers, octamers, nonamers, or random decamers. In some cases, non-standard bases which exhibit base pairing interactions with more than one partner such as inosine which can base pair with uracil, adenine, and cytosine may be used. The non-standard bases with greater propensity for multiple base pairing partners are referred to as wobble bases. In other cases, specific annealing sequences may be used in combination with such wobble bases. In some cases, all DNA primers comprise specific annealing sequences in which one or more positions within the annealing sequence are randomized. Such a

IDT Exhibit 1020.3452

combination of specific sequences with randomized positions may be used for example to amplify a set of related or homologous nucleic acid sequences. In still other cases, all DNA primers comprise specific annealing sequences, or a set of specific sequences. In an exemplary embodiment, the annealing sequence comprises a poly thymine sequence such that messenger RNA molecules are preferentially amplified.

[00180] In some cases, all DNA primers may comprise modification of bases or the introduction of nonstandard bases in addition to or other than the inosine base previously described. Non standard bases may be used to increase the thermal stability of the hybridized primer to its target nucleic acid, to increase specificity, or to increase stability (i.e. resistance to degradation). Non standard bases may include but are not limited to locked nucleic acids, peptide nucleic acids, pyranosyl nucleic acids, methylated nucleic acids, or any other modified or nonstandard nucleic acid known to the art. In some cases, the first primer, second primer, composite primer, or chimeric primer comprises non-standard bases which can be recognized by a polymerase such as for example a DNA-dependent DNA polymerase or an RNA-dependent DNA polymerase and copied in a consistent and efficient manner. In some cases, an amplification primer, composite amplification primer, or chimeric amplification primer may comprise any variety of nonstandard bases as described herein.

[00181] All DNA primers may also comprise a tail sequence. The tail sequence may provide a useful way for further manipulation of amplified products such as purification, functionalization, capture, immobilization or other downstream processes such as sequence determination or further amplification, the tag may also be integral to the amplification process itself such as but not limited to the process described in **Figures 3** and **8**. The tail sequence may comprise a sequence that does not hybridize or does not substantially hybridize to the target nucleic acid under the conditions of the method. In some cases, the tail sequence may be designed such that it does not hybridize or substantially hybridize to other primers, oligonucleotides, or tail sequences employed in the method such as sequence (B) of **Figure 4** or sequence (C) of **Figure 10**. . In other cases, the tail sequence may be designed such that it does hybridize to other primers, oligonucleotides, or tail sequences employed in the method. The tail sequence may comprise a sequence between about 1 or 2 nucleotides in length to about 200 nucleotides in length including about 3, 4, 5, 8, 10, 15, 20, 25, 30, 50, 75, 100, 150, or 200 nucleotides in length.

[00182] In some embodiments, all DNA primers comprising a 3'-template annealing sequence and a 5'-sequence tail (A) are used in a first primer extension reaction. The all DNA first primer may be annealed to the target nucleic acid subsequent to an elevated temperature denaturing step to remove secondary structure and or separate complementary strands. Annealing of the all DNA first primer may be followed by addition of polymerase and nucleotide triphosphates as well as buffers, cations, anions, and solute modifying reagents such as glycerol, polyethylene glycol, bovine serum albumin, dimethyl sulfoxide, or single-stranded binding proteins, and the like to initiate a first primer extension reaction. In some cases, a strand displacing polymerase may be used, while in other cases it may be advantageous to utilize a polymerase that does not exhibit substantial strand displacing activity. The first primer extension reaction may provide a template for further manipulation including but not limited to annealing of a chimeric DNA-RNA second primer.

[00183] In some aspects of the present invention, chimeric primers (also referred to as composite primers) are utilized. Chimeric primers can be for example a second primer in the methods of the invention, chimeric extension primers or chimeric amplification primers. Chimeric primers of the present invention include primers that comprise an RNA segment and a DNA segment. The RNA segment may be generally located at the 5' end of the chimeric primer and the DNA segment may generally be located at the 3' end of the

IDT Exhibit 1020.3453

chimeric primer. In some cases the RNA and DNA segments are adjacent. In other cases, there is an intervening element between the RNA and the DNA segment. The intervening element may comprise additional nucleic acid sequence. In one embodiment, a portion of the DNA segment comprises a template annealing sequence. The template annealing sequence as described for the all DNA primer may be target specific, a set of target specific sequences, or a random sequence.

[00184] In some embodiments of the present invention, the DNA-RNA chimeric primer annealing sequences may be between about 2 nucleotides in length to about 100 nucleotides in length. In some cases, annealing sequences may be between about 3 nucleotides in length to about 50 nucleotides in length. In still other cases, annealing sequences may be about 4 to about 30 nucleotides in length. For example, but without limitation, DNA-RNA chimeric primer annealing sequences may be about 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, or 25 nucleotides in length. In some cases, random annealing sequences comprise random pentamers, hexamers, heptamers, octomers, decamers, or larger.

[00185] In some cases, for example with the use of random template annealing sequences, it may be advantageous to extend the chimeric primer with a polymerase that does not exhibit substantial strand displacement activity. In other cases, it may be advantageous to extend the chimeric primer with a polymerase that does exhibit substantial strand displacement activity. In some cases, the use of a polymerase that exhibits substantial strand displacement activity may be useful for whole genome and or whole transcriptome amplification.

[00186] In some embodiments of the present invention, a portion of the RNA segment of the chimeric primer comprises a sequence tail. The sequence tail may in some cases comprise the same sequence (A) as the sequence tail used for an all DNA first primer extension reaction. In other cases, the sequence tail may comprise a different sequence (B) as compared to the sequence tail used for an all DNA first primer extension reaction.

[00187] In some embodiments of the present invention, a chimeric primer comprising a 5'-RNA segment comprising a sequence tail and a 3' DNA segment comprising a template annealing sequence is annealed to a first primer extension product, a polymerase and is added to the annealing reaction, and a second primer extension product is created. In some embodiments of the present invention, a 3'-exonuclease is added after the second primer extension reaction is completed. The 3'-exonuclease may remove the 3'-unhybridized portion of the first primer extension product. Removal of the exonuclease by heat, the addition of inhibitors, or nucleic acid purification methods may allow further extension of the first primer extension product. The further extension may provide sequence complementary to the sequence tail of the chimeric primer.

[00188] Amplification primers of the present invention include chimeric or composite amplification primers. Composite amplification primers are RNA/DNA composite primers that can be used to create multiple copies of (amplify) a polynucleotide sequence isothermally using RNA cleavage, and DNA polymerase activity with strand displacement. Amplification with such primers is described, for example in U. S. Patents 6,251,639, 6,692,918, and 6,946,251. The composite amplification primer comprises sequences capable of hybridizing to a portion of a DNA template, and most often comprises sequences hybridizable to a defined 3'-portion of the DNA.

[00189] A composite amplification primer comprises at least one RNA portion that is capable of (a) binding (hybridizing) to a sequence on a DNA template independent of hybridization of the DNA portion(s) to a sequence on the same extension product; and being cleaved with an RNase H when hybridized to the DNA

IDT Exhibit 1020.3454

template. The composite amplification primers bind to the DNA template to form a partial heteroduplex in which only the RNA portion of the primer is cleaved upon contact with a ribonuclease such as RNase H, while the DNA template remains intact, thus enabling annealing of another composite primer.

[00190] The composite amplification primers also comprise a 3'-DNA portion that is capable of hybridization to a sequence on the DNA template such that its hybridization to the DNA is favored over that of the nucleic acid strand that is displaced from the DNA template by the DNA polymerase. Such primers can be rationally designed based on well known factors that influence nucleic acid binding affinity, such as sequence length and/or identity, as well as hybridization conditions. In some aspects of the present invention, the 3' DNA portion of a composite amplification primer may be about 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, or 14 nucleotides in length.

[00191] In some embodiments of the present invention chimeric oligonucleotides are utilized. In some cases, the chimeric oligonucleotides may be used to provide a defined end sequence for the SPIA amplified DNA products. The chimeric oligonucleotides may comprise a 5' segment comprising RNA and a 3' segment comprising DNA. In some embodiments such as for example those depicted in **Figures 10 and 12**, a portion of the 3' segment of the chimeric oligonucleotide is substantially the same as the RNA sequence removed from the DNA-RNA heteroduplex of a SPIA substrate by RNase H. In other cases, such as for example that depicted in **Figure 16**, a portion of the 3' segment of the chimeric oligonucleotide is substantially the same as the complement of the tail sequence of the first primer.

Enzymes

[00192] Polymerases are used in the methods of the invention, for example to extend primers to produce primer extension products. A polymerase, or nucleotide polymerase, is a catalyst, usually an enzyme, for forming an extension of a polynucleotide along a DNA or RNA template where the extension is complementary thereto. The nucleotide polymerase is a template dependent polynucleotide polymerase and utilizes nucleoside triphosphates as building blocks for extending the 3'-end of a polynucleotide to provide a sequence complementary with the polynucleotide template. Usually, the catalysts are enzymes, such as DNA polymerases, for example, prokaryotic DNA polymerase (I, II, or III), T4 DNA polymerase, T7 DNA polymerase, Klenow fragment, Bst DNA polymerase, reverse transcriptase, Vent DNA polymerase, Pfu DNA polymerase, Taq DNA polymerase, and the like, derived from any source such as cells, bacteria, such as E. coli, plants, animals, virus, thermophilic bacteria, and so forth. RNA polymerases include T7 RNA polymerase, AMV polymerase, Q-beta-replicase, and so forth. In some cases, Bst DNA Polymerase Large Fragment can be used. Bst DNA polymerase Large Fragment is the portion of the *Bacillus stearothermophilus* DNA Polymerase protein that contains the 5′ → 3′ polymerase activity, but lacks the 5′ →3′ exonuclease domain. Where the polymerase forms an extension product on a DNA template, it is referred to herein as a DNA dependent polymerase. Where the polymerase forms an extension product on a RNA template, it is referred to herein as a RNA dependent polymerase.

[00193] In some embodiments, the methods of the present invention provide enzyme inhibitors. In some cases, polymerase inhibitors are provided such as for example DNA-dependent DNA polymerase inhibitors. In some cases, DNA-dependent DNA polymerase inhibitors such as actinomycin that do not substantially inhibit RNA-dependent DNA polymerization are provided. Actinomycin may be particularly useful in the methods of the present invention when the target sequence to be amplified comprises RNA that is present in a sample that also comprises DNA. In such cases actinomycin may be used to selectively inhibit the

IDT Exhibit 1020.3455

production of extension product complementary to the DNA during, for example extension of the first primer.

[00194] Exonucleases are used in the methods of the invention, for example an exonuclease specific for single-stranded DNA, , whereby treatment of the double stranded first and second primer extension products with the single-strand specific exonuclease enable the removal of the single stranded 3' nucleotides from the 3' region of a first primer extension product that are not hybridized to a second primer extension product. In another example an exonuclease may be used to remove unincorporated primers or target nucleic acids that are not hybridized to first or second primer extension products. In yet another example the treatment of the stem loop product of the second primer extension product generated using an all DNA tailed first primer and a chimeric second primer both comprising a sequence A at their 5'-ends, with a single strand specific 3'-exonuclease is used for clean-up of the reaction mixture as described previously, by the selective degradation of the non incorporated primers and other reaction products. Non-limiting examples of an exonuclease include single-strand specific 3'-exonucleases such as exonuclease 1. The exonuclease should remove all of the single-stranded 3' nucleotides which are not hybridized to form a double stranded nucleic acid. In some embodiments, the exonuclease may remove additional 3' nucleotides which are hybridized to a primer extension product. In other embodiments, a polymerase comprising exonuclease activity may be used. Non-limiting examples include a T4 polymerase comprising 3' exonuclease activity. . Exonucleases may be used in the methods of the present invention for clean up of reaction mixtures, such as amplification reaction mixtures, for example single strand specific DNA exonucleases may be used for clean-up of PCR or SPIA amplification products without the need for purification prior to performing sequencing reactions.

[00195] RNases are used in the methods of the present invention, for example in some cases a single and or double stranded RNA specific RNase is used to degrade single and double stranded RNA from template nucleic acid and or unincorporated primers. In some cases, single and double stranded RNA specific RNases includes but is not limited to RNase 1. In another example RNase which is specific for the cleavage of the RNA portion of a DNA-RNA heteroduplex is used in the methods of the present invention to degrade or cleave the RNA portion of the DNA-RNA heteroduplex of a partial heteroduplex generated by the methods of the invention. This may have the effect of making the DNA portion of the DNA-RNA heteroduplex available to hybridize with other oligonucleotides such as but not limited to primers, amplification primers and chimeric oligonucleotides. By way of example and without limitation an example of a RNase specific for cleaving the RNA portion of a DNA-RNA heteroduplex is RNase H.

Nucleic Acid Target

[00196] A nucleic acid target, also referred to as a template, as defined herein includes but is not limited to any length nucleotide, for example, DNA and or RNA. Examples of nucleic acid templates include but are not limited to complementary cDNA, mRNA, total RNA, genomic DNA, and DNA. In some cases, the nucleic acid target may be derived from a natural source such as for example an organism or from the environment. In other cases, the nucleic acid target may be derived from a synthetic source such as a nucleic acid synthesizer. In still other cases, the nucleic acid target may comprise both natural and synthetic nucleic acids in any proportion. In some embodiments of the present invention, target nucleic acid may be purified prior to use by the methods of the present invention. In some cases, target nucleic acid may be pre-fractionated according to size and or sequence by electrophoresis, ion exchange, gel filtration, reverse phase, or hybridization techniques. In some cases, target nucleic acid may be amplified by other methods

IDT Exhibit 1020.3456

known to the art as described herein prior to application of the methods of the present invention. Sources for target nucleic acid include but are not limited to animals including but not limited to humans, non-human primates, dogs, cats, cows, pigs, sheep, guinea pigs, hamsters, birds, frogs, fish, rats, mice, rodents and any other animal from which nucleic acid may be obtained. Other sources for target nucleic acid include but are not limited to bacteria, viruses, plants, insects, worms, fruit flies, and fungi. Target nucleic acid may be obtained from tissue or bodily fluids such as skin, blood, cells, hair follicles, sebaceous glands, cells obtained from fine needle aspiration or other surgical and non surgical biopsy techniques, saliva, urine, stool, menses and the like, or may be obtained from cultures, environmental samples and the like..

Surface or Solid Substrate

[00197] In various exemplary embodiments, a solid surface may have a wide variety of forms, including membranes, slides, plates, micromachined chips, microparticles, beads and the like. Solid surfaces may comprise a wide variety of compositions including, but not limited to, glass, plastic, silicon, alkanethiolate derivatized gold, cellulose, low cross linked and high cross linked polystyrene, silica gel, polyamide, and the like, and can have various shapes and features (e.g., wells, indentations, channels, etc.). As used herein, the terms "solid surface" and "solid substrate" are used interchangeably. In some cases these may be referred to as the surface or the support. The surface can be hydrophilic or capable of being rendered hydrophilic and may comprise inorganic powders such as silica, magnesium sulfate, and alumina; natural polymeric materials, particularly cellulosic materials and materials derived from cellulose, such as fiber containing papers, e.g., filter paper, chromatographic paper, etc.; synthetic or modified naturally occurring polymers, such as nitrocellulose, cellulose acetate, poly (vinyl chloride), polyacrylamide, cross linked dextran, agarose, polyacrylate, polyethylene, polypropylene, poly(4-methylbutene), polystyrene, polymethacrylate, poly(ethylene terephthalate), nylon, poly(vinyl butyrate), etc.; either used by themselves or in conjunction with other materials; glass available as Bioglass, ceramics, metals, and the like. Natural or synthetic assemblies such as liposomes, phospholipid vesicles, and cells can also be employed. The surface can have any one of a number of shapes, such as strip, rod, particle, including bead, and the like.

[00198] In some embodiments, the solid surface comprises a bead or plurality of beads. The beads may be of any convenient size and fabricated from any number of known materials. Example of such materials include: inorganics, natural polymers, and synthetic polymers. Specific examples of these materials include: cellulose, cellulose derivatives, acrylic resins, glass, silica gels, polystyrene, gelatin, polyvinyl pyrrolidone, co- polymers of vinyl and acrylamide, polystyrene cross-linked with divinylbenzene or the like (as described, e. g, in Merrifield, Biochemistry 1964,3, 1385-1390), polyacrylamides, latex gels, polystyrene, dextran, rubber, silicon, plastics, nitrocellulose, natural sponges, silica gels, control pore glass, metals, cross-linked dextrans (e. g., Sephadex) agarose gel (SepharoseTM), polystyrene and other solid phase supports known to those of skill in the art. The beads are generally about 2 to about 100 um in diameter, or about 5 to about 80 pm in diameter, in some cases, about 10 to about 40, um in diameter. In some embodiments the beads can be magnetic. Having magnetic beads can be useful for isolation and purification of the beads comprising nucleic acids described herein. Other methods to separate beads can also be used. For example, the capture beads may be labeled with a fluorescent moiety which would make the nucleic acid- bead complex fluorescent. The target capture bead complex may be separated, for example, by flow cytometry or fluorescence cell sorter.

IDT Exhibit 1020.3457

Attachment of oligonucleotides to solid surfaces

[00199] One aspect of the invention involves attaching oligonucleotides to solid surfaces such that the oligonucleotides can hybridize with polynucleotides produced by the methods of the invention. Many methods of attaching oligonucleotides to surfaces are known. In some embodiments, the oligonucleotide is attached covalently to the solid surface. In some embodiments such attached oligonucleotides act to capture nucleic acids such as amplification products, and in some cases such attached oligonucleotides also act as primers. Methods of attaching oligonucleotides and primers to a surface are known in the art (see, e.g., Beier et al., 1999, Nucleic Acids Res. 27(9):1970-1977; Brison et al., 1982, Molecular and Cellular Biology 2:578 587; Cheung et al., 1999, Nat. Genet 21(1 Suppl):15-19; Chrisey et al., 1996, Nucleic Acids Res. 24(15):3031-3039; Cohen et. al., 1997, Nucleic Acids Res. 1997 Feb. 15; 25(4):911-912; Devivar et al., 1999, Bioorg. Med. Chem. Lett. 9(9):1239-1242; Heme et al., 1997. J. Am. Chem. Soc. 119:8916-8920; Kumar et al., 2000, Nucleic Acis Res. 28(14):e71; Lipshutz et al., 1999, Nat. Genet. 21(1 Suppl):20-24; Milner et al., 1997, Nat. Biotechnol. June;15(6):537-541; Morozov et al., 1999, Anal. Chem. 71(15):3110-3117; Proudnikov et al., 1998, Anal Biochem. 259(1):34-41; Rasmussen et al., 1991, Anal Biochem. 198(1):138-142; Rogers et al., 1999, Anal. Biochem. 266(1):23-30; Salo et al., 1999, Bioconjug Chem. 10(5):815-823; Singh-Gasson et al., 1999, Nat. Biotechnol. 17(10):974-978, and Pierce Chemical Company Catalog 1994, pp. 155-200), incorporated herein by reference). Attached oligonucleotides may also be synthesized directly on the solid surface.

Uses of the present invention

Multiplex Amplification of Desired Nucleic Acid Targets from a Complex Sample

[00200] The methods of the present invention are useful for amplification of multiple different nucleic acid targets from a complex sample. The amplification of multiple different nucleic acid targets from a sample may be referred to as multiplex amplification. In some embodiments, multiplex amplification may be carried out by the methods of the present invention by the use of a set of first primers with different annealing sequences each of which is specific for a different nucleic acid target. In some cases, the template may be removed or degraded after the first primer extension product has been formed by the methods of the present invention. The multiple first primer extension products may then be contacted with a second primer with a random, degenerate, or specific annealing sequence. In other embodiments, the annealing sequence of the first primer may be random or degenerate, while a set of second primers may be used with different annealing sequences each of which is specific for a different nucleic acid target. Multiplex amplified products of the present invention may be detected by sequencing, nucleic acid array hybridization, electrophoresis, capture on solid phase (e.g. beads), or chromatography.

[00201] In a particular embodiment, multiplex amplification of a set of targets in a nucleic acid sample may be carried out to reduce the complexity in that sample. For example, the nucleic acid sample may be a whole genome or whole transcriptome. In some embodiments of the present invention it may be desirable to amplify only a set of the sequences in the whole genome or whole transcriptome such a set may include for example, the exons in a genome, the set of kinases, the set of protein kinases, the set of lipid kinases, proteases, serine proteases, cysteine proteases, oncogenes, transcription factors, polymorphic regions related to a particular disease or condition such as single nucleotide polymorphisms, or any set or subset of target sequences within a sample. In some cases, multiplex amplification may be used to amplify 5 or more sequences from a complex sample, 10 or more sequences, 25 or more, 50 or more, 100 or more sequences,

IDT Exhibit 1020.3458

200 or more sequences, 300 or more sequences, 500 or more sequences, 1000 or more sequences, 2000 or more sequences, 5000 or more sequences, 10,000 or more sequences, 25,000 or more sequences, 50,000 or more sequences, or 100,000 or more sequences. In some cases, the methods of the present invention provide for improved multiplex amplification of target sequences as compared to other methods known in the art such as for example PCR because the linear amplification methods provided herein do not substantially alter the relative abundance of different sequences in a sample by preferentially amplifying some sequences over others. In some cases, the methods of the present invention are useful for processing nucleic acid samples by reducing their complexity by amplifying sequences of interest and/or removing sequences that are not of interest and preparing a sample for downstream processing such as sequencing, quantifying, and array hybridization.

Capture and Amplification of Single or Multiple Nucleic Acid Target or Targets from a Sample

[00202] The methods of the present invention may be used for capture and amplification of single or multiple nucleic acid targets from a sample. In some embodiments, the first primer of the present invention may be attached to a solid phase such as a bead for example and used to capture a nucleic acid target from a sample such as a complex sample by hybridization. In some cases, un-captured nucleic acid may then be removed by washing way the nucleic acid not bound to the solid phase or by removing the solid phase from the nucleic acid that is not bound to the bead, or by a combination thereof. Methods and compositions for hybridization of specific sequences are provided herein. In some embodiments, the first primer and the target nucleic acid may hybridize in solution and the resulting hybridized product may then be captured. In some cases, the capture step may include the use of a solid phase that hybridizes to or has an affinity for a ligand that is bound to, or covalently bound to, the first primer. For example, the first primer may include a tail sequence that is complementary to a sequence bound to a bead. For another example, the first primer may include a ligand such as biotin that may bind to a solid phase (e.g. a bead) derivatized with streptavidin. In still another example, the first primer may include a streptavidin molecule and the solid phase may comprise biotin. In some embodiments the first primer has a specific annealing sequence for capture and amplification of a specific target nucleic acid. In other cases, the first primer may provide a degenerate annealing sequence for capture and amplification of multiple target nucleic acids. In some cases, the first primer may be provided as a set of first primers each with a different specific annealing sequence for capture and amplification of multiple target nucleic acids.

[00203] The captured target nucleic acid may be amplified by the methods provided herein. In some embodiments, a first primer extension product may be formed from the first primer bound to a captured target nucleic acid using the methods of the present invention by the use of a polymerase. The first primer extension product may be contacted with a second primer and a second primer extension product may be formed using the methods of the present invention. The second primer extension product may then be amplified using an amplification primer of the present invention, a polymerase, and RNAse H as described herein. The resulting amplification product may be useful for simplifying complex nucleic acid samples, (i.e. complexity reduction) by capturing and amplifying a subset of the nucleic acid that is desired for further analysis such as sequencing, array based genome or transcriptome analysis, expression profiling and analysis, genetic analysis, SNP analysis etc.

IDT Exhibit 1020.3459

Characterization of Nucleic Acids

[00204] The methods of the invention are amenable to quantitative analysis, as in some embodiments, amplification can yield sufficient single stranded polynucleotide (generally, DNA and RNA) products which accurately reflect the representation of the various DNA or RNA sequences (e.g. genomic DNA or mRNA) in the starting material. The amplified products can be analyzed using, for example, probe hybridization techniques known in the art, such as Northern blotting, and hybridizing to probe arrays. In addition, the single stranded polynucleotide products may serve as starting material for other starting material for other analytical and/or quantification methods known in the art, such as real time PCR, quantitative TaqMan, quantitative PCR using molecular beacons, methods described in Kurn, U.S. Pat. No. 6,251,639, etc. Thus, the invention includes those further analytical and/or quantification methods as applied to any of the products of the methods herein.

[00205] In another embodiment, the amplification methods of the invention are utilized to generate multiple copies of single stranded polynucleotide products from RNA or DNA targets that are labeled by the incorporation of labeled nucleotides during DNA polymerization. For example, amplification according to the methods of the invention can be carried out with suitable labeled dNTPs or rNTPs. These labeled nucleotides can be directly attached to a label, or can comprise a moiety which could be attached to a label. The label may be attached covalently or non-covalently to the amplification products. Suitable labels are known in the art, and include, for example, a ligand which is a member of a specific binding pair which can be detected/quantified using a detectable second member of the binding pair. Thus, amplification of total RNA or mRNA according to the methods of the invention in the presence of, for example, Cy3-dUTP or Cy5-dUTP results in the incorporation of these nucleotides into the amplification products. Labeling can also be accomplished by random priming of amplification products with the incorporation of labeled dNTPs. Other methods for labeling and or fragmenting of the amplification products of the invention are known in the art, for example the methods provided by U.S. Patent Application Serial Nos. 10/304,035 filed November 22, 2002, and 10/441,663 filed May 19, 2003 which are herby incorporated by reference in their entirety and which are useful for the generation of products suitable for analysis using various designs of microarrays, such as for example high density oligonucleotides microarrays .

[00206] The labeled amplified products are suitable for analysis (for example, detection and/or quantification) by contacting them with, for example, microarrays (of any suitable surface, which includes glass, chips, plastic), beads, or particles, that comprise suitable probes such as cDNA and/or oligonucleotide probes. Thus, the invention provides methods to characterize (for example, detect and/or quantify) a DNA or RNA sequence of interest by generating labeled polynucleotide (generally, DNA) products using amplification methods of the invention, and analyzing the labeled products. Analysis of labeled products can be performed by, for example, hybridization of the labeled amplification products to, for example, probes immobilized at, for example, specific locations on a solid or semi-solid substrate, probes immobilized on defined particles, or probes immobilized on blots (such as a membrane), for example arrays. Other methods of analyzing labeled products are known in the art, such as, for example, by contacting them with a solution comprising probes, followed by extraction of complexes comprising the labeled amplification products and probes from solution. The identity of the probes provides characterization of the sequence identity of the amplified products, and thus by extrapolation the identity of the target DNA or target RNA present in a sample. Hybridization of the labeled products is detectable, and the amount of specific labels that are detected is proportional to the amount of the labeled amplification products of a specific DNA or

IDT Exhibit 1020.3460

RNA sequence of interest. This measurement is useful for, for example, measuring the relative amounts of the various RNA species in a sample, which are related to the relative levels of gene expression, as described herein or to detect the presence or absence of defined target DNA or RNA in a sample. The measurement is also useful for measuring the relative amounts of various DNA sequences corresponding, for example, to genetic regions in the sample. The amount of labeled products (as indicated by, for example, detectable signal associated with the label) hybridized at defined locations on an array can be indicative of the detection and/or quantification of the corresponding target DNA or target RNA species in the sample.

Sequencing of the polynucleotide products of the invention

[00207] As described above, the methods can be used to obtain sequence information about a target RNA or target DNA of interest. The sequencing can be carried out on the primer extension products or amplification products produced by the methods herein. In some embodiments the sequencing is performed on the polynucleotides attached to solid surfaces as described herein. In one embodiment sequencing is performed on polynucleotides that are attached to the beads through oligonucleotides attached to the beads which capture amplified product, and are extended to produce a polynucleotide attached to the surface comprising a defined sequence at its 3' end.

[00208] The methods of the invention are useful, for example, for sequencing of a polynucleotide sequence of interest. The sequencing process may be carried out as described for the methods described herein. In some cases, the methods of the invention are useful for providing amplified products having defined 3' and 5' ends that are useful for further analysis such as for example for sequencing, array hybridization, or quantitative PCR. In some cases, the methods provide for multiplex amplification such as from complex mixtures such as whole genome or whole transcriptome or a portion thereof, or multiple whole genomes or whole transcriptomes or portions thereof.

[00209] Known methods for sequencing include, for example, those described in:, Sanger, F. et al. , Proc. Natl. Acad. Sci. U. S. A. 75,5463-5467 (1977); Maxam, A. M. & Gilbert, W. Proc Natl Acad Sci USA 74,560-564 (1977); Ronaghi, M. et al. , Science 281, 363,365 (1998); Lysov, 1. et al., Dokl Akad Nauk SSSR 303, 1508-1511 (1988); Bains W. & Smith G. C. J. Theor Biol 135,303-307 (1988); Drnanac, R. et al. , Genomics 4,114-128 (1989); Khrapko, K. R. et al. , FEBS Lett 256.118-122 (1989); Pevzner P. A. J Biomol Struct Dyn 7, 63-73 (1989); and Southern, E. M. et al., Genomics 13, 1008-1017 (1992). Pyrophosphate-based sequencing reaction as described, e. g., in US patent 6,274,320; 6,258, 568 and 6,210,891), may also be used.

[00210] In some cases, the methods and compositions of the present invention may be useful for capture and amplification of a plurality of target sequences from a complex sample, such as for example a whole genome or genomes, or a whole transcriptome or transcriptomes. The capture and amplification of target sequences from a complex sample may be desirable for complexity reduction. For example, complexity reduction may be advantageous for further use of the amplified products in highly parallel analytical techniques such as the use of high density microarrays, or high throughput sequencing methods including for example the use of the methods and instrumentation developed by Roche 454 Life Sciences, Illumina, Sequenom, ABI, or any of the sequencing methods described herein such as those described in US Patent Nos. 7,211,390; 6,355,431; 5,750,341; 5,969,119; 6,274,320; 6,258, 568; 6,210, 891; and 6,306,597. In

IDT Exhibit 1020.3461

some embodiments, this complexity reduction may be used to capture and in some cases amplify a single RNA or DNA sequence of interest from a complex sample.

[00211] The methods of the invention are useful, for example, for sequencing of an RNA sequence of interest. The sequencing process can be carried out by processing and amplifying a target RNA containing the sequence of interest by any of the methods described herein. Addition of nucleotides during primer extension can be analyzed using methods known in the art, for example, incorporation of a terminator nucleotide or sequencing by synthesis (e.g. pyrosequencing).

[00212] In embodiments wherein the end product is in the form of DNA primer extension products, in addition to the nucleotides, such as natural deoxyribonucleotide triphosphates (dNTPs), that are used in the amplification methods, appropriate nucleotide triphosphate analogs, which may be labeled or unlabeled, that upon incorporation into a primer extension product effect termination of primer extension, may be added to the reaction mixture. Preferably, the dNTP analogs are added after a sufficient amount of reaction time has elapsed since the initiation of the amplification reaction such that a desired amount of second primer extension product or fragment extension product has been generated. The amount of time can be determined empirically by one skilled in the art.

[00213] Suitable dNTP analogs include those commonly used in other sequencing methods and are well known in the art. Examples of dNTP analogs include dideoxyribonucleotides. Examples of rNTP analogs (such as RNA polymerase terminators) include 3'-dNTP. Sasaki et al., Biochemistry (1998) 95:3455-3460. These analogs may be labeled, for example, with fluorochromes or radioisotopes. The labels may also be labels which are suitable for mass spectroscopy. The label may also be a small molecule which is a member of a specific binding pair, and can be detected following binding of the other member of the specific binding pair, such as biotin and streptavidin, respectively, with the last member of the binding pair conjugated to an enzyme that catalyzes the generation of a detectable signal that could be detected by methods such as colorimetry, fluorometry or chemiluminescence. All of the above examples are well known in the art. These are incorporated into the primer extension product or RNA transcripts by the polymerase and serve to stop further extension along a template sequence. The resulting truncated polymerization products are labeled. The accumulated truncated products vary in length, according to the site of incorporation of each of the analogs, which represent the various sequence locations of a complementary nucleotide on the template sequence.

[00214] Analysis of the reaction products for elucidation of sequence information can be carried out using any of various methods known in the art. Such methods include gel electrophoresis and detection of the labeled bands using appropriate scanner, sequencing gel electrophoresis and detection of the radiolabeled band directly by phosphorescence, capillary electrophoresis adapted with a detector specific for the labels used in the reaction, and the like. The label can also be a ligand for a binding protein which is used for detection of the label in combination with an enzyme conjugated to the binding protein, such as biotin-labeled chain terminator and streptavidin conjugated to an enzyme. The label is detected by the enzymatic activity of the enzyme, which generates a detectable signal. As with other sequencing methods known in the art, the sequencing reactions for the various nucleotide types (A, C, G, T or U) are carried out either in a single reaction vessel, or in separate reaction vessels (each representing one of the various nucleotide types). The choice of method to be used is dependent on practical considerations readily apparent to one skilled in the art, such as the nucleotide tri phosphate analogs and/or label used. Thus, for example, when each of the analogs is differentially labeled, the sequencing reaction can be carried out in a single vessel. The

IDT Exhibit 1020.3462

considerations for choice of reagent and reaction conditions for optimal performance of sequencing analysis according to the methods of the invention are similar to those for other previously described sequencing methods. The reagent and reaction conditions should be as described above for the nucleic acid amplification methods of the invention.

[00215] In other embodiments, the methods of the present invention may be useful for sequencing by synthesis methods such as pyrosequencing. In some cases, high throughput pyrosequencing may be carried out by the methods described in Margulies et al. 2005 *Nature* 437, 376-380. In this method, target nucleic acid such as genomic DNA is fragmented and oligonucleotide linkers are attached. Adapters are ligated to the fragments and the fragments are bound to beads. These immobilized fragments are then amplified. Amplification results in a large number of copies of each immobilized fragment. The amplified immobilized fragments are then sequenced by pyrosequencing as outlined in Ronaghi et al. 1996 *Analytical Biochemistry* 242, 84-89. Ronaghi describes combining a template, a sequencing primer, dNTPs, an ATP sulfurylase, a DNA polymerase, and a luciferase. Addition of dNTPs to the 3' end of the sequencing primer by the DNA polymerase results in the release of pyrophosphate. Pyrophosphate is used as a substrate by ATP sulfurylase to create ATP. ATP is then used as a substrate by luciferase to generate light which is detected by a sensor. The sequence is then determined by correlating the generation of light by luciferase with the identity of the dNTP added to the reaction. For example, if a dCTP is added to the reaction, and light is generated then the sequenced strand comprises the complement guanosine at that position. The methods of the present invention may also be useful for high throughput sequencing by ligation, sequencing by hybridization and sequencing by synthesis using reversible dyes. Such sequencing methods are well known in the art and have been described in the literature such as for example in *Nature Biotechnology* **26**, 1135 - 1145 (2008).

[00216] In some cases, the amplification step may be performed by the methods of the present invention. In other cases, the methods of the present invention are used to immobilize the target nucleic acid onto the beads as depicted in **Figure 4** for example. In some cases, the methods of the present invention may be used to immobilize a sequence specific subset of the target nucleic acid onto beads for high throughput sequencing. Sequence specific immobilization may be particularly useful for reducing the number of different sequences to be examined in a single sequencing run. In some cases, this may be advantageous for increasing the redundancy of the sequences obtained, or increasing the number of overlapping sequences.

Determination of Gene Expression Profile

[00217] The amplification methods of the invention can be used for use in determining the levels of expression of multiple genes in a sample since the methods described herein are capable of amplifying multiple target RNAs in the same sample. As described above, amplification products can be detected and quantified by various methods, as described herein and/or known in the art. Since RNA is a product of gene expression, the levels of the various RNA species, such as whole transcriptome or total RNAs, in a sample is indicative of the relative expression levels of the various genes (gene expression profile). Thus, determination of the amount of RNA sequences of interest present in a sample, as determined by quantifying amplification products of the sequences, provides for determination of the gene expression profile of the sample source.

[00218] In some embodiments, the methods of the present invention allow for the storage and subsequent analysis of samples, allowing for a sample to be bound to a solid substrate for archiving, then later to be analyzed by

IDT Exhibit 1020.3463

the methods described herein to determine a gene expression profile. In some embodiments, the sample can be analyzed multiple times, and stored between analyses.

[00219] Accordingly, the invention provides methods of determining gene expression profile in a sample, the method comprising: amplifying single stranded product from at least one RNA sequence of interest in the sample, using any of the methods described herein; and determining amount of amplification products of each RNA sequence of interest, wherein each amount is indicative of the amount of each RNA sequence of interest in the sample, whereby the expression profile in the sample is determined. Generally, labeled products are generated. In one embodiment, the target RNA is mRNA. It is understood that amount of amplification product may be determined using quantitative and/or qualitative methods. Determining amount of amplification product includes determining whether amplification product is present or absent. Thus, an expression profile can includes information about presence or absence of one or more RNA sequence of interest. "Absent" or "absence" of product, and "lack of detection of product" as used herein includes insignificant, or de minimus levels.

[00220] The methods of expression profiling are useful in a wide variety of molecular diagnostic, and especially in the study of gene expression in essentially any mammalian cell (including a single cell) or cell population. A cell or cell population (e.g. a tissue) may be from, for example, blood, brain, spleen, bone, heart, vascular, lung, kidney, pituitary, endocrine gland, embryonic cells, tumors, or the like. Expression profiling is also useful for comparing a control (normal) sample to a test sample, including test samples collected at different times, including before, after, and/or during development, a treatment, and the like.

Libraries

[00221] In another embodiment, the invention encompasses a library comprising a plurality of nucleic acid molecules. In some cases, each nucleic acid molecule is separately immobilized to a different bead. In another embodiment, the invention encompasses a library comprising a plurality of nucleic acid molecules, wherein each nucleic acid molecule is separately immobilized to a different bead and wherein each bead comprises over 100, 000 clonal amplification copies of each nucleic acid molecule, wherein the library is contained in a single vessel. As examples, the nucleic acid molecules may be genomic DNA, cDNA, episomal DNA, BAC DNA, or YAC DNA. The genomic DNA may be animal, plant, viral, bacterial, or fungal genomic DNA. Preferably, the genomic DNA is human genomic DNA or human cDNA.

Genetic Analysis

[00222] The methods of the present invention may be used in the analysis of genetic information. Amplification methods as disclosed herein may be used in the devices, kits, and methods known to the art for genetic analysis such as but not limited to those found in US Patent Nos. 6449562, 6287766, 7361468, 7414117, 6225109, 6110709. In some cases, amplification methods of the present invention may be used to amplify target nucleic acid for DNA hybridization studies to determine the presence or absence of polymorphisms. The polymorphisms, or alleles, may be associated with diseases or conditions such as genetic disesase. In other cases the polymorphisms may be associated with susceptibility to diseases or conditions, for example polymorphisms associated with addiction, degenerative and age related conditions, cancer, and the like. In other cases, the polymorphisms may be associated with beneficial traits such as increased coronary health, or resistance to diseases such as HIV or malaria, or resistance to degenerative diseases such as osteoporosis, Alzheimer's or dementia.

IDT Exhibit 1020.3464

KITS

[00223] One aspect of the invention comprises kits useful for carrying out the methods of the invention.

[00224] The kits of the invention can comprise the components described for the methods disclosed herein. The components can be packaged in a manner that allows for efficiently carrying out the methods of the invention.

[00225] In one aspect, the kit comprises (a) a first primer comprising a 3'-DNA portion and a 5'-DNA portion, wherein the 3'-DNA portion comprises a random sequence or a specific sequence, and the 5' DNA portion further comprises sequence (A), and (b) a second primer comprising a 5'-RNA portion and a 3' DNA portion, wherein the 3' DNA portion comprises a random sequence or a specific sequence and the 5' RNA portion further comprises sequence (A) or (B). In some embodiments the kit may further comprise (c) an RNA dependent DNA polymerase, (d) a DNA dependent DNA polymerase with strand displacement activity, (e) RNase H, (f) an amplification chimeric primer comprising a 3'-DNA portion and a 5'-RNA portion wherein the sequence of the amplification primer is the substantial the same sequence as the (A) or (B) sequence, or a combination thereof. In some embodiments of this kit the RNA dependent DNA polymerase of component (c) possesses substantial strand displacement activity. In other embodiments, the RNA dependent DNA polymerase of kit component (c) does not possess substantial strand displacement activity. In some cases, component (d) consists of two separate DNA dependent DNA polymerases (d1) and (d2) one of which possesses substantial strand displacement activity and one of which does not possess substantial strand displacement activity. In some cases, the kit does not contain an RNA dependent DNA polymerase.

[00226] In some cases, a kit provided by the present invention may comprise an RNA-dependent DNA polymerase and a DNA-dependent DNA polymerase with and without strand displacement activity as described above. In some cases, the kit further comprises single strand specific 3' exonuclease. In some cases, the kit further comprises an RNase that is specific for single or double-stranded RNA and devoid of RNase H activity (i.e. does not substantially degrade the RNA portion in an RNA-DNA heteroduplex).

[00227] In one embodiment, the second primer comprises the tail sequence (A), and the amplification primer is substantially the same sequence as sequence (A). These kits are useful for the formation and purification of stem loop structures as described herein. In some embodiments, these kits comprise the enzymes that degrade single stranded DNA and RNA as described herein for clean-up of reaction mixtures comprising the stem loop structures. In some cases, the kit further contains instructions for the use of the kit.

[00228] In one aspect, the kit comprises (a) a chimeric oligonucleotide comprising a 3'-DNA portion substantially comprising sequence (A) or (B) and a 5'-RNA sequence (C), and (b) a first primer comprising a 3'-DNA portion and a 5'-DNA portion, wherein the 3'-DNA portion comprises a random or a specific sequence, and the 5' DNA portion comprises sequence (A). In some embodiments, the kit may further comprise a second primer, such as for example an RNA-DNA chimeric second primer, a polymerase having substantial strand displacement activity, RNase H, or a combination thereof.

[00229] In one aspect, the kit comprises (a) a chimeric oligonucleotide comprising a 3'-DNA portion substantially comprising sequence (A) or (B) and a 5'-RNA sequence (C), (b) a first primer comprising a 3'-DNA portion and a 5'-DNA portion, wherein the 3'-DNA portion comprises a random or a specific sequence, the 5' DNA portion comprises sequence (A), and the 5' DNA portion further comprises a ligand, and (c) a second primer comprising a 5'-RNA portion and a 3' DNA portion, wherein the 3' DNA portion comprises

IDT Exhibit 1020.3465

a random sequence or a specific sequence and the 5' RNA portion further comprises sequence (A) or (B). In some embodiments, the kit may further comprise (e) an RNA dependent DNA polymerase, (d) a DNA dependent DNA polymerase with strand displacement activity, (f) RNase H, (g) a chimeric amplification primer comprising a 3'-DNA portion and a 5'-RNA portion, wherein the chimeric amplification primer comprises a sequence which is substantially the same as sequence (C), or a combination thereof. In some embodiments of this kit the RNA dependent DNA polymerase of component (c) possesses substantial strand displacement activity. In other embodiments, the RNA dependent DNA polymerase of kit component (c) does not possess substantial strand displacement activity. In some cases, component (d) consists of two separate DNA dependent DNA polymerases (d1) and (d2) one of which possesses substantial strand displacement activity and one of which does not possess substantial strand displacement activity. In some cases, the kit does not contain an RNA dependent DNA polymerase. In some cases, the kit further contains instructions for the use of the kit.

[00230] In some embodiments the kit further comprises solid surface with an oligonucleotide attached to the surface by the 5'-end comprising a sequence (A) or (B) and further comprising a random or specific annealing sequence (P). In another embodiment the oligonucleotide attached to the solid surface comprises a sequence hybridizable to sequence (A) and the oligonucleotide is attached by the 5'-end.

[00231] In one aspect, the kit comprises (a) a first primer comprising a 3'-DNA portion and a 5'-DNA portion, wherein the 3'-DNA portion comprises a random sequence or a specific sequence, and the 5' DNA portion further comprises sequence (A), (b) a second primer comprising a 5'-RNA portion and a 3' DNA portion, wherein the 3' DNA portion comprises a random sequence or a specific sequence and the 5' RNA portion further comprises sequence (A) or (B), (c) a DNA dependent DNA polymerase, (d) a DNA dependent DNA polymerase with strand displacement activity, (e) RNase H, and (f) an amplification chimeric primer comprising a 3'-DNA portion and a 5'-RNA portion wherein the sequence of the amplification primer is substantially the same sequence as the (A) or (B) sequence. In some cases, the kit further contains instructions for the use of the kit.

[00232] In some cases, a kit provided by the present invention may comprise (a) a first primer comprising a 3'-DNA portion and a 5'-DNA portion, wherein the 3'-DNA portion comprises a random sequence or a specific sequence, and the 5'-DNA portion further comprises sequence (A) or (B). In some cases, the kit further comprises a ligand attached to the 5'-end of the first primer. The kit may further include (b) a tailed DNA-RNA chimeric second primer as described previously. In some cases, the kit may further include (c) a DNA or RNA dependent DNA polymerase, as described previously. In some cases, the kit may further include (d) a DNA dependent DNA polymerase with strand displacement activity. In some cases, the kit may further include (e) RNase H, and (f) an amplification chimeric primer comprising a 3'-DNA portion and a 5'-RNA portion wherein the sequence of the amplification primer is substantially the same as the (A) or (B) sequence.

[00233] In some embodiments, the kits useful for carrying out the methods of the invention may comprise (a) non-tailed DNA first primers comprising a specific or random annealing sequence, (b) tailed DNA-RNA chimeric second primers with specific or random annealing sequence and a tail sequence (A) or (B), (c) a DNA or RNA dependent DNA polymerase, (d) a DNA dependent DNA polymerase with strand displacement activity, (e) RNase H, and (f) an amplification chimeric primer comprising a 3' DNA portion and a 5' RNA portion wherein the sequence of the amplification primer is substantially the same sequence as the (A) or (B) sequence.

IDT Exhibit 1020.3466

[00234] In some embodiments the kits useful for carrying out the methods of the invention may further comprise an inhibitor of the DNA dependent DNA polymerase, such as Actinomycin. In some embodiments the kits useful for carrying out the methods of the invention may further comprise a single stranded DNA specific 3' exonuclease such as exonuclease 1, and or a single or double stranded RNA specific RNase such as RNase 1.

[00235] In one aspect, the kit comprises (a) reagents for forming an emulsion and (b) a DNA polymerase with substantial strand displacement activity. Reagents for forming a suitable water in oil emulsion are known and commercially available for example in emPCR kits II and III (454/Roche LifeSciences). Said emulsion forming reagents may include for example decamethylcyclopentasiloxane, polyphenylmethylsiloxane, water and/or buffer. In some cases, the kit may further comprise (c) one or more RNA-DNA chimeric primers, (d) an all DNA primer, (e) a solid surface such as a bead or set of beads, a substantially planar array, a well or wells in a plate, or an isolated surface or set of isolated surfaces, (f) RNase H, (g) a chimeric oligonucleotide, or a combination thereof. In some cases, the kit may further comprise instructions for the use of said kit.

[00236] The components of the kits may comprise the same aspects and embodiments as described above for the components in the description of methods. For example, the ligands and receptors, the primers, the enzymes and the oligonucleotides can be those described herein to carry out the methods of the invention.

[00237] The kits and methods herein can be used for preparing DNA and/or RNA samples for massively parallel sequencing, e.g., of whole genomes or parts of genomes as well as RNA expression analysis for whole transcriptome or parts of a transcriptome. Such kits and methods can be used to diagnosis, prognosis and/or theranostics of conditions including but not limited to cancer, inflammation, fetal abnormality, cardiovascular conditions, etc.

Examples

Example 1: Target specific amplification of multiple genomic sequences on E. coli genomic DNA

All-DNA tailed first primer

[00238] All DNA tailed first primers comprising 3'-end sequence complementary to defined E. coli genomic DNA sequence, and a tailed 5'-end which is not hybridizable to E. coli DNA, and is common to all tailed first primers are used in this example. Four different first all DNA tailed primers are used for this example, as listed in Table 1. The tail sequence of the first primer is the same sequence as that used for the 5'-end tail of the second chimeric primer. The second chimeric DNA-RNA primer comprises a 3'-end random DNA sequence which is hybridizable at random sites on the target DNA, and a 5'-RNA tail which is not hybridizable to the target DNA and is the same sequence as the 5'-tail of the all DNA first primer. Target specific amplification is carried out using the following protocol:

1. <u>First primer extension reaction mixture</u>:
- 1 µl of 1ng E. coli genomic DNA
- 1 µl of 1 µM all DNA first primer
- 1 µl of 10X standard Taq polymerase buffer (NEB)
- 0.2 µl 10mM dNTPs (each)
- 0.2 µl Taq DNA polymerase (NEB, 5 Units/µl)
- Water to a total of 10 µl

The reaction mixtures are incubated under the following temperature and durations in a thermocycler (MJ): 94^{0}C for 2 min, 50^{0}C for 30 sec, 72^{0}C for 5 min, and cooled down to 4^{0}C.

IDT Exhibit 1020.3467

2. Second chimeric primer annealing and extension:

2 µl of second tailed chimeric primer (50 µM stock-solution) is added to the first reaction mixtures and the mixtures are incubated in a thermocycler and heated to 98°C for 3 min (Denaturation) and cooled to 4°C. 8 µl of the following master mixture (T4 polymerase reaction mixtures) is added to each reaction mixture (above):

- 2 µl 10X NEB Buffer 2
- 0.1 µl dNTPs
- 0.2 µl 100X BSA (NEB)
- 0.2 µl T4 DNA polymerase (NEB, 3U/ µl)
- 5.5 µl water

The reaction mixtures are incubated in a thermocycler at the following temperatures and durations: 25°C for 5 min, 37°C for 30min, 95°C for 5 min (inactivation of T4 DNA polymerase) and slowly cooled down to 4°C.

3. Digestion with Exo 1:

0.5 µl of Exo 1 (NEB, 20 U/µl) are added to each reaction mixture followed by incubation in a thermocycler at the following temperature and durations:
25°C for 10 min, 37°C for 30 min, 80°C for 5 min (inactivation of Exo 1) and cooled down to 4°C.

4. Purification of the reaction products using Agencourt Magnetic beads

Reaction products are purified using an Agencourt® AMPure® kit according to the manufacturer's protocol.

5. SPIA amplification

[00239] SPIA amplification is carried out using buffer and enzyme mixtures from NuGEN's WT-Ovation Pico RNA amplification system (NuGen Technologies Inc, San Carlos CA, http://www.nugeninc.com/nugen/index.cfm/products/amplification-systems/wt-ovation-pico/):
3 µl chimeric amplification primer (50 µM stock; see sequence information below)

- 17 µl water
- 40 µl amplification buffer (WT-Ovation Pico System)
- 20 µl Amplification Enzyme Mix (WT-Ovation Pico System)

The reactions are incubated in a thermocycler at 50°C for 60 min, followed by 95°C for 5 min, and cooled down.

Results:

[00240] Quantification of amplification products is carried out by Real Time qPCR with SYBR Green, using MJ Opticon. Amplification reactions are diluted 1:100 in Tris-EDTA and 2 µl of the diluted DNA are analyzed by Real Time qPCR, with the corresponding forward and reverse PCR primer pairs as detailed in Table 1.

[00241] Quantification of the various E. coli genes is given as Ct value (inversely related to concentration of the interrogated sequence). The results as summarized in Figure 19 demonstrate specific amplification of the sequence targeted by the first all DNA tailed primer, with minimal amplification of other genes. The Ct value for the other genes, not targeted for amplification, reflects the input genomic DNA (input into the amplification reaction according to the method of the invention).

IDT Exhibit 1020.3468

[00242] The amplification amplitude of the targeted genes in the various reactions is indicated by delta Ct. The quantification of each of the non-targeted genes is calculated by the Avg Ct of the specific qPCR reaction of the reactions which do not target the specific gene. Delta Ct is the difference of the avg. Ct of the gene in the non targeted reactions and the Avg Ct for the specified gene in the targeted amplification reactions.

Target gene	Tailed, all DNA first primer sequence	PCR primer	PCR primer sequence
E. coli K12 dnaJ gene encoding a heat shock protein M12565	*GGTAATACGACTCACTATAGGCAGA*CACTGTCAGGGC CGCGGTACG*	EKGPF 1	GGGATTTTAACGGAC AGC
		EKGPR 1	CTGATCAAAGATCCG TGC
E. coli K12 dnaJ gene encoding a heat shock protein M12565	*GGTAATACGACTCACTATAGGCAGA*CGTACCGCGGCC CTGACAGTG*	EKGPF 2	TGGTTTTGCACCGCT AC
		EKGPR 2	GCTGTACGTGGCGTG AC
E.coli ATP-dependent C1p protease (clpA) gene M31045	*GGTAATACGACTCACTATAGGCAGA*GGCGACGGCAGA GAACCTGAATAGC*	clpA-F1	CTGCTGTTTGGTTCGC TGGT
		clpA-R1	GCTGCTTCCGCCTTGT GCT
E.coli ATP-dependent C1p protease (clpA) gene M31045	*GGTAATACGACTCACTATAGGCAGA*GCTATTCAGGTT CTCTGCCGTCGCC*	clpA-F2	CTCGATGTGGTGAAC TTTATCT
		clpA-R2	CTTGTTCTTCGCTGTT TGGCT

Non-amplified Ct values and the delta Ct of the targeted genes are shown in Figure 20. The delta Ct values for the four E. coli genes is very similar, 11.5 to 14 cycles corresponding to 2823-15654 fold amplification.

Table 1

*the 5'-tailed sequence is denoted by the underlined italic letters.

Table 2: Chimeric Primers

Chimeric Primer	Chimeric Primer Sequence
Second tailed chimeric primer	***gguaauacgacucacuauaggcaga*NNNNNN***
Chimeric amplification primer	***gguaauacgacucacuau*AGGCAGA***

* Lower case letters denotes ribonucleotides; Upper case letters denote deoxynucleotides

IDT Exhibit 1020.3469

Example 2: Single and multiplex sequence specific amplification

All DNA sequence specific tailed first primer

[00243] Single and multiplexed target specific amplification utilizing the methods of the invention is demonstrated in this example. Four specific gene sequences on genomic yeast DNA are targeted.

[00244] Target specific tailed all DNA first primers used for this example as well as the qPCR primer pairs utilized for quantification of the amplification products are described in Table 3. The input for all amplification reactions is 1ng genomic yeast DNA. Two experiments employing different enzymes and buffer system in the first and second primer extension reactions, prior to the linear isothermal amplification steps are described below.

[00245] I. The first set of reactions described herein, for single and multiplexed amplification of specific genomic sequences, are carried out as follows:

a. Genomic yeast DNA, 1ng/µl (1µl) is mixed with 1 µl of 10X Standard Taq buffer (NEB), 0.2 µl dNTPs (10mM each), 0.2 µl Taq DNA polymerase (NEB) 1 µl of the specified tailed target specific first primer (one each or a combination of primers), and 6.6 µl water, on ice. The reaction mixtures are incubated in a thermocycler at the following temperatures and durations: 94^{0}C for 2 min., 50^{0}C for 1 min., 72^{0}C for 5 min., and cooled down to 4^{0}C.

b. 2µl of a 50 µM stock solution of a chimeric second primer (the same sequence as that used in the E. coli example above), is added to each reaction mixture and the reactions are subjected to heat denaturation (98^{0}C for 3' min.) and cooled to 4^{0}C.

c. Second primer extension is carried out using T4 DNA polymerase (NEB) following the addition of the following reagents: 2µl of 10X NEB buffer 2 (T4 polymerase buffer), 0.1 µl 10mM dNTPs (each), 0.2µl 100X BSA (NEB), 0.2µl T4 DNA polymerase (NEB) and 5.5µl water. The reactions are incubated in a thermocycler at the following temperatures and durations: 25^{0}C for 5 min., 37^{0}C for 30 min., 98^{0}C for 3 min., and slowly cooled to 4^{0}C.

d. The products of the second primer extension reactions are than digested utilizing exonuclease 1 (0.5µl for each reaction) and RNase I (0.5µl for each reaction). The digestion reactions are carried out at 37^{0}C for 30 min. in a thermocycler, followed by inactivation of the enzymes (80^{0}C for 20 min.).

e. The reaction products are purified using Agencourt magnetic beads as per the manufacturer's instructions.

f. Isothermal linear amplification (SPIA) is carried out in a reaction mixture containing the above purified reaction products (10µl), 2µl chimeric amplification primer (100µM stock solution; as in the E. coli example), 18µ water, 40µl Amplification Buffer and 20µl Amplification Enzyme Mixture (from NuGEN Technologies Inc., as used in the WT-Ovation Pico RNA Amplification System). Amplification reactions are carried out as described in Example 1.

Results:

[00246] Quantification of the four targeted genes, targeted or not in the amplification reactions analyzed, is carried out by real-time PCR as described in Example 1, using the PCR primer pairs specific for the targets of interest (Table 3). Average Ct values (all reactions are carried out in triplicates or duplicates) are inversely related to the concentration of the target of interest in the reactions. Figure 21 depicts the results of Real-Time PCR quantification of the amplification products described above. The Avg Ct values for the different targets in the various reactions are grouped according to the type of tailed all DNA target specific first primer (or primers) employed. As is shown in Figure 20, amplification of the targeted gene of interest

IDT Exhibit 1020.3470

without amplification of the non targeted genes, is achieved in reactions carried out employing a tailed all DNA first primer specific for the target of interest. Thus, for example, reactions carried out employing the tailed all DNA first primer specific for a sequence of the MRK1 gene result in amplification of the MRK1 gene but does not lead to amplification of the non targeted genes PFK1, HAC1 and LCD. The amplification reactions carried out with single tailed all DNA first primer thus results in amplification of the single sequence of interest, as shown by the first four reactions, marked on the X axis of Figure 20 as Tailed-MRK1, Tailed-PFK1, Tailed-HAC1 and Tailed-LCD (demoting the specific tailed all DNA first primer employed in the amplification reaction according to the current invention).

[00247] The method of the invention also provide for simple and efficient multiplexed amplification of sequences of interest. As shown in Figure 21, efficient and equal amplification of four yeast genes in a sample comprising yeast genomic DNA, is achieved in reactions employing the four specific tailed all DNA primer (marked as "all 4"). The Avg Ct values for all four genes are similar and lower than that obtained for reactions carried out without the specific tailed all DNA first primer.

[00248] II. Single and multiplexed amplification of specific targeted yeast genomic sequences using the method of the invention: Amplification reactions are carried out similarly to the reactions described in I., with the exception that Taq DNA polymerase is employed for first and second primer extension steps and the buffer used in both reactions steps is ThermoPol (NEB). The second primer extension step is carried out by incubation of the reactions mixtures in a thermocycler at the following temperatures and durations: 50°C for 5 min., 55°C for 1 min., 72°C for 5 min., 98°C for 3' and cooled slowly to 40°C.

[00249] Single sequence specific amplification is carried out using the tailed all DNA first primer specific for the amplification of MRK1 gene sequence, Tailed-MARK1 (Table 3) and multiplexed amplification reactions are carried out using all four tailed all DNA target specific first primers: Tailed-MRK1, Tailed-PFK1, Tailed-HAC1 and Tailed-LCD (Table 3).

[00250] Results: Single or multiplexed amplification of the sequences of interest is determined by quantification of the targeted sequence by Real-Time PCR as described above. Avg Ct values for the various targeted and non targeted gene sequences (non-targeted in the case of amplification of a single targeted sequence, MRK1) are shown in Figure 22. Efficient and specific single and multiplexed amplification is achieved using Taq DNA polymerase, without strand displacement, for both primer extension products according to the method of the invention.

Table 3.

Seq name	Definition	Tailed first primer sequence * (5' to 3')	Q-PCR primers	Q-PCR primer sequences (5' to 3')	
MRK1	Saccharomyces cerevisiae putative protein kinase (MRK1) gene (U22348)	GGTAATACGACTCACTATAG GCAGAaatgcttgaaaccggatcag	Forward	MRK-F	ctcccgaactcatgtttg gt
			Reverse	MRK-R	cccaaaagcaactcag caat
PFK1	Saccharomyces cerevisiae phosphofructoki nase, alpha subunit (PFK1) gene (M26943)	GGTAATACGACTCACTATAG GCAGAccggtttaatcactggtgct	Forward	PFK-F	aacttccgtcacgacaa agg
			Reverse	PFK-R	tgatgtcagccaacaatt

IDT Exhibit 1020.3471

					gag
HAC1	Saccharomyces cerevisiae HAC1 gene for basic leucine zipper protein Hac1 (D26506)	GGTAATACGACTCACTATAG GCAGAaactggctgaccacgaagac	Forward	HAC-F	tcgcactcgtcgtctgat ac
			Reverse	HAC-R	acaaagtcgaggctcc attg
LCD	Synthetic construct Saccharomyces cerevisiae clone FLH202554.01X LCD1 gene (DQ332471)	GGTAATACGACTCACTATAG GCAGAtcaagcaggtagaaagccaa	Forward	LCD-F	tatgccaaaccgaaact gtg
			Reverse	LCD-R	caaggtgtcgatgatct ctca

* Capital underlined letter denote the 5;-end tail (all DNA tailed primer)

Example 3: Sequence specific multiplex amplification of RNA transcripts

All DNA tailed first primer

[00251] A subject presents to a health care provider symptoms indicative of lung cancer, including coughing up blood and shortness of breath. A chest X-ray is performed on the subject and a suspicious mass is detected in the lungs of the subject. A needle biopsy is performed on the subject to obtain tissue from the suspicious mass for further analysis. The biopsied tissue recovered from the subject is processed to extract and purify total RNA using a commercially available Qiagen RNeasy kit according to the manufacturer's instructions.

[00252] 500pg of total RNA representing at least a portion of the transcriptome of the biopsied material is amplified by the methods of the present invention as described herein using: 10 pmol of all-DNA first primers comprising a 5' segment and a 3' segment. The 3' DNA segment of these first primers further comprise an annealing sequence that is complementary to the sequence(s) of interest. In this example, three different first primers with annealing sequences specific for abl, ras, and her2 transcripts are included in a single primer extension reaction. The 5' DNA segment of the first primers comprise a tag sequence (A). Reaction mixture and reaction conditions can be those described by the protocols for the WT-Ovation Pico RNA Amplification system, the details of which are available on NuGEN's website: http://www.nugeninc.com/tasks/sites/nugen/assets/File/quick_protocols/qp_wt_ov_pico.pdf and http://www.nugeninc.com/tasks/sites/nugen/assets/File/user_guides/userguide_wt_ov_pico.pdf

[00253] The reaction mixture can be heated to 75°C for 2 minutes and cooled. 1µl of an RNA-dependant DNA polymerase (i.e. AMV reverse transcriptase) is introduced into the reaction mixture to create a first primer extension product hybridized to the RNA by incubating the reaction mixture at 37°C for 1 hour. The RNA is then removed from the first primer extension product by cooling the reaction mixture to 37°C, adding 2 units of RNase H and RNase I to the reaction mixture and incubating for 20 minutes at 37°C. The reaction mixture is then heated to 90°C for 3 minutes to inactivate the enzymes. The first primer extension product can be mixed with 10ul of the second primer extension mixture. The second strand polymerase can be one that does not have strand displacement, such as for example T4 DNA polymerase. Conditions as described in Examples 1 and 2 above may be applicable. 20pmol of a second primer comprising a 5' segment and a 3' segment, a portion of the 5' segment comprising RNA, and a portion of the 3' segment comprising DNA. A

IDT Exhibit 1020.3472

portion of the 5' RNA segment comprises a tag sequence (B), and a portion of the 3' DNA segment comprises an annealing sequence comprising a random sequence.

[00254] The reaction mixture can be incubated for 30 minutes at 37°C, followed by addition of exonuclease 1 (0.5 µl at 37°C for 30 min.), and finally heating to 75°C for 5 minutes to stop the reaction by inactivating the enzymes. The products are purified and subjected to SPIA amplification. The first and second primer extension products are purified using Agencourt magnetic beads as per the manufacturer's instructions.

[00255] The first primer extension product in the partial duplex is extended by the DNA polymerase in the amplification reaction mixture comprising an DNA dependent DNA polymerase and a DNA dependent DNA polymerase,and the resulting primer extension products comprise a double stranded DNA product with a DNA/RNA heteroduplex at one end of a sequence (B')/(B). The first and second primer extension products are then purified using Agencourt magnetic beads as per the manufacturer's instructions.

[00256] Isothermal linear amplification (SPIA) is carried out in a reaction mixture containing the above purified reaction products (10 µl), 2 µl chimeric amplification primer (100 µM stock solution), 18 µl water, 40 µl amplification buffer and 20 µl amplification enzyme mixture as provided in the WT-Ovation Pico RNA Amplification System (NuGen Technologies). The amplification is carried out according to the instructions provided for the WT-Ovation Pico RNA Amplification System.

[00257] The amplified product is then analyzed and quantitated by Real Time qPCR with SYBR Green, using an MJ Opticon thermocycler. Amplification reactions are diluted 1:100 in Tris-EDTA and 2 µl of the diluted DNA are analyzed using primer pairs specific for abl, ras, and her2 in three separate reactions.

[00258] The results of the qPCR are analyzed to determine that abl is overexpressed in the cells of the suspicious mass. The results are combined with immunohistochemical and cytological analysis to determine a suggested course of therapy.

Example 4: Sequence specific amplification of mammalian genomic DNA

All DNA tailed first primer

[00259] An individual is tested by a personal genomics business using the methods of the present invention for single nucleotide polymorphisms (SNPs) within the BRCA1, BRCA2, p53, MPO, NAT1, NAT2, and ras coding regions that are related to increased risks for specific types of cancer.

[00260] The individual supplies a small sample of tissue (i.e. a cheek swab) to the personal genomics business. Genomic DNA from the sample of tissue is isolated using a commercially available kit (i.e. Promega's Wizard ® Genomic DNA Purification Kit), according to the manufacturer's protocol.

[00261] 1ng of purified genomic DNA is used to specifically amplify the sequences corresponding to the genomic regions with known, cancer related, SNPs of the BRCA1, BRCA2, p53, MPO, NAT1, NAT2, and ras genes from the complex mixture of sequences present in the genomic DNA in separate reactions. Double stranded DNAs, each comprising a sequence of interest, an (A)/(A') duplex at one end, and a DNA-RNA heteroduplex of sequence (B')/(B) at the other end are created using the steps outlined in Figures 1 and 2.

[00262] The double stranded DNAs each comprising a sequence of interest are amplified by isothermal linear amplification using the methods of the present invention to generate single stranded amplified product comprising the complement of the sequences of interest and a sequence (A') at the 3' end. The amplified product is then sequenced by dye terminator sequencing using a primer complementary to the sequence (A').

IDT Exhibit 1020.3473

[00263] The resulting sequences are used to determine the presence or absence of SNPs related to cancer in the genes of interest. A report is generated that includes the SNPs identified, the impact of the SNPs on lifetime risk of developing specific diseases or conditions, and suggestions for prophylactic or therapeutic interventions.

Example 5: Sequence specific multiplex amplification of fetal mammalian genomic DNA

All DNA tailed first primer

[00264] A fetal sample is obtained by amniocentesis. DNA is extracted and purified from the sample using a commercially available kit (i.e. Promega's Wizard ® Genomic DNA Purification Kit). 1ng of the DNA is amplified in a multiplex fashion according to the method outlined in Figures 1 and 2. Twenty three sequence specific all DNA first primers are used, each complementary to a region of one the following genes:

ACADM which is present on chromosome 1;
ABCA12 which is present on chromosome 2;
ALAS1 which is present on chromosome 3;
ANK2 which is present on chromosome 4;
ADAMTS2 which is present on chromosome 5;
BCKDHB which is present on chromosome 6;
ASL which is present on chromosome 7;
FGFR1 which is present on chromosome 8;
ABO which is present on chromosome 9;
CDH23 which is present on chromosome 10;
ACAT1 which is present on chromosome 11;
ACVRL1 which is present on chromosome 12;
ATP7B which is present on chromosome 13;
COCH which is present on chromosome 14;
CAPN3 which is present on chromosome 15;
ABCC6 which is present on chromosome 16;
ACADVL which is present on chromosome 17;
FECH which is present on chromosome 18;
APOE which is present on chromosome 19;
ADA1 which is present on chromosome 20;
APP which is present on chromosome 21;
IGL which is present on chromosome 22;
pyruvate dehydrogenase E1 alpha which is present on chromosome X; and
SRY which is present on chromosome Y;

[00265] The amplified products are then analyzed and quantitated by Real Time qPCR with SYBR Green, using an MJ Opticon thermocycler. Amplification reactions are diluted 1:100 in Tris-EDTA and 2 µl of the diluted DNA are analyzed using primer pairs specific for the twenty three different genes in twenty-three separate reactions.

[00266] The results of the qPCR are analyzed to determine whether the fetus exhibits an aneuploid genotype.

[00267] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

IDT Exhibit 1020.3474

WHAT IS CLAIMED IS:

1. A method for creating a partial RNA-DNA heteroduplex comprising:

a) providing a template nucleic acid in a reaction mixture;

b) adding an RNA-DNA chimeric primer;

c) adding a DNA polymerase; and

d) adding an exonuclease;

wherein said DNA polymerase comprises RNA-dependent polymerase activity.

2. A kit comprising a container containing reagents for forming a water in oil emulsion and a DNA polymerase with substantial strand-displacement activity; the kit optionally further comprising one or more all an exonuclease, an RNA-DNA chimeric primer, DNA primers, an RNA-DNA chimeric oligonucleotide, RNase H, and instructions for use of said reagents.

3. A kit comprising an all DNA first primer, an RNA-DNA chimeric oligonucleotide; the kit optionally further comprising a second primer, a polymerase having substantial strand-displacement activity, and RNase H.

4. A method for producing amplified nucleic acid comprising:

a) providing a template nucleic acid;

b) annealing a first primer or set of first DNA primers comprising a 3' end whereby a portion of the 3' end comprises a template annealing sequence and a 5' tail sequence whereby a portion of the tail sequence comprises a sequence (A);

c) extending said first primer or set of first primers with a DNA polymerase to create a first primer extension product or first set of DNA polymerase products;

d) dissociating the primer extension product or products of said DNA polymerase from the template nucleic acid;

e) annealing a second primer comprising a 3' DNA sequence and a 5' RNA sequence, whereby a portion of the 3' end comprises a randomized annealing sequence and a portion of the 5' end comprises sequence (A) or sequence (B), to said primer extension product or products; and

f) extending said second primer with a DNA polymerase to create a second primer extension product or products to produce a double stranded DNA product or products comprising an A-A' DNA duplex at one end.

5. The method of claim 4 wherein the template nucleic acid is selected from the group consisting of DNA and RNA.

6. The method of claim 4 wherein the template nucleic acid is RNA and the DNA polymerase of step (c) is an RNA dependant DNA polymerase.

7. The method of claim 6 wherein the dissociation step of step (d) is performed by degrading the RNA.

8. The method of claim 4 wherein the template nucleic acid is DNA and the DNA polymerase of step (c) is a DNA dependant DNA polymerase.

9. The method of claim 8 wherein the dissociation step of step (d) is performed by heat or chemical denaturation.

IDT Exhibit 1020.3475

10. The method of claim 4 wherein the DNA polymerase does not exhibit substantial strand displacement activity

11. The method of claim 4 wherein each template annealing sequence of the set of primers is specific for a target or region of template nucleic acid.

12. The method of claim 4 wherein the template annealing sequence of the first primer comprises a random sequence.

13. The method of claim 4 wherein the members of the set of first primers each comprises a distinct 3' DNA annealing sequence, each specific for a target or region of template nucleic acid.

14. The method of claim 4 wherein the first primer tail sequence is not complementary to the template nucleic acid.

15. The method of claim 4 wherein the 5' tail sequence of the second primer comprises sequence (A).

16. The method of claim 4 wherein the 5' tail sequence of the second primer does not comprise tail sequence (A).

17. The method of claim 4 further comprising step (g), degradation of single stranded 3' end of said first primer extension product or products in the complex of first and second primer extension products with a single stranded DNA specific 3' exonuclease.

18. The method of claim 17 further comprising step (h), the addition of a DNA polymerase with reverse transcriptase activity to create a double stranded nucleic acid with an A-A' DNA duplex at one end and an B-B' DNA-RNA heteroduplex at the other end.

19. The method of claim 18 further comprising SPIA with a reaction mixture comprising RNAse H, a composite amplification primer comprising a 3' DNA portion and a 5' end whereby a portion of the 5' end comprises RNA wherein the RNase H reacts with the heteroduplex of step (h) to produce a single stranded end, and the amplification primer hybridizes to the single stranded end produced by RNaseH, and a DNA polymerase with strand displacement activity.

20. The method of claim 4 wherein the first primer or set of primers comprises a ligand portion at the 5'end.

21. The method of claim 18 further comprising:

(i) cleavage of the RNA from the DNA-RNA heteroduplex with RNase H;

(j) annealing a chimeric oligonucleotide comprising a 5' end whereby the 5' end comprises sequence (C) and a 3' end whereby the 3' end comprises the DNA sequence (B) to the double stranded nucleic acid product;

(k) extending the double stranded nucleic acid product with a DNA polymerase to produce a C-C' DNA-RNA heteroduplex; and

IDT Exhibit 1020.3476

(l) amplifying the second primer extension product with a reaction mixture comprising RNase H, an amplification primer comprising a 3' DNA portion and a 5' end whereby a portion of the 5' end comprises RNA that hybridizes to a portion of the single stranded end produced by RNase H, and a DNA polymerase with strand displacement activity to produce an amplification product comprising sequence (B) near its 3' end and sequence (A') at the 5' end.

22. A method for producing amplified nucleic acid comprising:

a) providing a template nucleic acid;

b) annealing a first DNA primer or set of first DNA primers comprising a 3' end whereby a portion of the 3' end comprises a template annealing sequence and a 5' end whereby a portion of the 5' end comprises a DNA tail sequence (A);

c) extending said first primer or set of first primers with a DNA polymerase to create a primer extension product or set of primer extension products;

d) dissociation of the primer extenstion product or products from the template nucleic acid;

e) annealing a second primer comprising a 3' DNA end, whereby a portion of the 3' end comprises a random template annealing sequence, and a 5' RNA end, whereby a portion of the 5' end comprises the tail sequence (A), to said primer extension product or products or to said template nucleic acid;

f) extending said second primer with a DNA polymerase to create a second primer extension product or set of second primer extension products; and

g) denaturing and annealing the second primer extension product or set of second primer extension products to form a stem loop structure comprising an A-A' RNA-DNA heteroduplex.

23. The method of claim 22 further comprising the addition of a 3' single stranded DNA specific exonuclease and single strand specific RNase to degrade unincorporated primers and template nucleic acid.

24. The method of claim 23 further comprising amplification with a reaction mixture comprising RNase H, an amplification primer comprising sequence (A'), and a DNA polymerase with strand displacement activity, wherein the amplification primer comprises a 3'-DNA sequence and a 5'-RNA sequence.

25. The method of claim 22 wherein said template nucleic acid is selected from the group consisting of RNA and DNA.

26. The method of claim 22 wherein the template nucleic acid is RNA and the DNA polymerase of step (c) is an RNA dependant DNA polymerase.

27. The method of claim 22 wherein the template nucleic acid is DNA and the DNA polymerase of step (c) is a DNA dependant DNA polymerase.

28. The method of claim 26 wherein the dissociation step of step (d) is performed by degradation of the RNA.

29. The method of claim 27 wherein the dissociation step of step (d) is performed by chemical or heat denaturation.

IDT Exhibit 1020.3477

30. The method of claim 22 wherein the DNA polymerase does not exhibit substantial strand displacement activity.

31. The method of claim 22 wherein each DNA annealing sequence in the set of first DNA primers is specific for a target or region of the template nucleic acid.

32. The method of claim 22 wherein a portion of the 5' tail sequence of the first primer is sequence (A) and a portion of the 5' tail sequence of the second primer is sequence (A).

33. The method of claim 22 wherein the tail sequence is not complimentary to the template nucleic acid.

34. A method of capturing and amplifying a target sequence or sequences comprising:

(a) providing a surface wherein a first primer or set of first primers comprising a 5' tail sequence comprising sequence (A) and a 3' template annealing sequence comprising sequence (P) are immobilized on said surface;

(b) annealing a target nucleic acid to said first primer or set of primers;

(c) extending the first primer or set of primers with a DNA polymerase to produce an immobilized primer extension product or set of immobilized primer extension products complimentary to and hybridized with the target nucleic acid to form a complex or set of complexes;

(d) dissociating the target nucleic acid-primer extension product complex or set of primer extension product complexes;

(e) annealing a second primer comprising a 3' DNA template annealing sequence (P) or set of template annealing sequences and a 5' RNA tail sequence comprising sequence (A) or sequence (B) to the immobilized DNA polymerase product or DNA polymerase products; and

(f) extending said second primer with a DNA polymerase to create a second primer extension product or set of primer extension products comprising a double stranded nucleic acid comprising a 3' DNA-DNA duplex A-A'.

35. The method of claim 34 wherein said template nucleic acid is selected from the group consisting of RNA and DNA.

36. The method of claim 34 wherein said first primer comprises a spacer element between the immobilizing surface and the remaining portion of the primer, a 3' DNA end comprising a 3' DNA annealing sequence (P), and a 5' comprising a common 5' DNA sequence tail (A) or sequence tail (B) between said spacer element and annealing sequence.

37. The method of claim 34 wherein the set of first primers comprises:

(a) a spacer element between the immobilizing surface and the remaining portion of the primer;

(b) a 3' DNA end wherein members of the set of first primers comprise unique target specific 3' DNA annealing sequences (P); and

(c) a 5' DNA end comprising a common 5' DNA sequence tail (A) or DNA sequence tail (B) between said spacer element and annealing sequence.

IDT Exhibit 1020.3478

38. The method of claim 34 wherein said annealing sequences (P) of said set of first primers are each specific to a target or region of the template nucleic acid.

39. The method of claim 34 wherein said set of first primers is specific to one or more of the following gene targets: kinases, protein kinases, lipid kinases, phosphatases, G-protein coupled receptors, proteases, serine proteases, metalloproteases, cysteine proteases, aspartyl proteases, or cytokines.

40. The method of claim 34 wherein the 5' tail sequence of the first primer is sequence (A) and the 5' tail sequence of the second primer is sequence (B).

41. The method of claim 34 wherein the 5' tail sequence of the first primer is sequence (A) and the 5' tail sequence of the second primer is sequence (A).

42. The method of claim 34 wherein said DNA polymerase does not exhibit strand displacement activity.

43. The method of claim 34 further comprising step (g) elution of the second primer extension product by heat or chemical denaturation.

44. The method of claim 34 wherein the surface is selected from the group consisting of a bead, a magnetic particle, a microarray, a gene chip, and an array.

45. The method of amplifying the eluted DNA polymerase product of claim 43 further comprising:

(h) annealing an amplification primer comprising a 5' RNA sequence and a 3'DNA sequence;

(i) extending said annealed primer with a DNA polymerase; and

(j) amplifying the eluted DNA polymerase product with a reaction mixture comprising RNAse H, an amplification primer comprising a DNA portion and a 5' RNA portion, and a DNA polymerase with strand displacement activity to produce an amplified DNA product or products having a defined 5' sequence (A') or (B').

46. The method of claim 34 further comprising:

(g) degradation of the single stranded 3' end of said first primer extension product and unextended immobilized first primers with a single stranded DNA specific 3' exonuclease;

(h) cleavage of spacer element with light or chemical cleavage to release the immobilized double stranded nucleic acid from the surface; and

(i) amplifying the released double stranded nucleic acid in solution with a reaction mixture comprising RNase H, an amplification primer comprising a DNA portion and a 5' RNA portion, and a DNA polymerase with strand displacement activity to produce an amplified DNA product or products having a defined 3' sequence comprising sequence (A').

47. The method of claim 34 further comprising:

(g) degrading the single stranded 3' end of said first primer extension product or products and unextended immobilized first primers with a single stranded DNA specific 3' exonuclease;

(h) extending said first primer extension product with a DNA polymerase; and

IDT Exhibit 1020.3479

(i) degrading RNA in the heteroduplex from the first primer extension product and the RNA portion of the second primer with RNAse H.

(j) annealing an amplification primer to the single stranded portion of the second primer extension product wherein the amplification primer has a DNA portion and a 5' RNA portion;

(k) extending the amplification primer with a DNA polymerase having strand displacement activity to produce an amplified product hybridized to the second primer extension product; and

(l) repeating steps (i)-(k) to produce multiple copies of amplified product or products having a defined 3' sequence comprising sequence (A').

48. The method of claim 41 further comprising:

(g) separating the second primer extension product or products from the first primer extension product or products by heat or chemical denaturation;

(h) annealing the second primer extension product in solution to create a stem loop structure with a DNA sequence (A')-RNA sequence (A) heteroduplex end;

(i) degrading RNA in the heteroduplex with RNase H;

(j) annealing an amplification primer to the single stranded portion of the second primer extension product wherein the amplification primer has a DNA portion and a 5' RNA portion, wherein the RNA portion of the amplification primer forms a heteroduplex with a DNA portion of the second primer extension product;

(k) extending the amplification primer with a DNA polymerase having strand displacement activity to produce an amplified product hybridized to the second primer extension product to form a double stranded nucleic acid, said amplified product forming an RNA-DNA heteroduplex at one end of the double stranded nucleic acid; and

(l) repeating steps (i)-(k) to produce amplified DNA.

49. The method of claim 48 wherein between step (f) and step (g), a single stranded DNA specific 3' exonuclease is added to degrade unincorporated primers.

50. A kit comprising:

(a) a first primer comprising a 3' DNA portion and a 5' DNA portion, wherein the 5' DNA portion further comprises sequence (A) and the 3' portion further comprises an annealing sequence (P);

(b) a second primer comprising a 3' DNA portion and a 5' RNA portion, wherein the 5' RNA portion further comprises sequence (A) or (B) and the 3' portion further comprises an annealing sequence (P);

(c) a DNA polymerase;

(d) a DNA dependent DNA polymerase with strand displacing activity;

(e) RNase H; and

(f) a chimeric amplification primer comprising a 3' DNA portion and a 5' RNA portion wherein the sequence of the amplification primer is substantially the same as the (A) sequence.

51. A kit comprising:

(a) a first primer comprising a 3' DNA portion and a 5' DNA portion, wherein the 5' DNA portion further comprises sequence (A) and the 3' DNA portion further comprises annealing sequence (P);

IDT Exhibit 1020.3480

(b) a second primer comprising a 3' DNA portion and a 5' RNA portion, wherein the 5' RNA portion further comprises sequence (A) or (B) and the 3' DNA portion further comprises annealing sequence (P);

(c) a DNA polymerase;

(d) a DNA dependent DNA polymerase with strand displacing activity;

(e) a chimeric oligonucleotide comprising a 3' DNA portion substantially comprising sequence (A) or (B) and a 5' RNA portion comprising sequence (C).

(f) RNase H; and

(g) a chimeric amplification primer comprising a 3' DNA portion and a 5' RNA portion wherein the sequence of the amplification primer is substantially the same as the (A) sequence.

52. A kit comprising:

(a) a first primer comprising a DNA portion wherein the DNA portion further comprises an annealing sequence (P);

(b) a second primer comprising a 3' DNA portion and a 5' RNA portion, wherein the 5' RNA portion further comprises sequence (A) or (B) and the 3' portion further comprises an annealing sequence (P);

(c) a DNA polymerase;

(d) a DNA dependent DNA polymerase with strand displacing activity;

(e) RNase H; and

(f) a chimeric amplification primer comprising a 3' DNA portion and a 5' RNA portion wherein the sequence of the amplification primer is substantially the same as the (A) sequence.

53. A kit comprising:

(a) an immobilized first primer comprising a 3' DNA portion and a 5' DNA portion, wherein the 5' DNA portion further comprises sequence (A), and the 3' DNA portion further comprises annealing sequence (P);

(b) a second primer comprising a 3' DNA portion and a 5' RNA portion, wherein the 5' RNA portion further comprises sequence (A) or (B);

(c) a DNA polymerase;

(d) a DNA dependent DNA polymerase with strand displacing activity;

(e) RNase H; and

(f) a chimeric amplification primer comprising a 3' DNA portion and a 5' RNA portion wherein the sequence of the amplification primer is substantially the same as the (A) sequence.

54. The kit of any one of claims 47 - 50 wherein the DNA polymerase of component (c) is an RNA dependant DNA polymerase.

55. The kit of any one of claims 47 - 50 wherein the DNA polymerase of component (c) is a DNA dependant DNA polymerase.

56. The kit of any one of claims 47 - 50 wherein the DNA polymerase of component (c) does or does not possess substantial strand displacing activity.

57. The kit of any one of claims 47 - 50 wherein the 3' DNA portion of the first primer comprises a random sequence.

IDT Exhibit 1020.3481

58. The kit of any one of claims 47 - 50 wherein the 3' DNA portion of the first primer comprises a specific sequence.

59. The kit of any one of claims 47 - 50 wherein the 3' DNA portion of the first primer comprises a set of specific sequences.

60. The kit of any one of claims 47 - 50 wherein the 3' DNA portion of the second primer comprises a random sequence.

61. The kit of any one of claims 47 - 50 wherein the 3' DNA portion of the second primer comprises a specific sequence.

62. The kit of any one of claims 47 - 50 wherein the 3' DNA portion of the second primer comprises a set of specific sequences.

63. The kit of any one of claims 47 - 50 wherein the first primer further comprises a ligand.

64. The kit of any one of claims 47 - 50 further comprising a single stranded DNA specific 3' exonuclease.

65. The kit of any one of claims 47 - 50 further comprising an RNase specific for single and double stranded RNA.

66. An amplification method comprising;

a) forming a polynucleotide strand comprising a sequence (A) at its 5' end and complementary sequence (A') at its 3' end wherein sequence (A) comprises RNA within a mixture comprising other polynucleotides that are single stranded;

b) treating said polynucleotide so as to intramolecularly anneal the (A) sequence to the (A') sequence;

c) purifying the polynucleotide by degrading the other polynucleotides by treating the mixture with one or more enzymes that degrade single stranded polynucleotides; and

d) performing SPIA amplification with the purified polynucleotide from step (c) using sequence (A') as a priming site for the SPIA amplification..

67. An amplification method comprising:

a) obtaining a polynucleotide strand comprising sequence (B') at its 3' end;

b) annealing a chimeric oligonucleotide comprising a 3' DNA portion and a 5' RNA portion, wherein the 5'portion comprises sequence (C), and the 3' portion comprises sequence (B);

c) extending the polynucleotide strand with a RNA dependent DNA polymerase to add a DNA sequence (C') to the polynucleotide strand; and

d) performing SPIA amplification using the product of step (c) as the SPIA substrate using sequence (C') as a priming site for SPIA amplification.

IDT Exhibit 1020.3482

Figure 1: Sequence specific amplification with multiplex capabilities
Tailed first and second primers with same or different sequence tail (A) or (B)

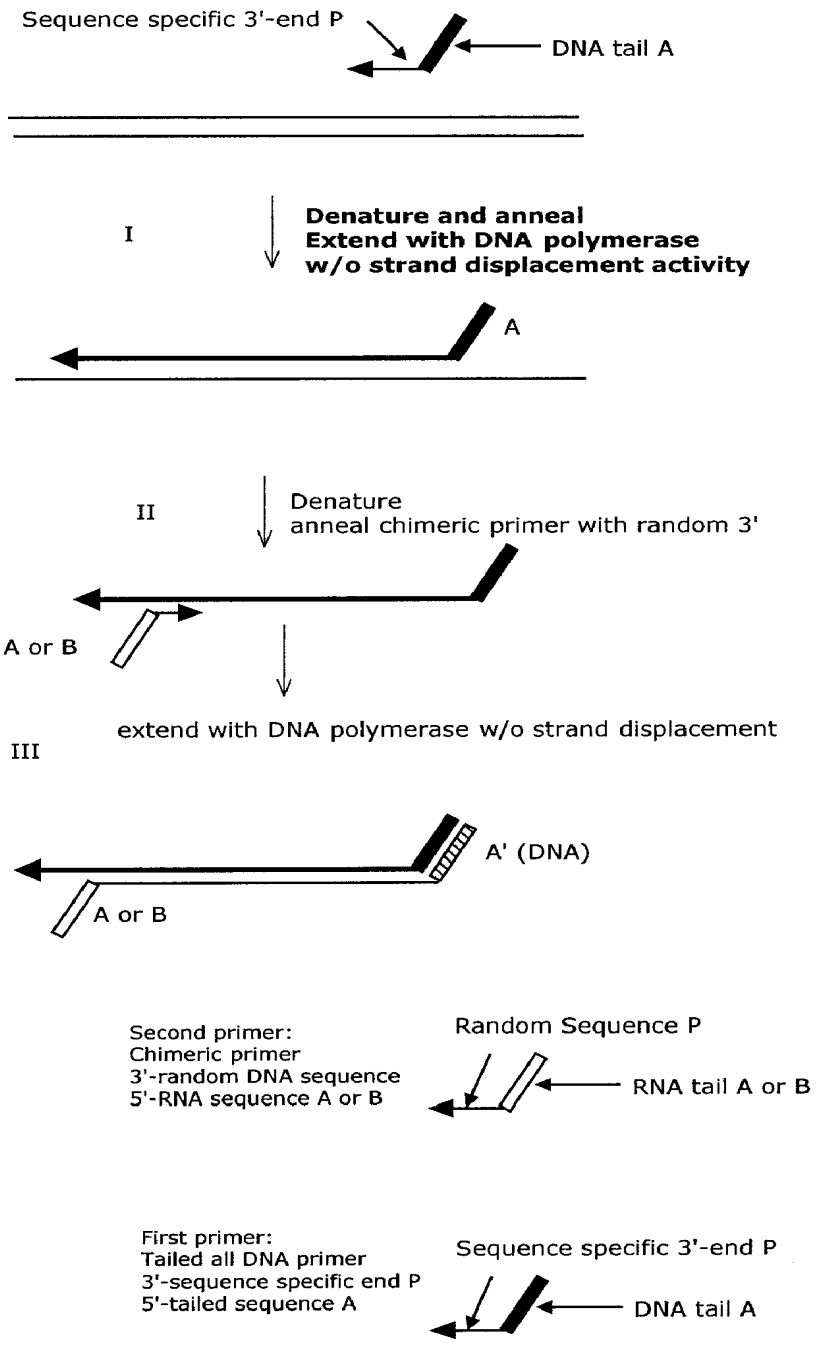


Figure 2: Formation of double stranded nucleic acid suitable for amplification by SPIA from primer extension products (described in Figure 1)

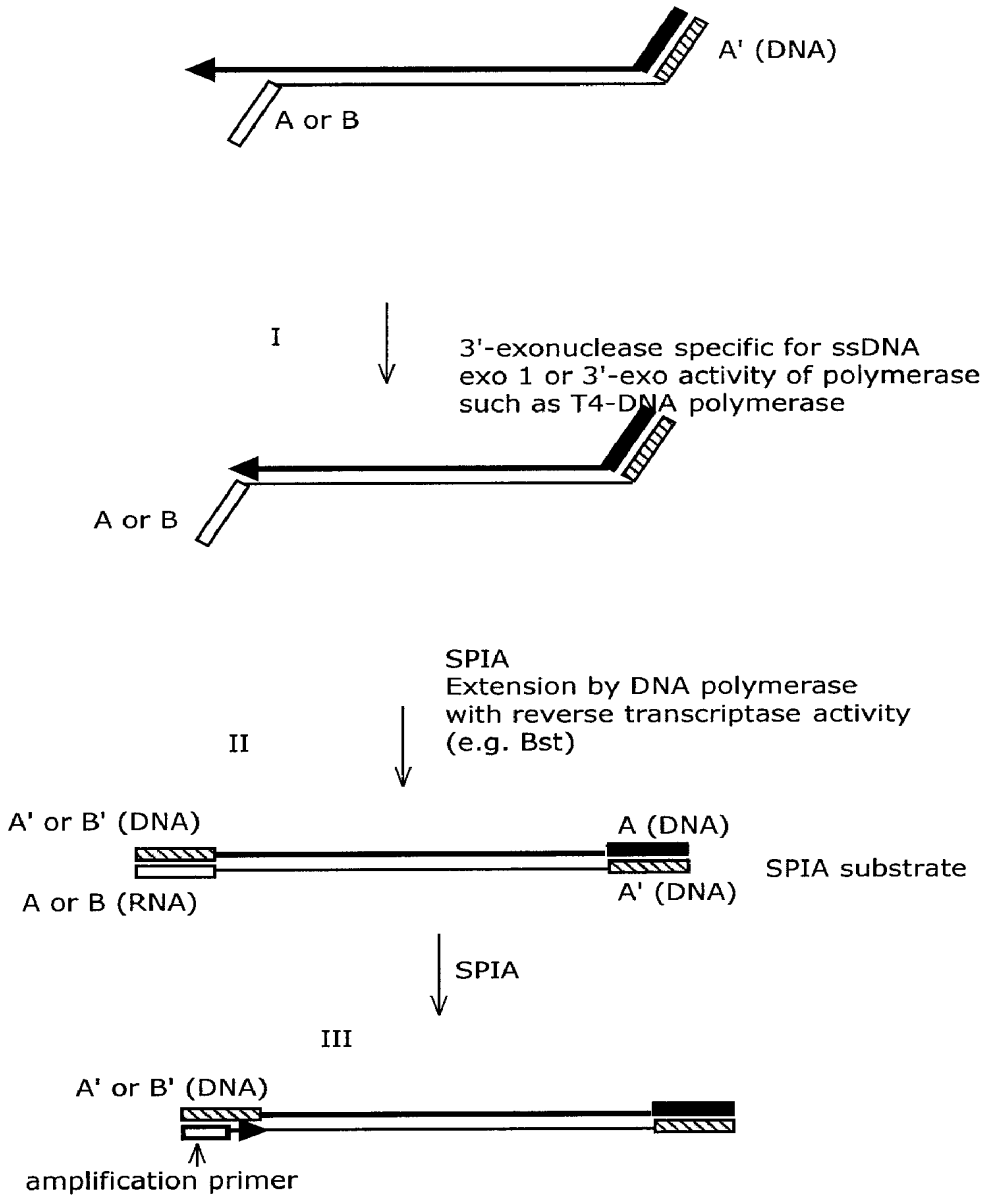


Figure 3: formation of stem-loop suitable for amplification by SPIA method from first and second primer extension products as described in Figure 1

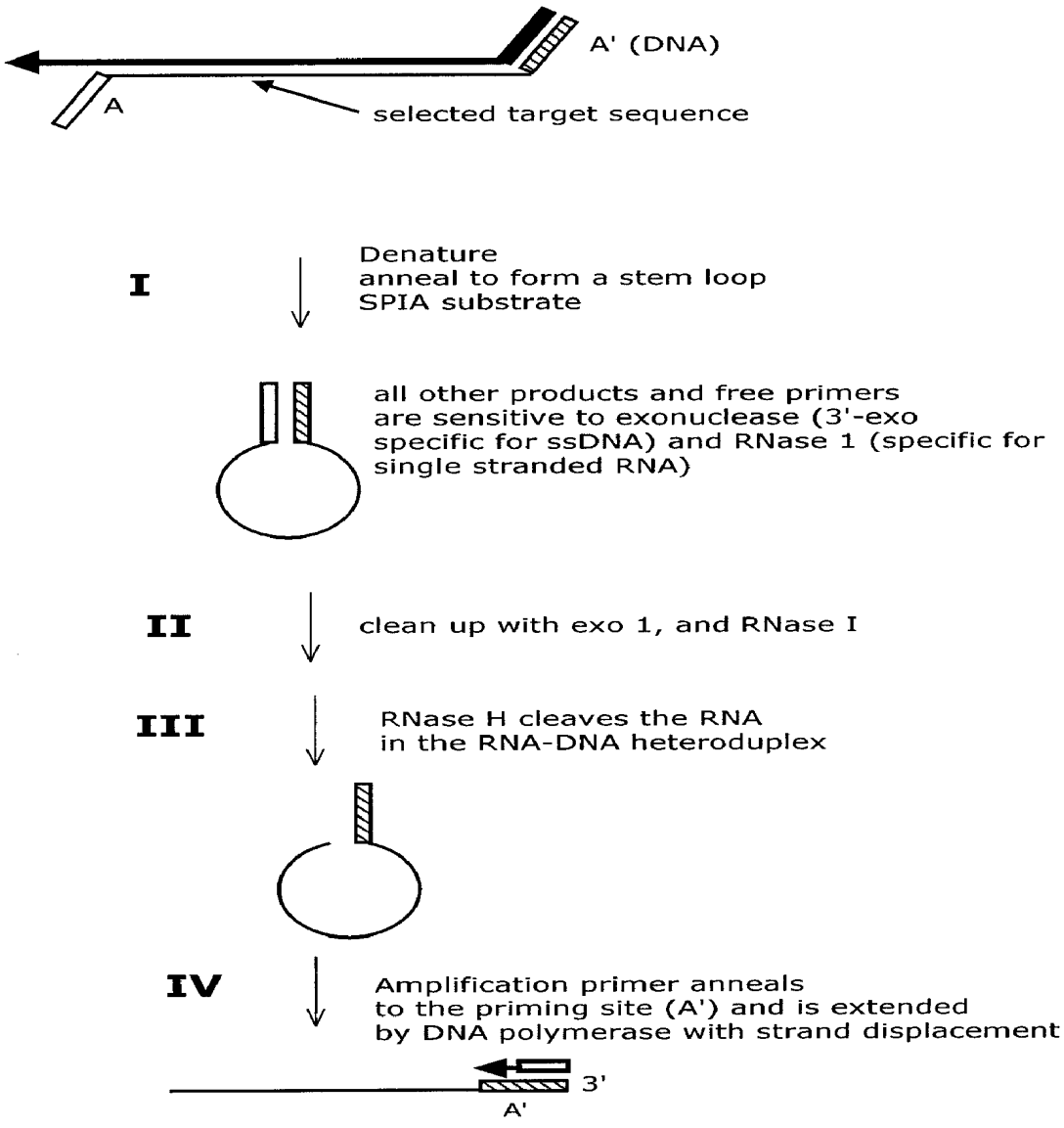


Figure 4
Capture of targeted sequence on immobilized primers
for amplification of specific captured sequences

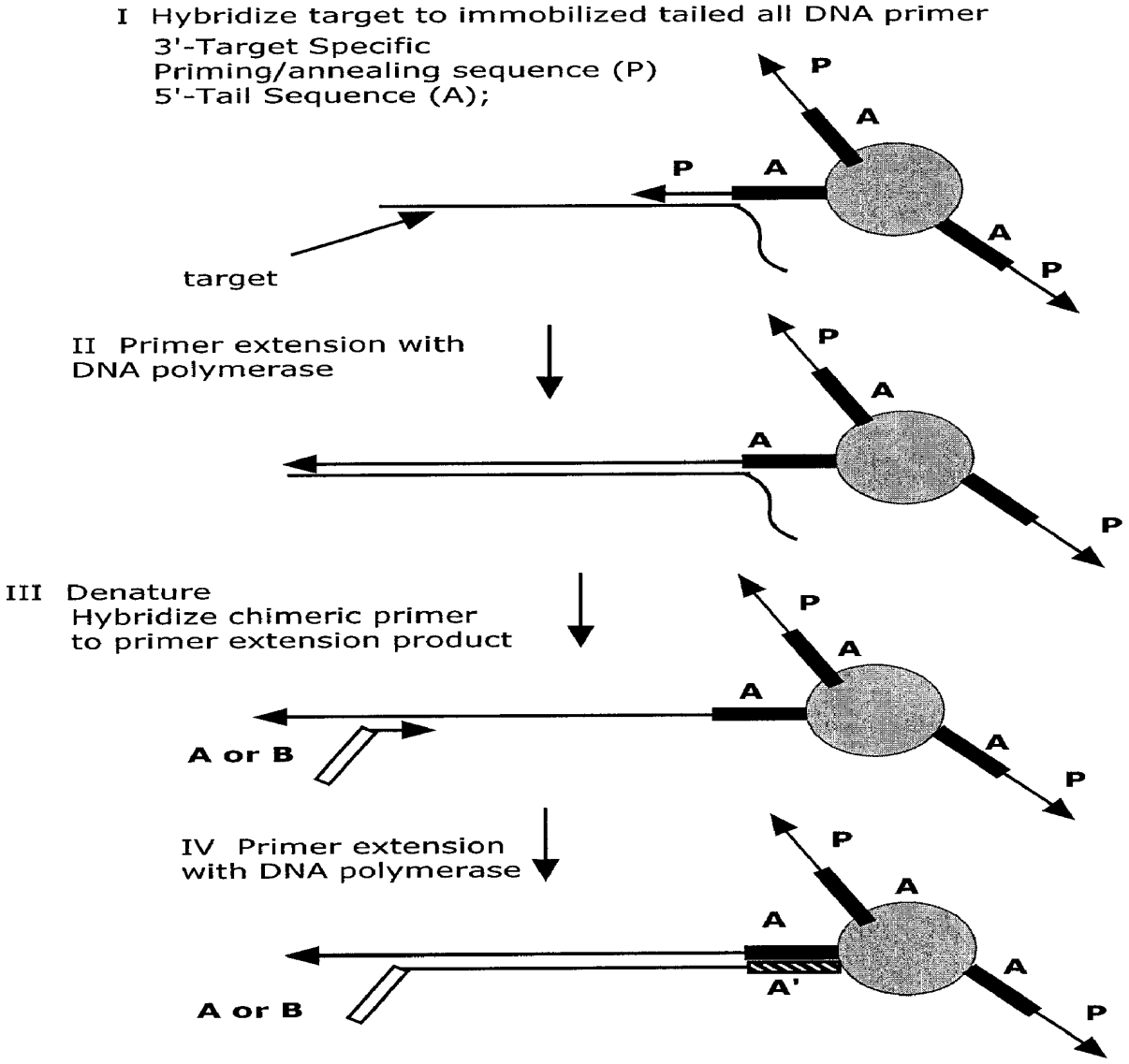


Figure 5 **Generation of amplified DNA from captured target sequences, as described in Figure 4, and further SPIA amplification**

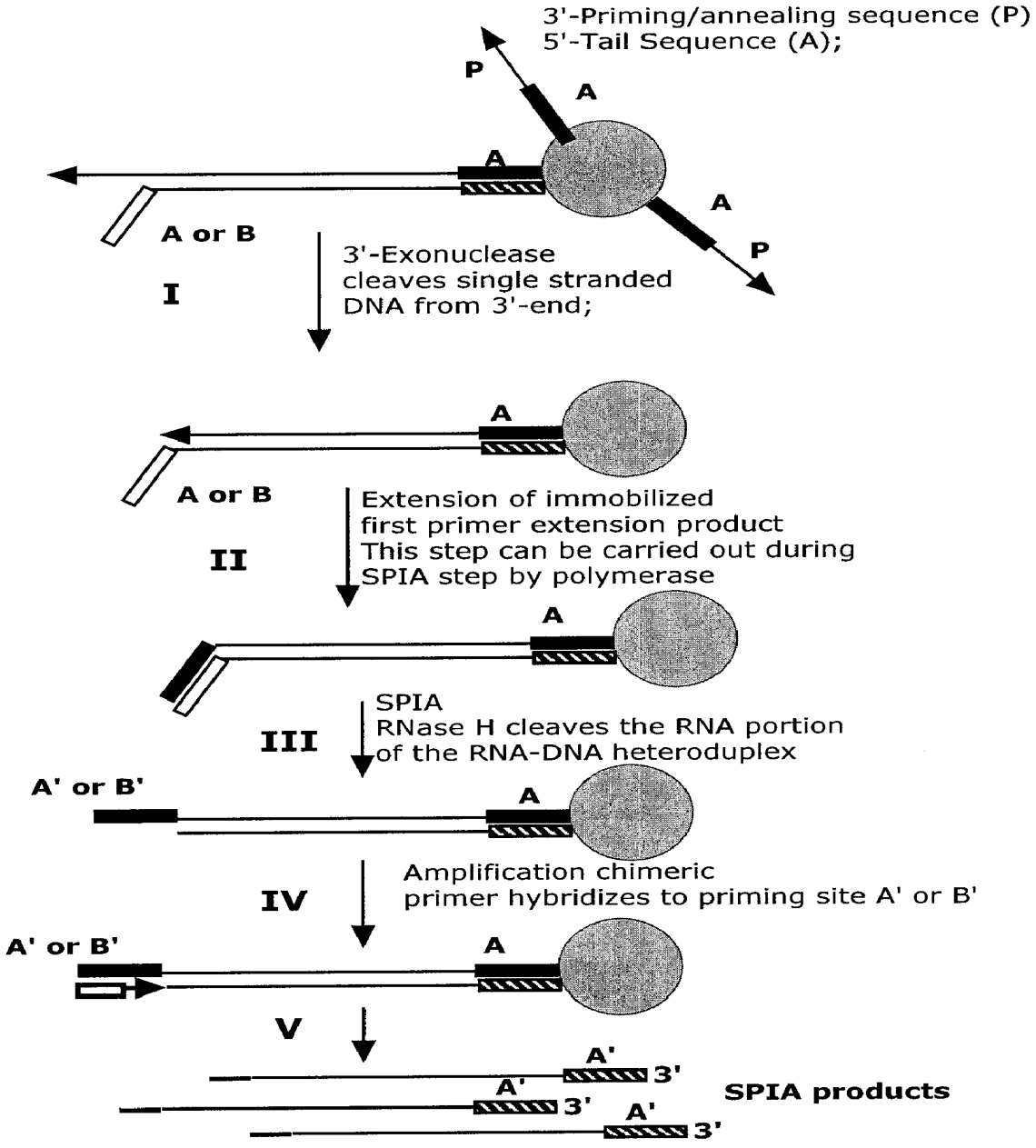


Figure 6: Generation of amplified DNA in solution from captured target sequences, as described in Figure 4 and further solution phase SPIA amplification

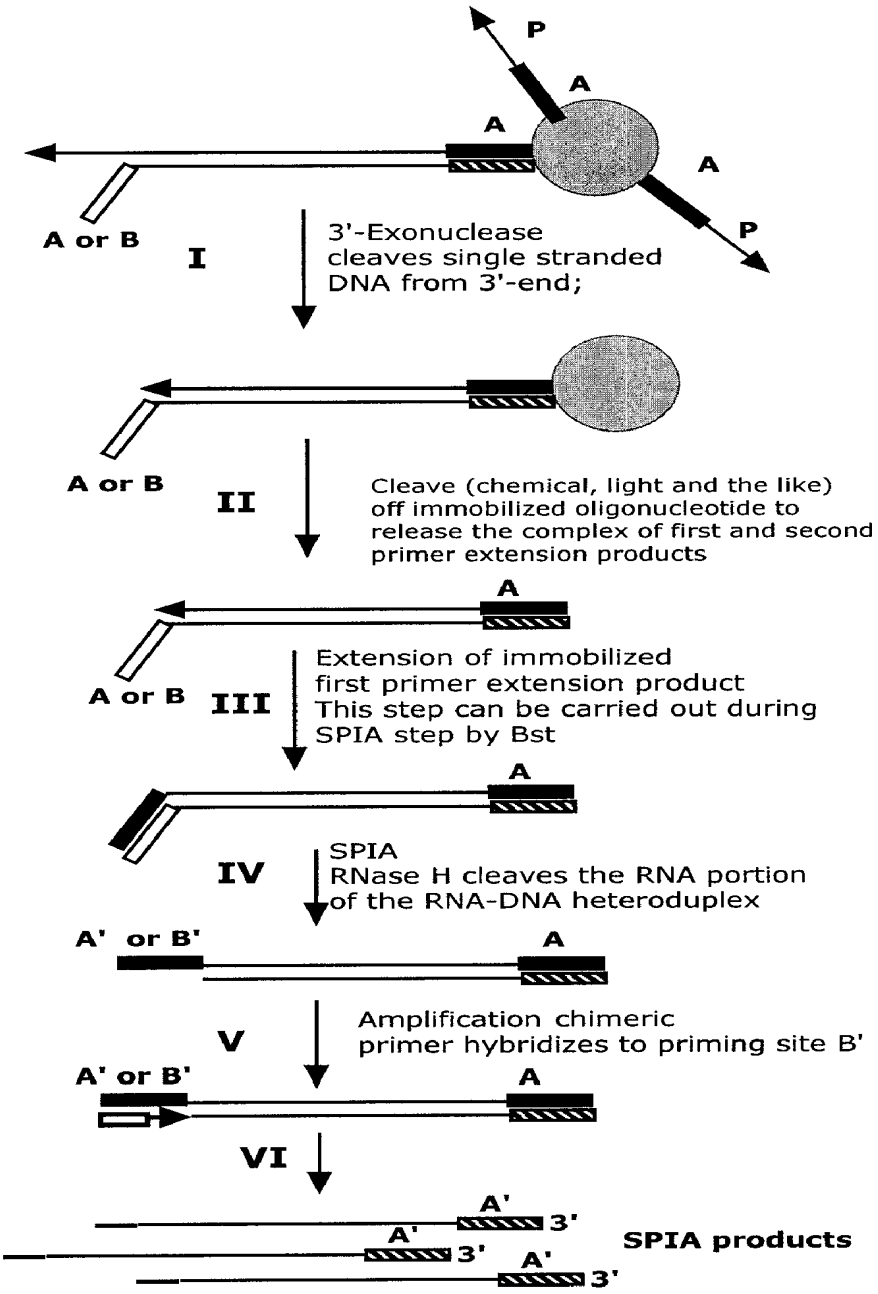


Figure 7
Capture of targeted sequence on immobilized primers for amplification of specific captured sequences similar to Figure 4 using chimeric primer with 5'-RNA sequence A, the same as DNA sequence A of the immobilized oligonucleotides

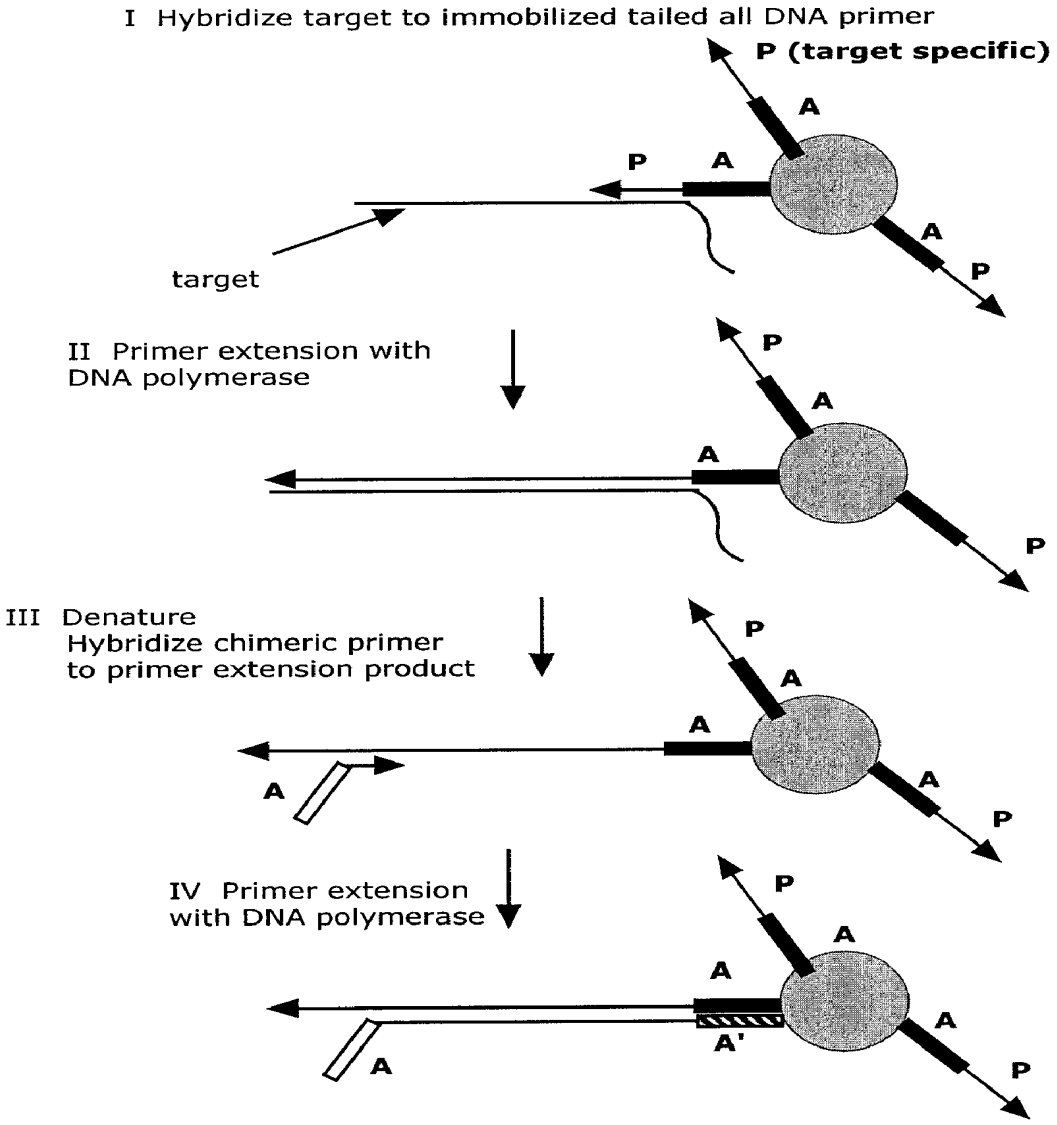


Figure 8: Generation of stem-loop amplification substrate from captured target sequences as described in Figure 7

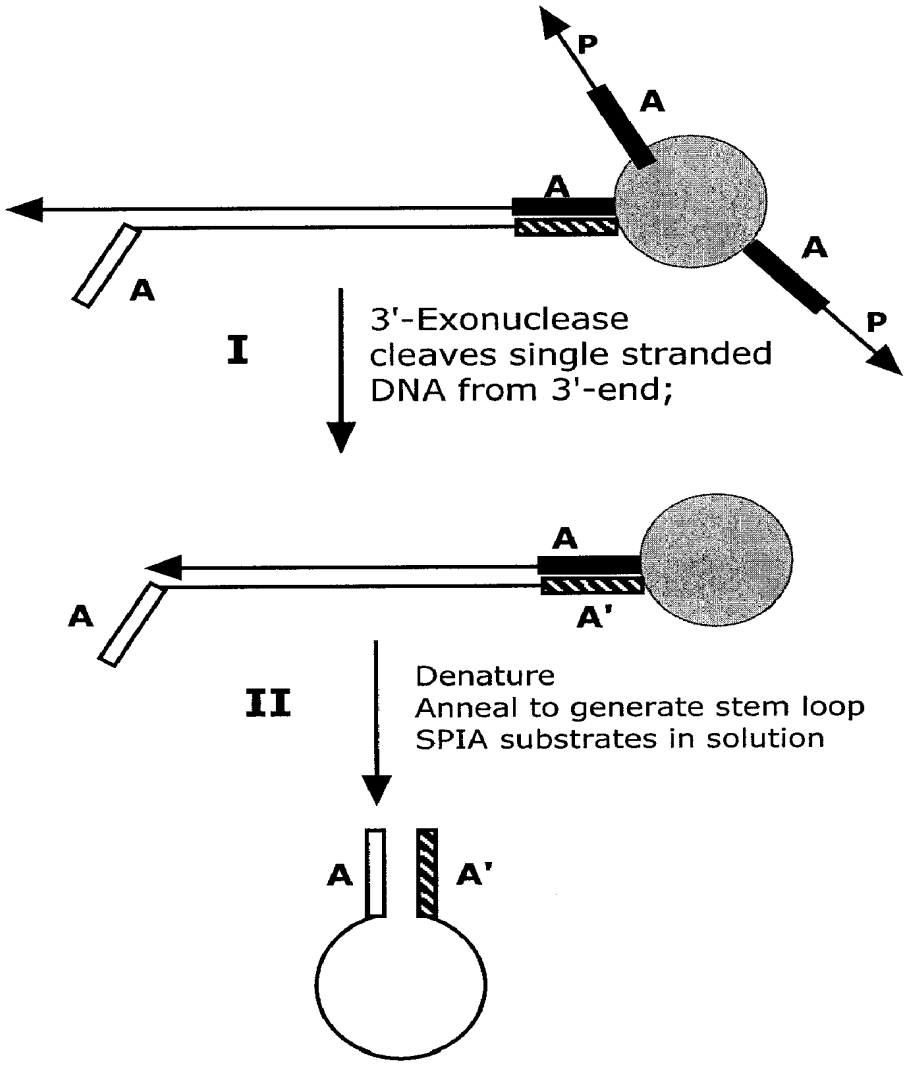


Figure 9
Amplification of RNA using a chimeric second primer

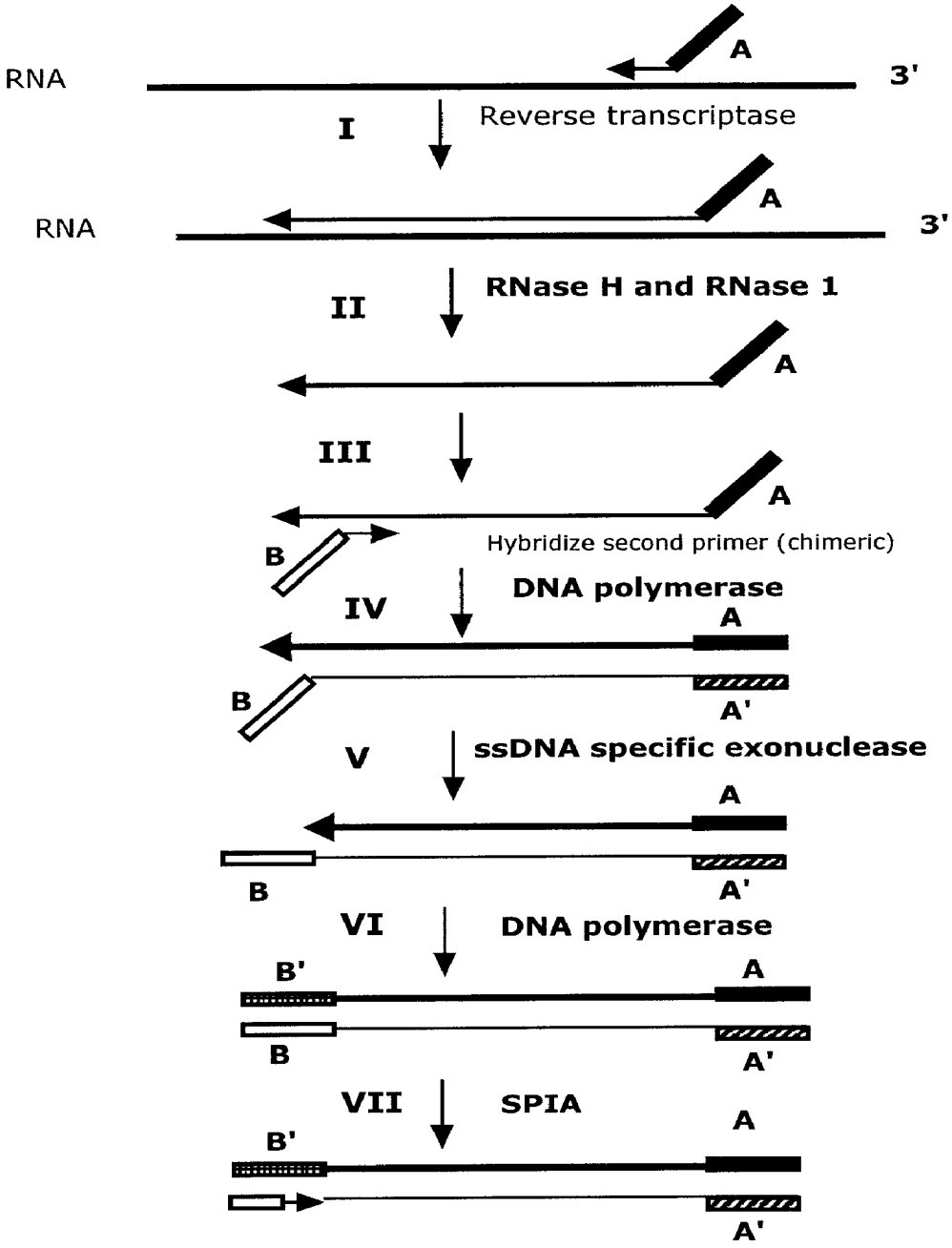


Figure 10: Expansion (from Figure 9) to generate DNA substrate suitable for generation of amplification products by SPIA comprising sequence A and B at the 3' and 5' ends.

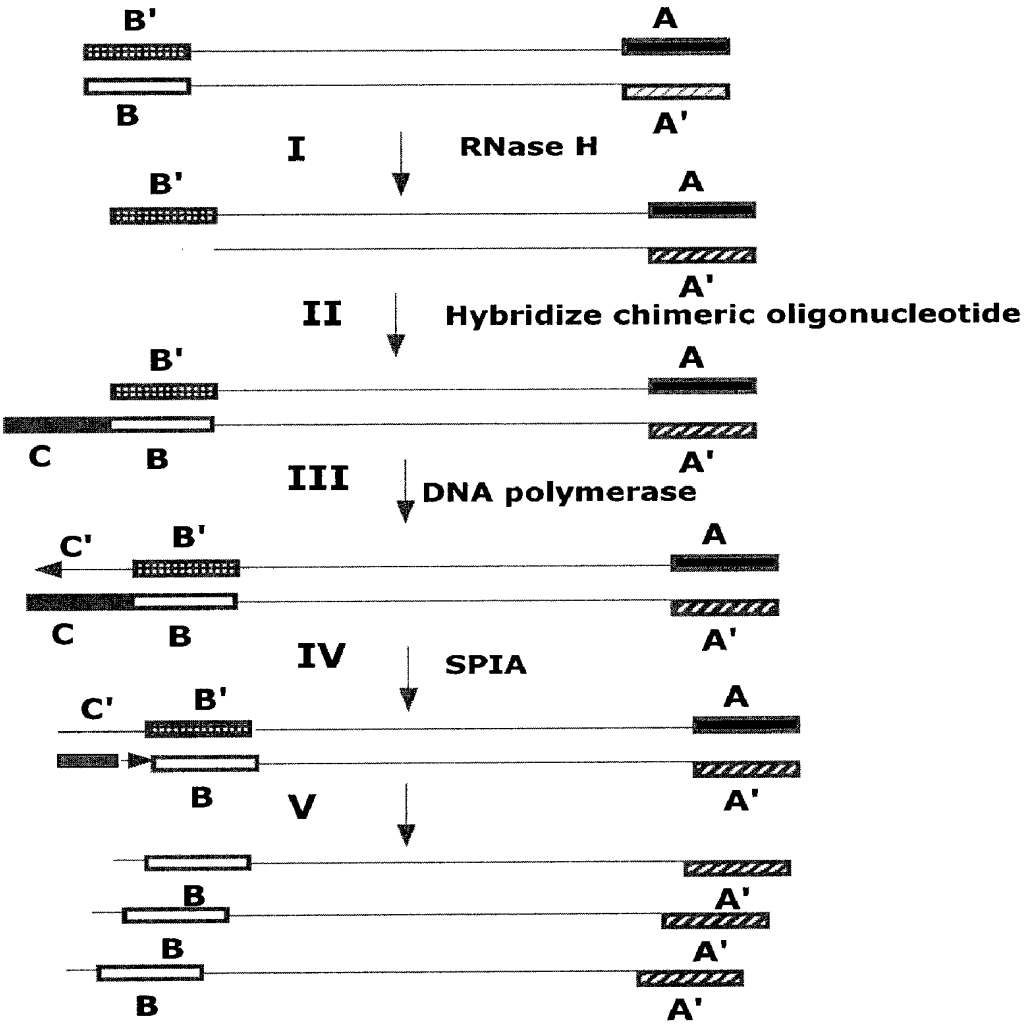


Figure 11 Method for generating a SPIA substrate with a ligand

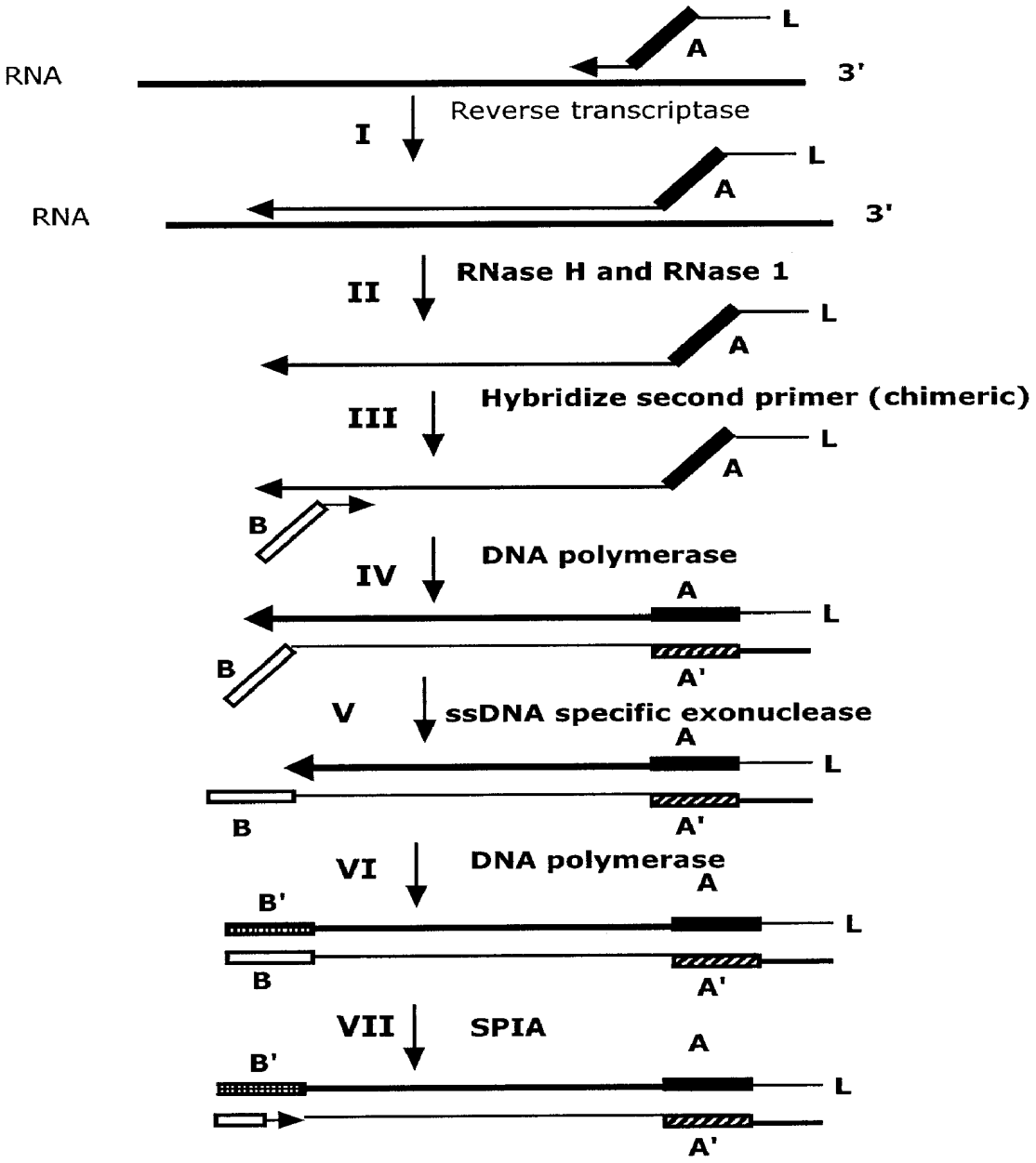
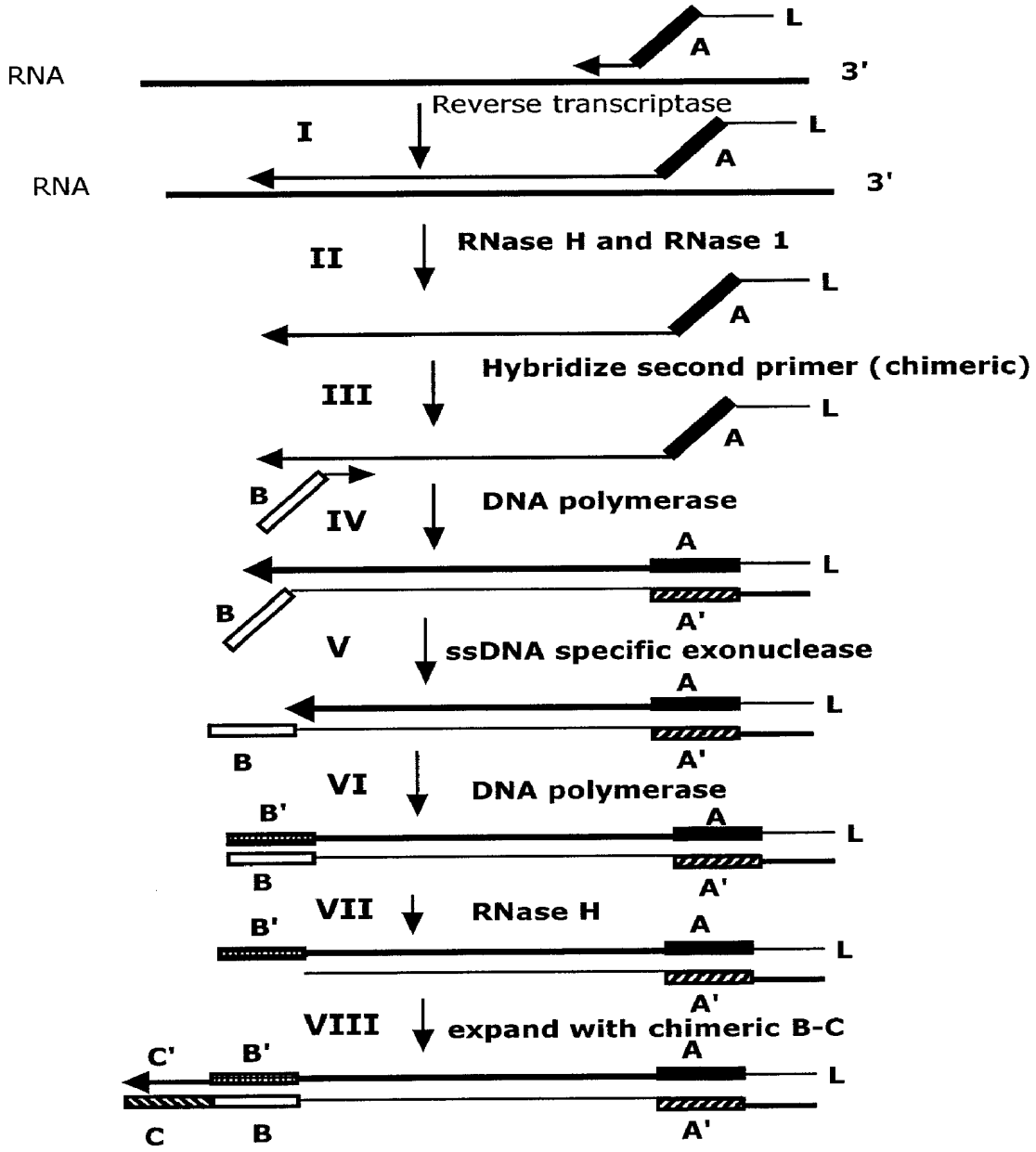


Figure 12 Method of producing a SPIA substrate with a ligand and two defined ends



I

Figure 13: DNA target; non tailed DNA first primer; tailed chimeric second primer

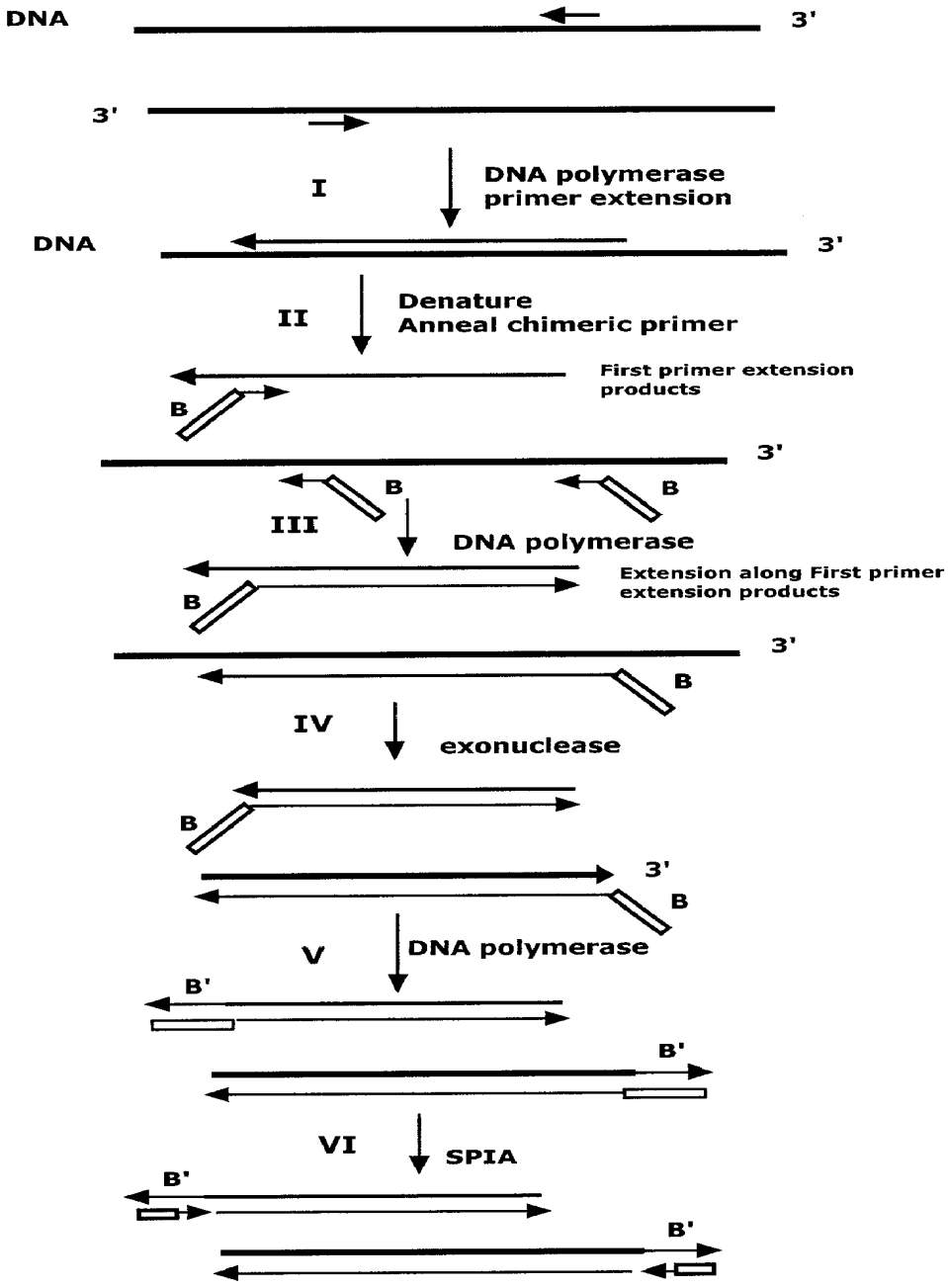


Figure 14: DNA target; non tailed, sequence specific DNA first primer; tailed chimeric second primer

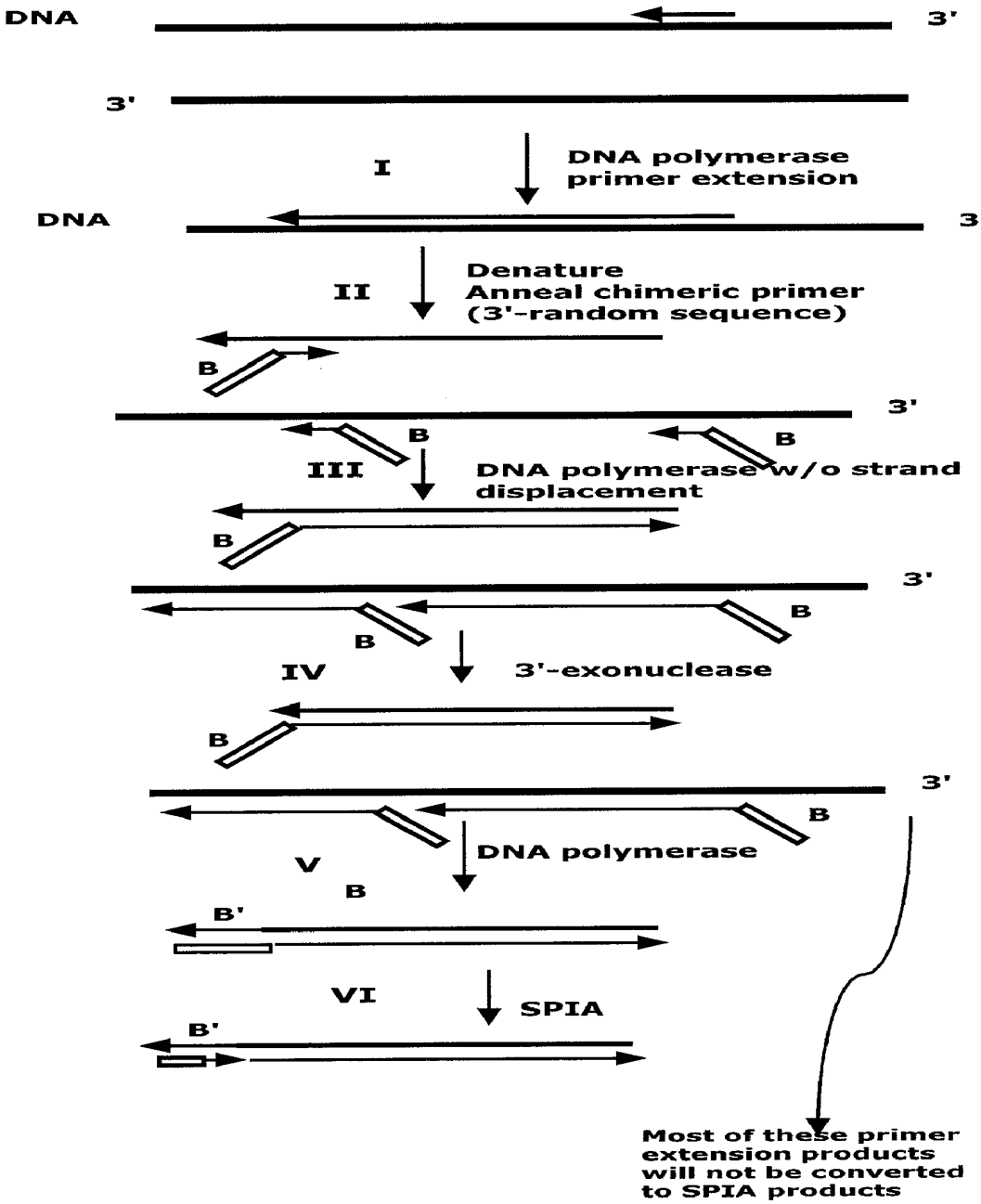


Figure 15 Amplification of a DNA target with a tailed DNA first primer with a ligand L and a chimeric tailed second primer

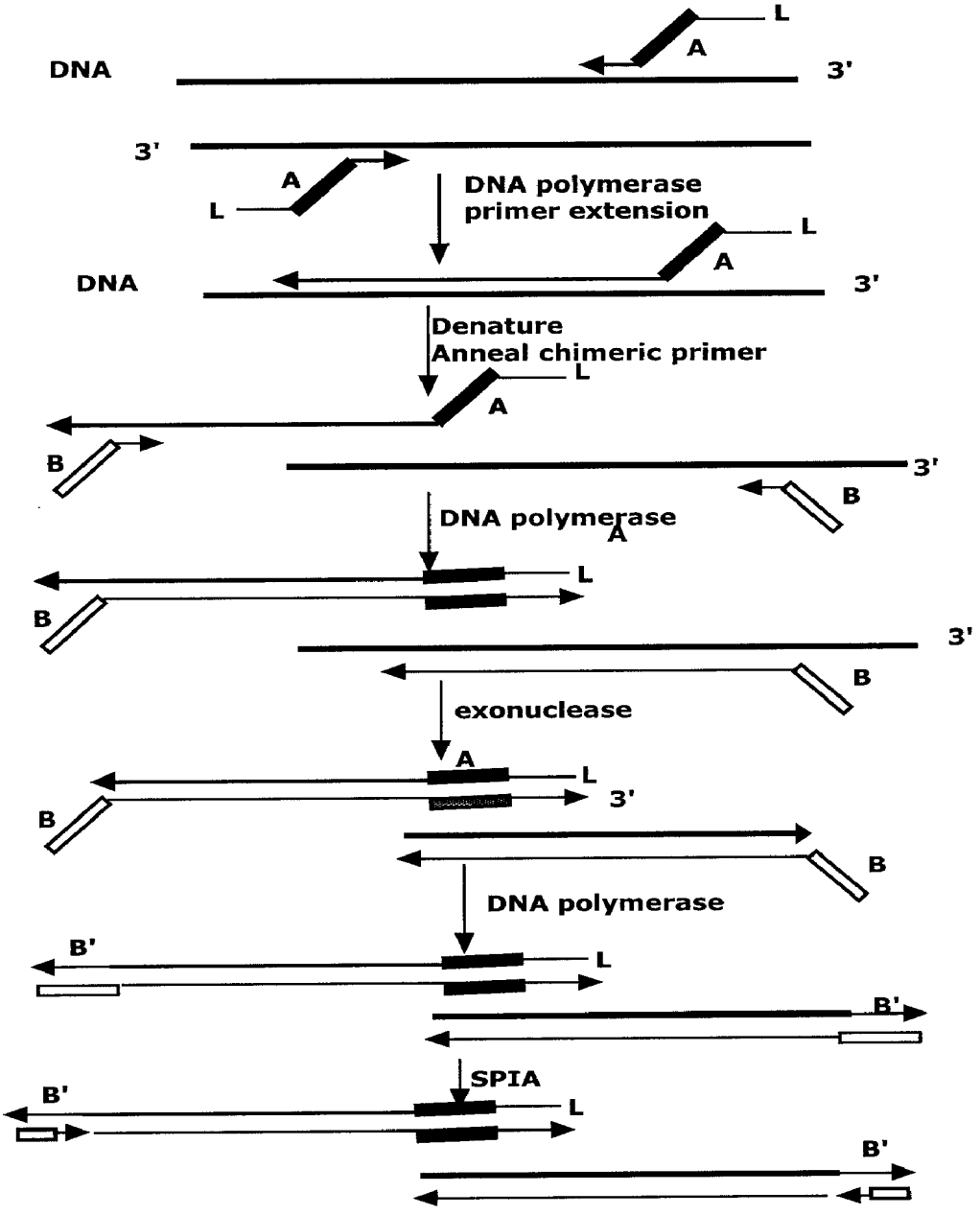
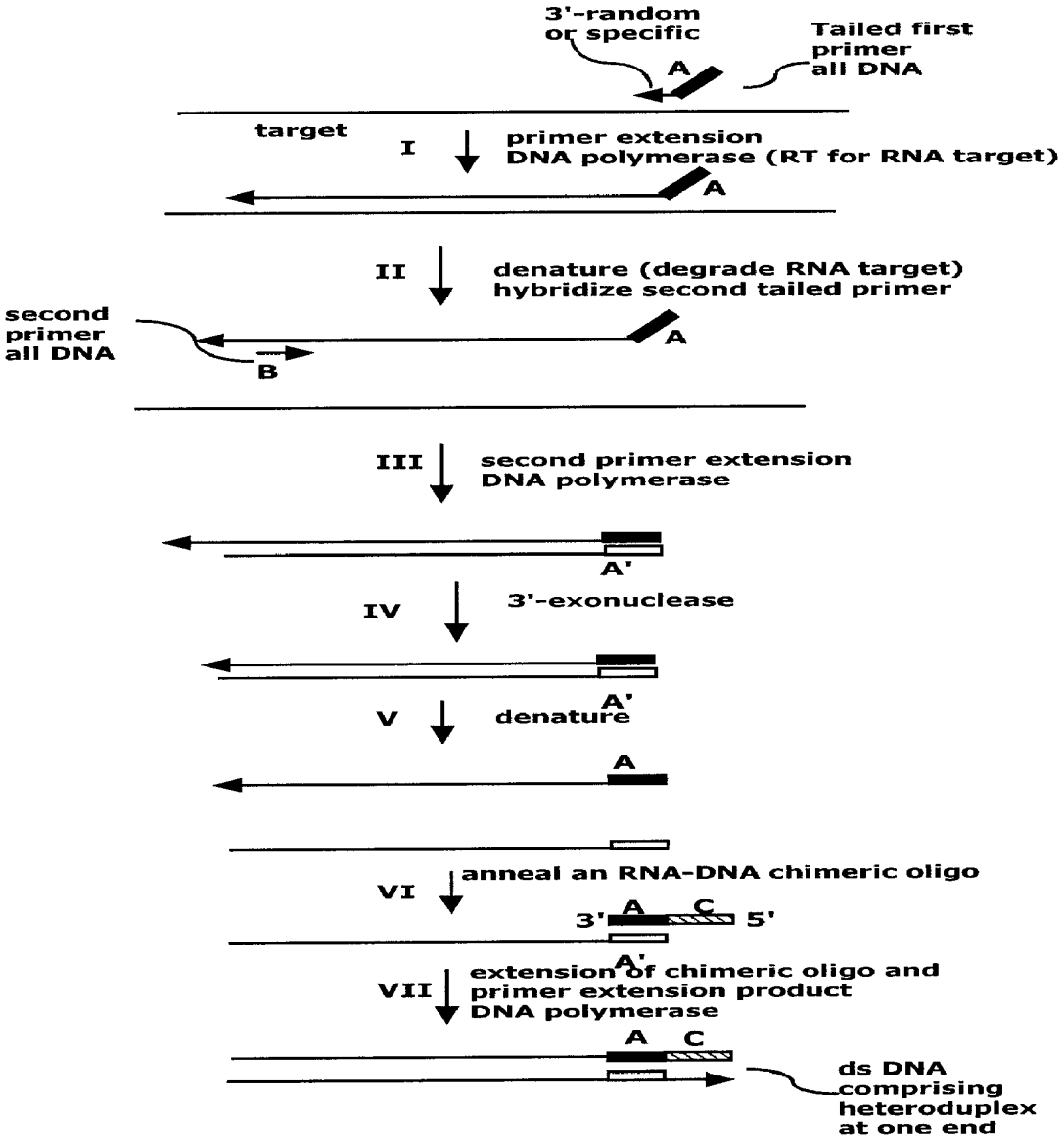
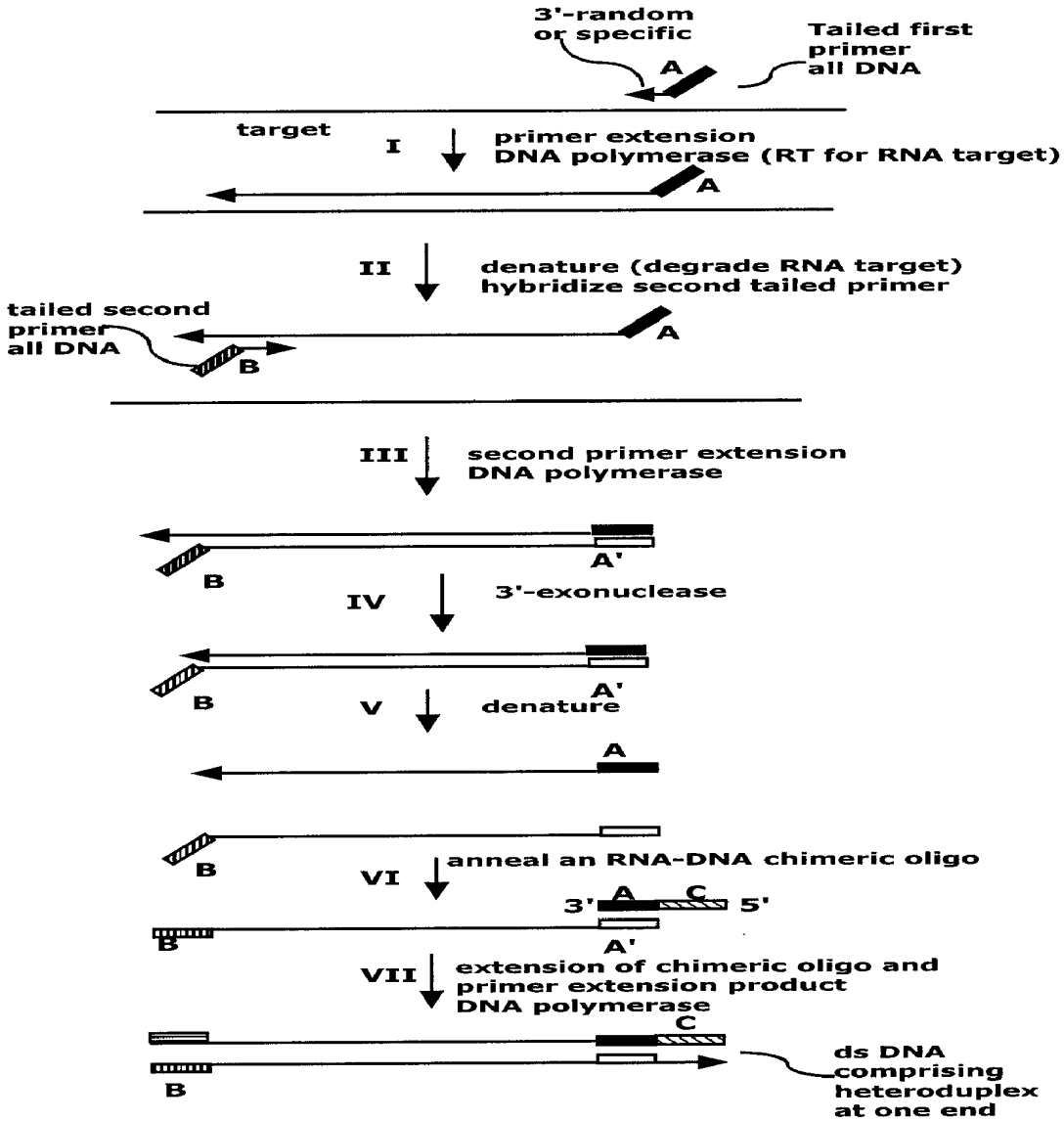


Figure 16
Generation of ds DNA complex with an RNA-DNA heteroduplex at one end suitable for SPIA amplification using tailed and non-tailed first and second DNA primers



A=5' tail of first primer (DNA) and 3'-end of the chimeric oligonucleotide
C= 5'-end RNA portion of the chimeric oligonucleotide

Figure 17
Generation of ds DNA complex with an RNA-DNA heteroduplex at one end suitable for SPIA amplification using tailed DNA primers



A=5' tail of first primer (DNA) and
3'-end of the chimeric oligonucleotide
B= 5'-tail of second primer (DNA)
C= 5'-end RNA portion of the
chimeric oligonucleotide

Figure 18: SPIA amplification from the ds DNA with Heteroduplex at one end generated as in Figures 16 and 17

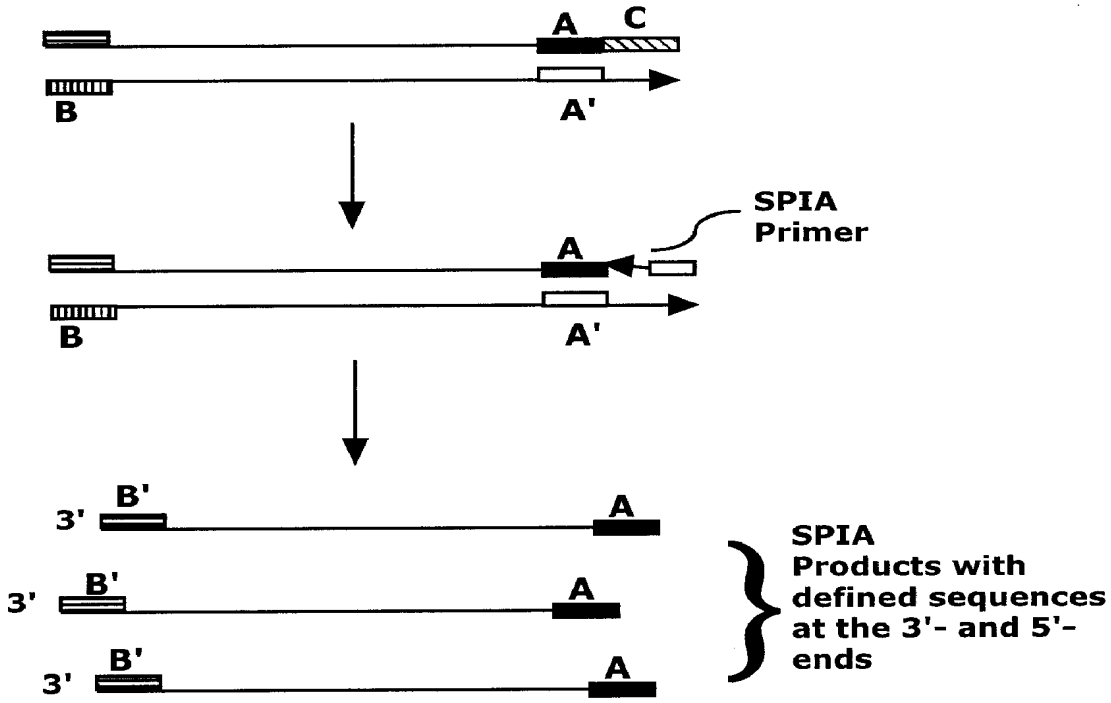


Figure 19

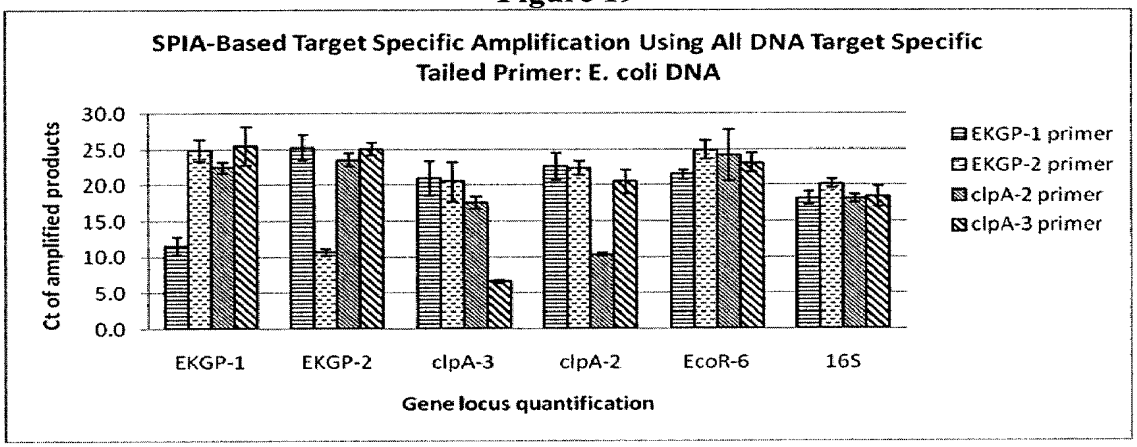

SPIA-Based Target Specific Amplification Using All DNA Target Specific Tailed Primer: E. coli DNA
Ct of amplified products
30.0
25.0
20.0
15.0
10.0
5.0
0.0
EKGP-1
EKGP-2
clpA-3
clpA-2
EcoR-6
16S
Gene locus quantification
EKGP-1 primer
EKGP-2 primer
clpA-2 primer
clpA-3 primer

Figure 20

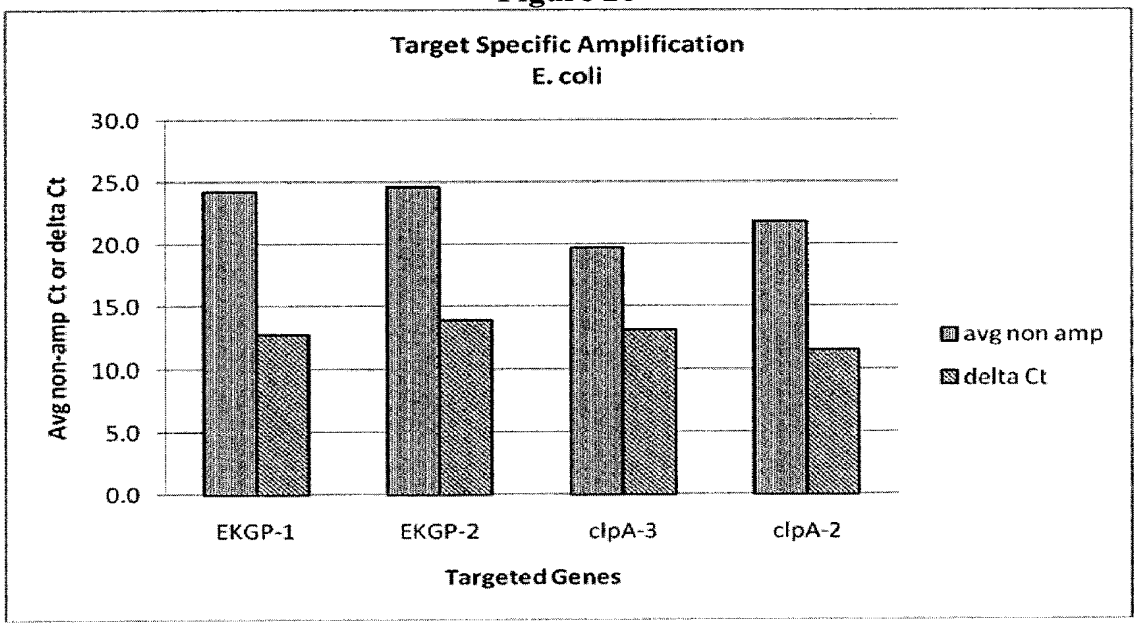

Target Specific Amplification
E. coli
Avg non-amp Ct or delta Ct
30.0
25.0
20.0
15.0
10.0
5.0
0.0
EKGP-1
EKGP-2
clpA-3
clpA-2
Targeted Genes
avg non amp
delta Ct

Figure 21

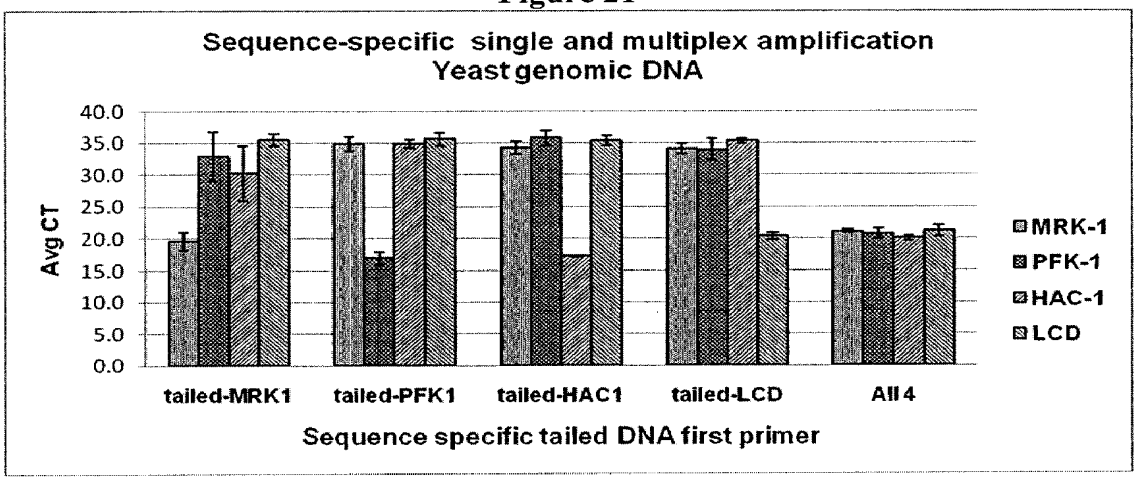
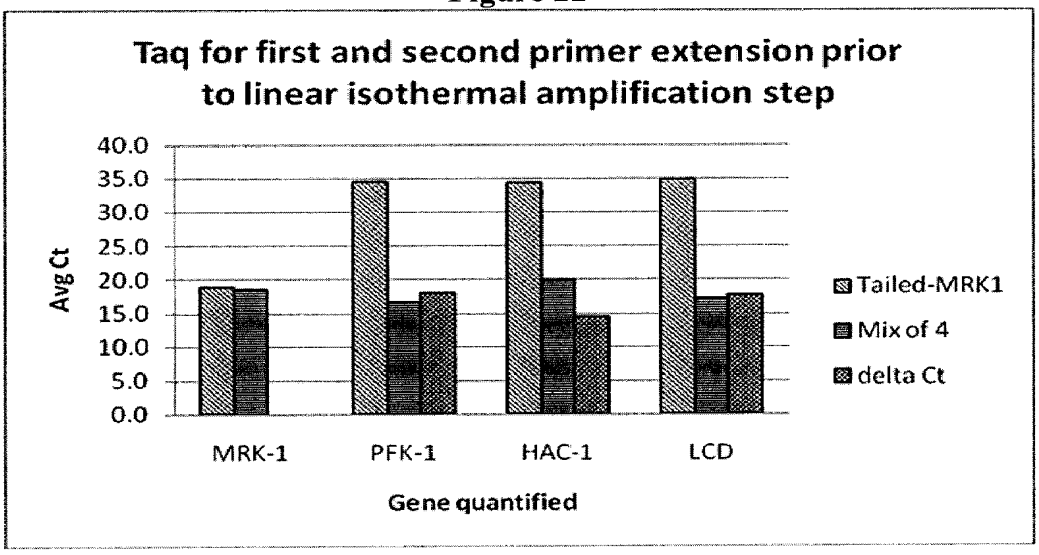

Sequence-specific single and multiplex amplification
Yeast genomic DNA
Avg CT
40.0
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0
tailed-MRK1
tailed-PFK1
tailed-HAC1
tailed-LCD
All 4
Sequence specific tailed DNA first primer
MRK-1
PFK-1
HAC-1
LCD

Figure 22


Taq for first and second primer extension prior to linear isothermal amplification step
40.0
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0
Avg Ct
MRK-1
PFK-1
HAC-1
LCD
Gene quantified
Tailed-MRK1
Mix of 4
delta Ct

Espacenet

Bibliographic data: WO2008015396 (A2) — 2008-02-07

METHOD OF LIBRARY PREPARATION AVOIDING THE FORMATION OF ADAPTOR DIMERS

Inventor(s): EARNSHAW DAVID JAMES [GB]; GORMLEY NIALL ANTHONY [GB]; BURGAR HELEN RACHEL [GB]; SMITH MELANIE ANNE [GB] ± (EARNSHAW, DAVID, JAMES, ; GORMLEY, NIALL, ANTHONY, ; BURGAR, HELEN, RACHEL, ; SMITH, MELANIE, ANNE)

Applicant(s): SOLEXA LTD [GB]; EARNSHAW DAVID JAMES [GB]; GORMLEY NIALL ANTHONY [GB]; BURGAR HELEN RACHEL [GB]; SMITH MELANIE ANNE [GB] ± (SOLEXA LIMITED, ; EARNSHAW, DAVID, JAMES, ; GORMLEY, NIALL, ANTHONY, ; BURGAR, HELEN, RACHEL, ; SMITH, MELANIE, ANNE)

Classification: **- international:** ***C12Q1/68***
- cooperative: **C12Q1/6806**; **C12Q1/6855**; C12Q1/6874 more

Application number: WO2007GB02865 20070730 Global Dossier

Priority number(s): US20060834424P 20060731

Also published as: WO2008015396 (A3) EP2049682 (A2) US2010167954 (A1) US9328378 (B2)

Abstract of WO2008015396 (A2)

The invention relates to a method of preparing a library of template polynucleotides which reduces and/or prevents the formation of adaptor-dimers. The invention also relates to the use of a library of templates prepared using the method of the invention for solid-phase nucleic acid amplification. In particular, the invention relates to a method of preparing a library of template polynucleotides which have common sequences at their 5' ends and at their 3' ends which is substantially free of adaptor-dimers.

IDT Exhibit 1020.3505

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 February 2008 (07.02.2008)

PCT

(10) International Publication Number
WO 2008/015396 A2

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

(21) International Application Number:
PCT/GB2007/002865

(22) International Filing Date: 30 July 2007 (30.07.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/834,424 31 July 2006 (31.07.2006) US

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

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **EARNSHAW, David, James** [GB/GB]; c/o Solexa Limited, Chesterford Research Park, Little Chesterford, Saffron Walden, Essex CB10 1XL (GB). **GORMLEY, Niall, Anthony** [IE/GB]; c/o Solexa Limited, Chesterford Research Park, Little Chesterford, Saffron Walden, Essex CB10 1XL (GB). **BURGAR, Helen, Rachel** [GB/GB]; c/o Solexa Limited, Chesterford Research Park, Little Chesterford, Saffron Walden, Essex CB10 1XL (GB). **SMITH, Melanie, Anne** [GB/GB]; Solexa Limited, Chesterford Research Park, Little Chesterford, Saffron Walden, Essex CB10 1XL (GB).

(74) Agents: **WHITE, Nina, Louise** et al.; Boult Wade Tennant, Verulam Gardens, 70 Gray's Inn Road, London WC1X 8BT (GB).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, 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, LT, 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.

(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, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

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

WO 2008/015396 A2

(54) Title: METHOD OF LIBRARY PREPARATION AVOIDING THE FORMATION OF ADAPTOR DIMERS

(57) Abstract: The invention relates to a method of preparing a library of template polynucleotides which reduces and/or prevents the formation of adaptor-dimers. The invention also relates to the use of a library of templates prepared using the method of the invention for solid-phase nucleic acid amplification. In particular, the invention relates to a method of preparing a library of template polynucleotides which have common sequences at their 5' ends and at their 3' ends which is substantially free of adaptor-dimers.

IDT Exhibit 1020.3506

Method of library preparation avoiding the formation of adaptor dimers

Field of the invention

The invention relates to a method of preparing a library of template polynucleotides which reduces and/or prevents the formation of adaptor-dimers. The invention also relates to the use of a library of templates prepared using the method of the invention for solid-phase nucleic acid amplification. In particular, the invention relates to a method of preparing a library of template polynucleotides which have common sequences at their 5' ends and at their 3' ends which is substantially free of adaptor dimers.

Background of 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 documents is incorporated by reference herein.

The ability to acquire and analyse DNA sequence data has increased phenomenally over the past few years. As a result nucleic acid analysis has become increasingly important in many areas of biology, biotechnology and medicine. Molecular biology and pharmaceutical drug development now make intensive use of nucleic acid analysis. The most challenging areas are whole genome sequencing, single nucleotide polymorphism detection, screening and gene expression monitoring, which typically require generation and analysis of large amounts of nucleic acid sequence data.

One area of technology which revolutionised the study of nucleic acids was the development of nucleic acid

IDT Exhibit 1020.3507

amplification techniques, such as the polymerase chain reaction (PCR). Amplification reactions, such as PCR, can enable the user to specifically and selectively amplify a particular target nucleic acid of interest from a complex mixture of nucleic acids. However, there is also an ongoing need for nucleic acid amplification techniques which enable simultaneous amplification of complex mixtures of templates of diverse sequence, such as genomic DNA fragments (e.g. 'whole genome' amplification) or cDNA libraries, in a single amplification reaction.

PCR amplification cannot occur in the absence of annealing of forward and reverse amplification primers to primer binding sequences in the template to be amplified under the conditions of the annealing steps of the PCR reaction, i.e. if there is insufficient complementarity between primers and template. Some prior knowledge of the sequence of the template is therefore required before one can carry out a PCR reaction to amplify a *specific* template, unless random primers are used with a consequential loss of specificity. The user must usually know the sequence of at least the primer-binding sites in the template in advance so that appropriate primers can be designed, although the remaining sequence of the template may be unknown. The need for prior knowledge of the sequence of the template increases the complexity and cost of PCR amplification of complex mixtures of templates, such as genomic DNA fragments.

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 the ends of the different fragments all contain

IDT Exhibit 1020.3508

the same DNA sequence. Fragments with universal ends can be amplified in a single reaction with a single pair of 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 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 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 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 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 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 in the template to which the primers anneal under conditions

IDT Exhibit 1020.3509

of the amplification reaction may be referred to herein as 'primer-binding' sequences.

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 '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 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 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. 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.

Such 'universal primer' approaches to PCR amplification, and in particular solid-phase bridging amplification, are advantageous since they enable multiple template molecules of the same or different, known or unknown sequence to be amplified in a single amplification reaction, which may be carried out on a solid support bearing a single pair of 'universal' primers. Simultaneous

IDT Exhibit 1020.3510

amplification of a mixture of templates of different sequences would otherwise require a plurality of primer pairs, each pair being complementary to each unique template in the mixture. The generation of a plurality of primer pairs for each individual template is not a viable option for complex mixtures of templates.

The addition of universal priming sequences onto the ends of targets to be amplified by PCR can be achieved by a variety of methods known to those skilled in the art. For example, a universal primer consisting of a universal sequence at its 5′ end and a degenerate sequence at its 3′ end can be used in a PCR (DOP-PCR, eg PNAS 1996 vol 93 pg 14676-14679) to amplify fragments randomly from a complex target sequence or a complex mixture of target sequences. The degenerate 3′ portion of the primer anneals at random positions on DNA and can be extended to generate a copy of the target that has the universal sequence at its 5′ end.

Alternatively, adaptors that contain universal priming sequences can be ligated onto the ends of the target sequences. The adaptors may be single-stranded or double-stranded. If double-stranded, they may have overhanging ends that are complementary to overhanging ends on the target molecules that may have been generated with a restriction endonuclease, or added with a DNA polymerase or terminal transferase. Alternatively, the double-stranded adaptors may be blunt, in which case the targets are also blunt ended. The blunt ends of the targets may have been formed during a process to shear the DNA into fragments, or they may have been formed by an end repair reaction, as would be well known to those skilled in the art.

A single adaptor or two different adaptors may be used in a ligation reaction with target sequences. If a target

IDT Exhibit 1020.3511

has been manipulated such that its ends are the same, i.e. both are blunt or both have the same overhang, then ligation of a single compatible adaptor will generate a template with that adaptor on both ends. However, if two compatible adaptors, adaptor A and adaptor B, are used, then three permutations of ligated products are formed: template with adaptor A on both ends, template with adaptor B on both ends, and template with adaptor A on one end and adaptor B on the other end. This last product is, under some circumstances, the only desired product from the ligation reaction and consequently additional purification steps are necessary following the ligation reaction to purify it from the ligation products that have the same adaptor at both ends.

A major drawback in preparing nucleic acid fragment libraries by ligating adaptors to the ends of template nucleic acid fragments is the formation of adaptor-dimers. Adaptor-dimers are formed by the ligation of two adaptors directly to each other such that they do not contain a template nucleic acid fragment as an insert. Such molecules are undesirable, in that during any amplification steps, for example during a universal amplification reaction, adaptor-dimers are amplified alongside the nucleic acid fragment library. Since adaptor-dimers are generally smaller than the fragments contained in the libraries they amplify and accumulate at a faster rate. This reduces the efficiency of the amplification reaction by limiting amplification of the library fragments by depletion of components, such as for example dNTP's and primers, in the amplification reaction. Another more serious concern that when such amplified fragments are sequenced they do not give useful sequence information since they contain no insert. In the case of

IDT Exhibit 1020.3512

clustered arrays, a significant population of clusters that have no target DNA sequence is undesirable due to the lower density of real sequence data obtained from a chip of finite size. Hence the efficiency of sequencing can be significantly reduced. Thus, the preparation of libraries with a low level of adaptor-dimers is highly advantageous in the sequencing of polynucleotides, particularly when such processes are high-throughput.

The invention presented herein is directed to a method of generating a library of template polynucleotides using a single adaptor construct in a ligation reaction which reduces and/or prevents the formation of adaptor-dimers. The method can be applied to preparing simple or complex populations of templates for amplification, for example on a solid surface, using primer sequences, with no prior knowledge of the target sequences. The invention is applicable to the preparation of templates from complex samples such as whole genomes or mixtures of cDNAs, as well as mono-template applications.

Summary of the invention

A first aspect of the invention provides a method for generating a library of template polynucleotide molecules from one or more primary polynucleotide molecules; said method comprising:

(a) fragmenting said one or more primary polynucleotide molecules to produce target polynucleotide duplexes;

(b) ligating an adaptor polynucleotide construct to both ends of the target polynucleotide duplexes to form combined ligated adaptor-target-adaptor sequences;

IDT Exhibit 1020.3513

(c) preparing an amplification reaction comprising said combined ligated adaptor-target-adaptor sequences and at least two different primer oligonucleotides wherein said primer oligonucleotides are complementary to both a part of the adaptor polynucleotide sequence portion of the combined ligated adaptor-target-adaptor sequences, and complementary to at least one base of either the target polynucleotide duplexes, or an overhang attached to the target duplex;

(d) carrying out an amplification reaction in which said at least two different primer oligonucleotides are annealed to complementary parts of the adaptor-target-adaptor sequence and extended by sequential addition of nucleotides to generate amplification products complementary to at least one strand of the combined ligated adaptor-target-adaptor sequences, wherein said amplification products have a first common sequence at their 5′ ends and a second common sequence at their 3′ ends and collectively provide a library of template polynucleotide molecules. See Figure 1 for schematic representation.

A second aspect of the invention provides a method for generating a library of template polynucleotide molecules from one or more primary polynucleotide molecules; said method comprising:

(a) fragmenting said one or more primary polynucleotide molecules to produce target polynucleotide duplexes;

(b) ligating an adaptor polynucleotide construct to both ends of the target polynucleotide duplexes to form combined ligated adaptor-target-adaptor sequences; wherein said

IDT Exhibit 1020.3514

adaptor polynucleotide construct contains at least one overhanging base that is modified such that the nucleic acid strand is resistant to exonucleolysis;

(c) preparing an amplification reaction comprising said combined ligated adaptor-target-adaptor sequences and at least two different primer oligonucleotides wherein said primer oligonucleotides are complementary to a part of the adaptor polynucleotide sequence portion of the combined ligated adaptor-target-adaptor sequences;

(d) carrying out an amplification reaction in which said at least two different primer oligonucleotides are annealed to complementary parts of the adaptor-target-adaptor sequence and extended by sequential addition of nucleotides to generate amplification products complementary to at least one strand of the combined ligated adaptor-target-adaptor sequences and wherein said amplification products have a common sequence at their 5′ ends and a common sequence at their 3′ ends and collectively provide a library of template polynucleotide molecules.

A third aspect of the invention combines the first and second aspects, and provides a method for generating a library of template polynucleotide molecules from one or more primary polynucleotide molecules; said method comprising:

(a) fragmenting said one or more primary polynucleotide molecules to produce target polynucleotide duplexes;

(b) ligating an adaptor polynucleotide construct to both ends of the target polynucleotide duplexes to form combined

IDT Exhibit 1020.3515

ligated adaptor-target-adaptor sequences; wherein said adaptor polynucleotide construct contains at least one overhanging base that is modified such that the nucleic acid strand is resistant to exonucleolysis;

(c) preparing an amplification reaction comprising said combined ligated adaptor-target-adaptor sequences and at least two different primer oligonucleotides wherein said primer oligonucleotides are complementary to both a part of the adaptor polynucleotide sequence portion of the combined ligated adaptor-target-adaptor sequences, and complementary to at least one base of either the target polynucleotide duplexes, or an overhang attached to the target duplex;

(d) carrying out an amplification reaction in which said at least two different primer oligonucleotides are annealed to complementary parts of the adaptor-target-adaptor sequence and extended by sequential addition of nucleotides to generate amplification products complementary to at least one strand of the combined ligated adaptor-target-adaptor sequences and wherein said amplification products have common sequences at their 5′ ends and common sequences at their 3′ ends and collectively provide a library of template polynucleotide molecules.

A fourth aspect of the invention relates to the use of a library of template polynucleotide molecules prepared according to the method of the first, second or third aspects of the invention as a template for solid-phase nucleic acid amplification. Thus, in a particular embodiment the invention provides a method of solid-phase nucleic acid amplification of template polynucleotide molecules which comprises: preparing a library of template polynucleotide

IDT Exhibit 1020.3516

molecules which have common sequences at their 5′ and 3′ ends using the method according to the first, second or third aspects of the invention and carrying out a solid-phase nucleic acid amplification reaction wherein said template polynucleotide molecules are amplified.

Accordingly, the fourth aspect of the invention relates to methods wherein a library of template polynucleotide molecules prepared according to the method of the first, second or third aspects of the invention is used as a template for solid-phase nucleic acid amplification.

In a fifth aspect, the invention provides a kit for use in preparing a library of template polynucleotide molecules which have common sequences at their 5' and 3' ends wherein the common sequence at the 5' end of each individual template in the library is not identical and not fully complementary to the common sequence at the 3' end of said template, the kit comprising mismatched adaptor polynucleotides as defined herein in relation to the first, second or third aspects of the invention and oligonucleotide amplification primers capable of annealing to the mismatched adaptor polynucleotides.

Brief Description of the Drawings

Figure 1 shows a schematic of a modified PCR step whereby adaptor dimer amplification is reduced and an increased length of primers used in PCR.

Figure 2 shows a scan of a polyacrylamide gel stained to visualize DNA following electrophoresis. The lanes are as follows: M) marker lane; a) Amplification of library using unmodified PCR primers; b) Amplification of library using PCR primers having a phosphorothioate bond

IDT Exhibit 1020.3517

modification; c) Amplification of library using PCR primers having a phosphorothioate bond and 2'-OMe modification.

Figure 3 shows a scan of a polyacrylamide gel stained to visualize DNA following electrophoresis. The lanes are as follows: M) marker lane; a) Amplification of library using unmodified PCR primers; b) Amplification of library using unmodified long PCR primers; c) Amplification of library using PCR primers having a phosphorothioate bond modification;
d) Amplification of library using exonuclease I treated PCR primers, which have a phosphorothioate modification.

Detailed description of the invention

The invention relates to a method of preparing a library of template polynucleotides which reduces and/or prevents the formation of adaptor-dimers during preparation of the library. As a result, the method of the invention is advantageous because it reduces the number of adaptor sequences recorded during any subsequent sequencing process.

As explained in further detail below, all templates within the library will contain regions of common sequence at (or proximal to) their 5' and 3' ends, wherein the common sequence at the 5' end of each individual template in the library is not identical and not fully complementary to the common sequence at the 3' end of said template.

The term 'library' refers to a collection or plurality of template molecules which share common sequences at their 5' ends and common sequences at their 3' ends. Use of the term 'library' to refer to a collection or plurality of template molecules should not be taken to imply that the templates making up the library are derived from a particular source, or that the 'library' has a particular

IDT Exhibit 1020.3518

composition. By way of example, use of the term 'library' should not be taken to imply that the individual templates within the library must be of different nucleotide sequence or that the templates be related in terms of sequence and/or source.

In its various embodiments, the invention encompasses formation of so-called 'monotemplate' libraries, which comprise multiple copies of a single type of template molecule, each having common sequences at their 5′ ends and their 3′ ends, as well as 'complex' libraries wherein many of the individual template molecules comprise different target sequences, although all share common sequences at their 5′ ends and 3′ ends. Such complex template libraries may be prepared using the method of the invention starting from a complex mixture of target polynucleotides such as (but not limited to) random genomic DNA fragments, cDNA libraries etc. The invention also extends to 'complex' libraries formed by mixing together several individual 'monotemplate' libraries, each of which has been prepared separately using the method of the invention starting from a single type of target molecule (i.e. a monotemplate), or libraries made from different 'complex' libraries further pooled. All templates in a given library will share common sequence at their 5′ ends and common sequence at their 3′ ends.

Use of the term 'template' to refer to individual polynucleotide molecules in the library merely indicates that one or both strands of the polynucleotides in the library are capable of acting as templates for template-dependent nucleic acid polymerisation catalysed by a polymerase. Use of this term should not be taken as limiting the scope of the invention to libraries of polynucleotides

IDT Exhibit 1020.3519

which are actually used as templates in a subsequent enzyme-catalysed polymerisation reaction.

The library of the invention is formed by first ligating identical adaptor polynucleotide molecules ('mismatched adaptors', the general features of which are defined below) to the 5' and 3' ends of one or more target polynucleotide duplexes (which may be of known, 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 optionally amplified copies thereof, collectively provide a library of template polynucleotides.

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 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 strands so as to provide, when the two strands are annealed, at least one double-stranded region and at least one unmatched region.

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

IDT Exhibit 1020.3520

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

IDT Exhibit 1020.3521

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 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 should be 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 a enzyme-catalysed ligation reaction, provided that the two strands revert to single stranded form under annealing conditions in an amplification reaction.

The adaptor constructs may contain exonuclease resistant modifications as described below. Such modifications lower the level of adaptor-dimers present in the library as the two adaptors can not undergo ligation

IDT Exhibit 1020.3522

without removal of their non complementary overhangs, which are non-complementary. The adaptors can be treated with an exonuclease enzyme, as described below, 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.

Each strand of each template molecule in the library formed in the primer extension reaction will therefore have the following structure, when viewed as a single strand:

5'-[common sequence I]-[target sequence]-[common sequence II]-3'

wherein 'common sequence I' represents a sequence derived from copying a first strand of the mismatched adaptor and is common to all template molecules in the library generated in the initial primer extension reaction; 'target' represents a sequence derived from one strand of the target polynucleotide duplex and may be different in different individual template molecules within the library; and 'common sequence II' represents a sequence derived from copying of a second strand of the mismatched adaptor and is also common to all template molecules in the library generated in the initial primer extension reaction.

Since 'common sequence I' and 'common sequence II' are common to all template strands in the library they may include 'universal' primer-binding sequences, enabling all templates in the library to be ultimately amplified in a solid-phase PCR procedure using universal primers.

IDT Exhibit 1020.3523

It is a key feature of the invention, however, that the common 5' and 3' end sequences denoted 'common sequence I' and 'common sequence II' are not *fully* complementary to each other, meaning that each individual template strand can contain different (and non-complementary) universal primer sequences at its 5' and 3' ends.

It is generally advantageous for complex libraries of templates to be amplified, for example by PCR or isothermal amplification (e.g. whole genome amplification), either in solution or on a solid support, to include regions of 'different' sequence at their 5' and 3' ends, which are nevertheless common to all template molecules in the library, especially if the amplification products are to be ultimately sequenced. For example, the presence of common unique sequence at one end only of each template in the library can provide a binding site for a sequencing primer, enabling one strand of each template in the amplified form of the library to be sequenced in a single sequencing reaction using a single type of sequencing primer. Equally the method of the invention may be applied to the preparation of libraries which are amplified *in-vivo*, such as for example bacterial cDNA libraries and the like.

Typically 'common sequence I' and 'common sequence II' will consist of no more than 100, or no more than 50, or no more than 40 consecutive nucleotides at the 5' and 3' ends, respectively, of each strand of each template polynucleotide. The precise length of the two sequences may or may not be identical. The nucleotide sequences of 'common sequence I' and 'common sequence II' in the template polynucleotides will be determined in part by the sequences of the adaptor strands ligated to the target polynucleotides and in part by the sequence of the primer used in the

IDT Exhibit 1020.3524

initial primer extension reaction, and any subsequent rounds of nucleic acid amplification.

In embodiments wherein the initial primer extension product is subjected to further amplification by conventional PCR, then the products of the amplification reaction will be double-stranded polynucleotides, one strand of which has the structure:

5'-[common sequence I]-[target sequence]-[common sequence II]-3'

It will be appreciated that 'common sequence II' in the amplification products may differ somewhat to the 'common sequence II' present in the products of the initial primer extension reaction, since the former will be determined in part by the sequence of the PCR primer used to prime synthesis of a polynucleotide strand complementary to the initial primer extension product, whereas the latter will be determined solely by copying of the adaptor sequences at the 3' ends of the adaptor-template constructs in the initial primer extension. Nevertheless, since the PCR primer is designed to anneal to a sequence in the initial extension products which is complementary to the 3' adaptor, the two forms of 'common sequence II' will contain identical sequence, at least at the 3' end. Additional sequence may be included at the 5' end of 'common sequence II' in the amplified products, for example by the use of 'tailed' PCR primers, as described in detail below. In other embodiments the common sequences present in the amplification products may actually be shorter than the common sequences included in the adaptors originally ligated to the target.

The precise nucleotide sequences of the common regions of the template molecules in the library are generally not

IDT Exhibit 1020.3525

material to the invention and may be selected by the user. The common sequences must at least comprise 'primer-binding' sequences which enable specific annealing of amplification primers when the templates are in use in a solid-phase amplification reaction. The primer-binding sequences are thus determined by the sequence of the primers to be ultimately used for solid-phase amplification. The sequence of these primers in turn is advantageously selected to avoid or minimise binding of the primers to the target portions of the templates within the library under the conditions of the amplification reaction, but is otherwise not particularly limited. By way of example, if the target portions of the templates are derived from human genomic DNA, then the sequences of the primers to be used in solid phase amplification should ideally be selected to minimise non-specific binding to any human genomic sequence.

The conditions encountered during the annealing steps of an amplification reaction will be generally known to one skilled in the art, although the precise annealing conditions will vary from reaction to reaction (see 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.*). Typically such conditions may comprise, but are not limited to, (following a denaturing step at a temperature of about 94°C for about one minute) exposure to a temperature in the range of from 40°C to 72°C (preferably 50-68°C) for a period of about 1 minute in standard PCR reaction buffer.

Different annealing conditions may be used for a single primer extension reaction not forming part of a PCR reaction (again see Sambrook *et al.*, 2001, *Molecular Cloning, A Laboratory Manual*, 3rd Ed, Cold Spring Harbor Laboratory

IDT Exhibit 1020.3526

Press, Cold Spring Harbor Laboratory Press, NY; Current Protocols, eds Ausubel *et al.*). Conditions for primer annealing in a single primer extension include, for example, exposure to a temperature in the range of from 30 to 37°C in standard primer extension buffer. It will be appreciated that different enzymes, and hence different reaction buffers, may be used for a single primer extension reaction as opposed to a PCR reaction. There is no requirement to use a thermostable polymerase for a single primer extension 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 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

IDT Exhibit 1020.3527

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 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, for example to 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. The adaptors may further include 'tag' sequences, which can be used to tag or mark template molecules derived from a particular source. The general features and use of such tag sequences is described in the applicant's pending application published as WO 05/068656.

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,

IDT Exhibit 1020.3528

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 method comprises a first step of fragmenting one or more primary polynucleotide molecules to produce target polynucleotide duplexes.

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

IDT Exhibit 1020.3529

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 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 sequence (coding 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 particular chromosomes, for example. Yet more particularly, the 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 fragmentation processes, and prior to or subsequent to the ligation of the adaptor sequences.

The sequence of the primary polynucleotide molecules may be the same or different, for example, a mixture of primary polynucleotide molecules of different sequences may be prepared by mixing a plurality of individual primary polynucleotide molecules. For example, DNA from more than one source can be prepared if each DNA sample is first tagged to enable its identification after it has been sequenced. Many different suitable DNA-tag methodologies already exist in the art and are well within the purview of the skilled person.

IDT Exhibit 1020.3530

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 because the larger piece of nucleic acid sequence remains in intact (i.e., is not fragmented by the PCR amplification). The random fragmentation is designed to produce fragments irrespective of the sequence identity or position 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 to 700 base 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) 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, 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

IDT Exhibit 1020.3531

be introduced during an enzymatic treatment, for example using 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 has a nontemplate-dependent terminal transferase activity that adds a single deoxynucleotide, for example deoxyadenosine (A) to the 3′ ends of, for example, PCR 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 prevents self-ligation of both vector and target such that there is a bias towards formation of the combined ligated adaptor-target sequences.

The term 'target polynucleotide duplexes' refers to nucleic acid molecules that it is desired to sequence. The term 'template' refers to the target sequences ligated to the adaptor sequences; and thus the 'templates' are suitable for amplification; whereas the 'targets' without the adaptors are not.

The second step of the method comprises ligating a double stranded adaptor polynucleotide sequence to both ends of the target polynucleotide duplexes to form combined ligated adaptor-target-adaptor polynucleotide sequences. It is particularly advantageous to use the same adaptor

IDT Exhibit 1020.3532

construct for both ends of the target duplex, although two sets of adaptors can also be utilised.

Ligation methods are known in the art and utilise standard methods (Sambrook and Russell, Molecular Cloning, A Laboratory Manual, third edition). Such methods utilise 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 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 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 phosphodiester linkage between the two polynucleotide strands, but other means of covalent linkage (e.g. non-phosphodiester backbone linkages) may be used.

Whilst the method of TA ligation (or TA cloning) is known in the art, the presence of the 'T' overhang does not fully remove the formation and presence of adaptor-dimer constructs in the library. The Inventors have made the surprising discovery that during ligation, even with an enzyme which is purported not to have measurable exonuclease activity, the overhanging nucleotide(s) is/are removed from the adaptors with surprising frequency. Thus even with the TA method of cloning, adaptor-dimers are not prevented. The

IDT Exhibit 1020.3533

deficiencies of the TA ligation protocol are addressed by the methods disclosed herein.

According to a particular aspect of the invention, in addition to the mismatched region previously described, the adaptor construct or constructs contains an overhanging base or bases at the 3'- end of one of the strands that is complementary to the overhanging base or bases at the 3'-end of the target duplexes, and a 5'-phosphate moiety on the complementary hybridised strand, The adaptor constructs may also contain a region on one, or both, of the strands that does not hybridise with a sequence on the other strand of the adaptor. Such 'mismatched' adaptors can serve as priming sites for further amplification reactions, and may allow for amplification with primers extending beyond the sequence of the ligated adaptor. Thus the region of known sequence in the template for amplification may be longer than the adaptor sequence ligated to the target. See Figure 1.

The library of template polynucleotide molecules is particularly suitable for use in solid phase sequencing methods. Because sequence reads may be short, that is around 25-50 base pairs in length, unlike conventional methods of library preparation, it is of no consequence if multiple different target polynucleotide duplexes are ligated into a single template polynucleotide. Because the sequence read is shorter than the length of the individual target polynucleotide duplexes, there is no risk of artificial concatamers of sequence data being produced. The formation of target concatamers is minimised by the presence of an excess of the adaptor constructs.

Optionally the combined ligated polynucleotide sequences and unligated adaptor polynucleotide constructs may be purified from any components of the ligation

IDT Exhibit 1020.3534

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).

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 an amplification reaction. Generally amplification reactions require at least two amplification primers, often denoted 'forward' and 'reverse' primers (primer oligonucleotides) 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.

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 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 'adaptor-target specific portion' relates to the length of

IDT Exhibit 1020.3535

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 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 starting adaptor-target strands. Typically the invention relates to the use of forward and reverse primers of sufficient length to hybridise to the whole of the adaptor sequence and at least one base of the target sequence. 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 length. The forward and reverse primers may be of significantly different lengths; for example one may be 20-40 bases, and 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.

IDT Exhibit 1020.3536

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 a new polynucleotide strand complementary to the template strand.

Primers may additionally comprise non-nucleotide chemical modifications, 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.

To reduce the amplification of ligated adaptor-dimers, the Inventors have made the discovery that the forward and reverse amplification primers should anneal to at least one base 'originating' from the target nucleic acid fragment (described below). See Figure 1. Such a base, or bases, can be part of the sequence of the target nucleic acid fragment,

IDT Exhibit 1020.3537

such as for example genomic sequence, or may be attached by a nucleotide tailing reaction, or by a ligation reaction and a subsequent cleavage reaction, for example with a restriction endonuclease. Attachment of a single nucleotide 3'-overhang is preferred. Accordingly the forward and reverse primers therefore contain the complementary base to the nucleotide triphosphate used in the target tailing reaction. The use of dATP in the enzymatic tailing reaction of the target nucleic acids means that the primers should contain a single 3'-'T' overhang. This base can be chemically introduced during primer synthesis, or introduced enzymatically in a similar tailing reaction used to tail the blunt ended target nucleic acids. For the avoidance of doubt, any bases added to the target in treatment steps carried out before the adaptor ligation step are considered to originate from the attached target during the adaptor attachment process.

In order to accurately copy the target DNA sequence, it is desirable to use a DNA polymerase with a high fidelity. Such high fidelity polymerases often have a 'proofreading' exonuclease domain to read and remove incorrectly incorporated nucleotides. Such high fidelity polymerases include Phusion™ DNA polymerase, PfuUltra™ DNA polymerase, Deep Vent® DNA polymerase or KOD DNA polymerase

The invention is directed to methods of using forward and reverse amplification primers, as described herein, designed such that the 3'-terminal base of the primers only hybridises to constructs that contain the target nucleic acid sequence, and not to adaptor dimers. This reduces amplification of the adaptor dimers, which is desirable to lower the number of adaptor-dimers from the template library. Not wishing to be bound by hypothesis, it is

IDT Exhibit 1020.3538

believed that since the primers have an additional 3′ terminal base which only hybridises to the target nucleic acid sequence, and because such a base is not present in the adaptor-dimers, the 3′ end of the primer forms a mis-matched end when hybridised to an adaptor-primer. The efficiency of the polymerase to extend this mis-matched end during amplification appears to be reduced. Hence the efficiency of amplification of adaptor-dimers is also reduced and the number of adaptor-primers contaminating a library is similarly lowered.

The Inventors have also made the further discovery that exonuclease activity of some polymerase enzymes used in an amplification reaction may, however, remove the terminal non hybridised bases from the amplification primers. This has the effect that amplification of the adaptor-dimers can occur, albeit at a reduced level. Therefore in a further particular embodiment of the invention, the amplification primers are modified to prevent removal of nucleotides from the 3′- end.

In a particular embodiment, the modification is a chemical modification. Exonucleoytic attack on the primer molecules can be efficiently prevented by the introduction of a single phosphorothioate bond at their 3′- termini (Nucleic Acids Research, 1992, 29, (14), 3551-4). Other such exonuclease resistant modifications may include phosphorodithioates, methyl phosphonates and 2′-O-methyl sugars, either separately or in combination. A number of other modifications are known to reduce the exonuclease degradation of single DNA strands, including phosphoramidites (P-NR2), phosphorofluoridates (P-F), boranophosphanes (P-BH3) or phosphoroselenoates (P-Se), and modifications to the sugar rings, such as 2′-O alkyl groups,

IDT Exhibit 1020.3539

2'- fluoro groups, 2'-amino groups such as 2-amino propyl (PNAS, 1999, 96 (25) p14240-45) or locked nucleic acids (LNA) where the 2' and 4' sugar positions are connected.

The amplification primers or adaptor constructs can optionally be treated with an exonuclease enzyme, for example a DNA polymerase with exonuclease activity, or exonuclease I, prior to use in the amplification reaction. In the case of the phosphorothioate modifications this removes the synthetic failure phosphate sequences, and also the phosphorothioate isomer that remains susceptible to exonucleolysis. The remaining material can be re-purified, if desired, then used in the amplification reaction. Such material is thereby completely resistant to exonucleolysis in the amplification reaction. Therefore the ability of the ligated adaptor-dimers to amplify is significantly reduced or even prevented, and they are thus removed from the library of ligated material. The result is that the efficiency of both amplification and/or sequencing of a library prepared according to the methods of the invention is significantly improved. This results in both a reduction of the costs of sequencing whilst increasing the quantity of useful sequence data.

The combination of both processes, namely the exonuclease treatment of the adaptor construct prior to ligation, to prevent the formation of the adaptor-dimers, and the exonuclease treatment of the amplification primers that overlap the target, to prevent amplification of the adaptor dimers is especially prefered, although both methods in isolation are within the scope of this invention.

IDT Exhibit 1020.3540

Use of the template library

Template libraries prepared using any of the methods of the invention may be used in essentially any method of nucleic acid analysis which requires further amplification of the templates and/or sequencing of the templates or amplification products thereof. Exemplary uses of the template libraries include, but are not limited to, providing templates for whole genome amplification and also solid-phase PCR amplification or solid-phase isothermal amplification (of either monotemplate or complex template libraries). A particularly preferred use is in whole-genome amplification carried out on a solid-support.

Whole-genome amplification

Template libraries prepared according to the methods of the invention starting from a complex mixture of genomic DNA fragments representing a whole or substantially whole genome provide suitable templates for so-called 'whole-genome' amplification. The term 'whole-genome amplification' refers to a nucleic acid amplification reaction (e.g. PCR) in which the template to be amplified comprises a complex mixture of nucleic acid fragments representative of a whole (or substantially whole) genome.

Solid-phase amplification

Once formed, the library of templates prepared according to the methods described above can be used for solid-phase nucleic acid amplification.

Thus, in further aspects the inventions provide use of the methods in solid-phase nucleic acid amplification of template polynucleotide molecules which comprises:

IDT Exhibit 1020.3541

preparing a library of template polynucleotide molecules which have common sequences at their 5' and 3' ends using a method according to the first, second or third aspects of the invention described herein and carrying out a solid-phase nucleic acid amplification reaction wherein said template polynucleotide molecules are amplified.

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

IDT Exhibit 1020.3542

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 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 other embodiments of the invention the forward and reverse primers may contain template-specific portions of different sequence.

In all embodiments of the invention, primers for solid-phase amplification are preferably immobilised by 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 one particularly

IDT Exhibit 1020.3543

preferred 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 and *N*-(5-bromoacetamidylpentyl) acrylamide (BRAPA).

The library may be amplified on beads wherein each bead contains a forward and reverse amplification primer. It is preferred to use the library of templates prepared according to the first, second or third aspects of the invention to prepare clustered arrays of nucleic acid colonies, analogous to those described in WO 00/18957 and WO 98/44151, 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 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.

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 are amplified; or an array of beads on a surface after the beads have been amplified.

IDT Exhibit 1020.3544

Clustered arrays can be prepared using either a process of thermocycling, as described in patent WO9844151, or a process whereby the temperature is maintained as a constant, and the cycles of extension and denaturing are performed using changes of reagents. Such isothermal amplification methods are described in patent application number US60/783618 (Isothermal methods for creating clonal single molecule arrays). 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 amplified nucleic acids generated by whole genome or solid-phase amplification. Thus, the invention provides a method of nucleic acid sequencing comprising amplifying a library of nucleic acid templates using whole genome or solid-phase amplification as described above and carrying out a nucleic acid sequencing reaction to determine the sequence of the whole or a part of at least one amplified nucleic acid strand produced in the whole genome or solid-phase amplification reaction.

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

IDT Exhibit 1020.3545

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 whole genome or 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.

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 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 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 acid sequencing it is preferred to remove substantially all or 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

IDT Exhibit 1020.3546

the template which is single-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'.

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 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 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 not be essential if the solid-phase amplification reaction is performed with only one primer covalently immobilised and the other in free solution.

In order to generate a linearised template suitable for sequencing it is necessary to remove 'unequal' amounts of the complementary strands in the bridged structure formed by amplification so as to leave behind a linearised template for sequencing which is fully or partially single stranded. Most preferably one strand of the bridged structure is substantially or completely removed.

Following the cleavage step, regardless of the method used for cleavage, the product of the cleavage reaction may be subjected to denaturing conditions in order to remove the portion(s) of the cleaved strand(s) that are not attached to

IDT Exhibit 1020.3547

the solid support. Suitable denaturing conditions, for example sodium hydroxide solution, formamide solution or heat, will be apparent to the skilled reader with reference 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 (and subsequent re-annealing of the cleaved strands) results in the production of a sequencing 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.

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 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.

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. Once 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 nature of the base incorporated into the growing chain has been determined, the

IDT Exhibit 1020.3548

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. 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 may be carried out containing each of the modified nucleotides separately.

The modified nucleotides may carry a label to facilitate their detection. Preferably this is a fluorescent label. Each nucleotide type may 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). However the detectable label need not 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).

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

IDT Exhibit 1020.3549

be used. Suitable alternative techniques include, for example, Pyrosequencing™, FISSEQ (fluorescent in situ sequencing), MPSS and sequencing by ligation-based methods.

The target polynucleotide to be sequenced using the method of the invention may be any polynucleotide that it is desired to sequence. Using the template library preparation method described in detail herein it is possible to prepare template libraries starting from essentially any double or single-stranded target polynucleotide of known, unknown or partially known sequence. With the use of clustered arrays prepared by solid-phase amplification it is possible to sequence multiple targets of the same or different sequence in parallel.

Kits

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

Preferred 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 at least one base originating from the attached target, and priming synthesis of an extension product, which extension product would include any target sequence ligated to the adaptor when the adaptor is in use.

The preferred 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 and properties of amplification primers will be well known to those skilled in the art. Suitable primers of appropriate nucleotide sequence for use with the adaptors included in

IDT Exhibit 1020.3550

the kit can be readily prepared using standard automated nucleic acid synthesis equipment and reagents in routine use 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 phase and/or on a suitable solid support (i.e. solid-phase amplification).

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 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 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 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, 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

IDT Exhibit 1020.3551

include 'universal' sequencing primers suitable for sequencing templates prepared using the mismatched adaptors and primers.

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

Example 1

The DNA source used was a purified 10kb region of Human BAC. The DNA was first prepared for ligation to forked adaptors by: fragmentation of the DNA by nebulisation, end repair of the DNA ends to make them blunt-ended and phosphorylation, 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). 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 adaptor at both ends of the fragments.

Materials and Methods

DNA Sample Preparation

Amplified 10kb region of human BAC DNA (140K human chromosome 6 insert in a pTARBAC vector) using two 5' phosphorylated primers (LRPCR primer 1: 5'Phosphate TGGAACAGCCGCTCTCACCT SEQ ID NO:1 and LRPCR primer 2: 5'Phosphate TCCTGGAGGGAAGTGACTAT SEQ ID NO:2). Used 0.5 µM each primer, 200 µM each dNTP, 0.02 ng/µl BAC DNA, 0.02U/µl Phusion™ polymerase (NEB/Finnzymes #F530S), diluted in 1X Finnzymes GC buffer.

PCR Programme :

IDT Exhibit 1020.3552

98°C 2 mins
98°C 30 secs
60°C 30 secs } 35 cycles
72°C 5.5 mins
72°C 10 mins
Hold at 4°C

Gel purified 10kb product on 0.7% TAE agarose gel, using Qiagen gel extraction kit (Qiagen #28706), according to manufacturer's instructions.

Nebulization

Materials:

- 10kb region of human BAC (from above)
- Buffer (glycerol 53.1 ml, water 42.1 ml, 1 M TrisHCl (pH7.5) 3.7 ml, 0.5 M EDTA 1.1 ml)
- Nebulizer Invitrogen (#K7025-05)
- Qiagen columns PCR purification kit #28104

Mix: 100µl (0.8 micrograms) of DNA
660µl Buffer

Procedure:

Chilled DNA solution was fragmented in the nebulizer on ice for 6 minutes under 32 pounds per square inch (psi) of pressure. The recovered volume was purified with a Qiagen PCR purification kit column and eluted in 30µl of EB (Qiagen).

End-repair

Materials:

- T4 DNA Polymerase NEB #M0203L

IDT Exhibit 1020.3553

- 10xNEB 2 buffer NEB #B7002S
- 100x BSA NEB #B9001S
- dNTPs mix (10mM each) NEB #N0447S
- E. coli DNA Pol I large fragment (Klenow) NEB #M0210S
- T4 polynucleotide kinase NEB #M0201L
- T4 PNK buffer NEB #B0201S
- 100 mM ATP Amersham #27-2056-01
- Qiagen columns PCR purification kit #28104

End repair mix was assembled as follows:

DNA	30µl
Water	7.5µl
10xNEB2	5µl
100xBSA	0.5µl
10mM dNTPs	2µl
T4 DNA pol (3U/µl)	5µl
	50µl total

The reaction was incubated for 15 min at room temperature, then 1µl of E. coli DNA Pol I large fragment (Klenow) was added and the reaction incubated for a further 15 min at room temperature. The DNA was purified from enzymes, buffer, etc. by loading the reaction mix on a Qiagen column, and eluting in 30µl EB (Qiagen). The 5' ends of the DNA were then phosphorylated using polynucleotide kinase as follows:

DNA	30µl
Water	9.5µl
10xPNK buffer	5µl

IDT Exhibit 1020.3554

100mM ATP	0.5µl
T4 PNK (10U/µl)	5µl
	50µl total

The reaction was incubated for 30 min at 37°C, then heat inactivated at 65°C for 20 min. DNA was then purified from enzymes, buffer, etc by loading the reaction mix on a Qiagen column, finally eluting in 30µl EB (Qiagen).

A- tailing reaction

Materials:

- Taq DNA polymerase NEB #M0267S
- 10x thermopol buffer NEB #B9004S
- 1mM dATP Amersham-Pharmacia #272050
- Qiagen MinElute column PCR purification kit #28004

The following reaction mix was assembled:

DNA	30µl
10x thermopol buffer	5µl
1mM dATP	10µl
Taq pol (5U/µl)	3µl
	~50µl total

The reaction was incubated for 30 min at 70°C, then the DNA purified from enzymes, buffer, etc. by loading the reaction mix on a Qiagen MinElute column, eluting in 10µl EB (Qiagen).

IDT Exhibit 1020.3555

Anneal forked adaptor

Materials:

- Oligo A: 5'ACACTCTTTCCCTACACGACGCTCTTCCGATCxT (x = phosphorothioate bond) SEQ ID NO:3
- Oligo B: 5'Phosphate-GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG SEQ ID NO:4
- 50mM Tris/50mM NaCl pH7
- PCR machine

100µM Oligo A	20µl
100µm Oligo B	20µl
Tris/NaCl	10µl
	50µl at 40µM duplex in 10mM Tris/10mM NaCl (pH 7.5)

The adaptor strands were annealed in a PCR machine programmed as 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
Forked adaptor was diluted to 15 µM final concentration in 10mM Tris/10mM NaCl (pH 7.5).

Ligation reaction

Materials:

- 15µM forked adaptor (from above)
- A-tailed genomic DNA (from above)
- Quick Ligase NEB #M2200L
- Quick Ligase 2x buffer NEB #B2200S
- PCR machine
- Qiagen columns PCR purification kit #28104

IDT Exhibit 1020.3556

Reaction mix was assembled as follows:

DNA	10µl
2x buffer	25µl
15 µM adaptor	10µl
Quick Ligase	5µl
	50µl total

The reaction was incubated for 30 min at room temperature, then the DNA purified from enzymes, buffer, etc. by loading the reaction mix on a Qiagen column, and eluting in 30µl EB (Qiagen).

Gel purification

Materials:

- Agarose Biorad #161-3101
- 100 base pair ladder NEB #N3231L
- TAE
- Loading buffer (50 mM Tris pH8, 40 mM EDTA, 40% w/v sucrose)
- Ethidium bromide
- Gel trays and tank. Electrophoresis unit
- Qiagen MinElute columns Qiagen(#28004)

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 50 min. The gel was then viewed on a 'White-light' box and fragments from 300bp to 700bp excised and purified with two minElute columns, eluting each in 15µl EB (Qiagen) and pooled.

IDT Exhibit 1020.3557

PCR amplification

Materials:

- Gel purified ligated DNA (from above)
- PCR PRIMER 1: 5′ AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCyT, where y= phosphorothioate bond and 2'-OMe SEQ ID NO:5
- PCR PRIMER 2: 5′ CAAGCAGAAGACGGCATACGAGCTCTTCCGATCyT, where y= phosphorothioate bond and 2'-OMe SEQ ID NO:6
- DeepVent NEB #M0258L
- 10x Thermopol buffer NEB #B9004S
- dNTPs mix (10mM each) NEB #N0447S
- PCR machine
- Qiagen columns PCR purification kit #28104

The PCR reaction was prepared as follows:

DNA	1µl
10x Thermopol buffer	5µl
dNTPs (10mM each)	1ul
25 µM PCR primer 1	1µl
25 µM PCR primer 2	1µl
DeepVent	1ul
Water	40µl
	50µl total

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

- 2 min @ 94°C
- [45 sec@ 94°C, 45 sec @ 65°C, 2 min @ 70°C] 18 cycles
- 5 min @ 70°C
- Hold @ 4°C

IDT Exhibit 1020.3558

PCR products were purified from enzymes, buffer, etc. on a Qiagen column, eluting in 30µl EB (Qiagen). The resulting DNA library is ready for amplification on a surface amplification platform.

Validation of Library by Gel Analysis

3ul of library was run on a 4-20% TBE PAGE gel (Invitrogen #EC62252) and stained in Vistra green stain according to manufacturer's instructions (Amersham, RPN5786). The stained DNA was visualised using a Typhoon scanner and is shown below. The library was amplified (visualised as a smear from 300-700bp) and no adaptor dimer (forked adaptor that has ligated to itself) was visible by eye using PCR primers 1 and 2 (from above), which have a phosphorothioate bond and 2'-OMe modification. Using the other 2 sets of primers (unmodified and modified with 3' phosphorothioate), the genomic library was amplified, but so was the adaptor dimer. The phosphorothioate bond and 2'-OMe modification of the primers prevents adaptor dimer amplification with DeepVent polymerase.

Figure 2 shows a 4-20% TBE PAGE gel (Invitrogen, EC62252) of 3 µl of purified library, stained in Vistra green stain according to manufacturer's instructions (Amersham, RPN5786).

M) Marker lane

a) Amplification of library using unmodified PCR primers (AATGATACGGCGACCACCGACACTCTTTCCCTACACGACGCTCTTCCGATCT SEQ ID NO:7 and CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT SEQ ID NO:8).

b) Amplification of library using PCR primers, which have a phosphorothioate bond modification (AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG

IDT Exhibit 1020.3559

ATCxT SEQ ID NO:9 and
CAAGCAGAAGACGGCATACGAGCTCTTCCGATCxT SEQ ID NO:10, where x is a phosphorothioate bond).

c) Amplification of library using PCR primers, which have a phosphorothioate bond and 2'-OMe modification (see PCR amplification method above).

Validation of Library by Conventional Sanger Sequencing

4 µl of the library was cloned into a plasmid vector (Zero Blunt TOPO PCR cloning kit, Invitrogen #K2800-20) and plated out on agar, according to manufacturer's instructions. Colonies were picked, mini-prepped and the cloned inserts sequenced by conventional Sanger sequencing.

These results confirm that the library preparation method produces a library of 'sequenceable' DNA templates containing a mixture of genomic fragments of different sequence, flanked by the two adaptors (AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT SEQ ID NO:7 and AGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG SEQ ID NO:11), required for cluster formation and SBS sequencing. The insert DNA from each of the clones sequenced was found to align to the human BAC reference sequence. The genomic DNA is amplified with low error rates by using Deep Vent®, a high fidelity polymerase. The adaptor sequences have no mutations/deletions at adaptor-genomic DNA boundaries, which should allow the hybridisation of the SBS sequencing primer and successful SBS sequencing of inserts. None of the clones were adaptor dimers (lacked genomic insert).

Conventional Sanger sequencing of 20 clones from library.

Clone 1 SEQ ID NO:16
CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTGCTGGAGGGAAATCCAAACTCAGGGGTGCCTGTGCCACAGCAAACACTCTCCCTCT
CACACCACCTGGAATAGAGATCAGCTAGAGCAGAGGCTGCTAAGAGAGGGAACAGAGGCTCCTTGTGACAGGGAGACTAGGATCAGAAGT
CAGGGAAGGGACAGCCGGGTGAAATGACTGGAAAGAGGAGCAATCACTCAGCAGTAAGGCAGGTTCTTCCAAAGACAAAAGGACACAGAG

IDT Exhibit 1020.3560

ATAAGTCAGGGCACTTCCAAGGAACCCAACTACCTACTCCACACTCCCAAATTTATTCTGGGTTGGGCCCTTTTTGGTTCCAATATCACC
TCGGATACCATAACTTGTCCAAGGTCTCTTCTTACCTCTCCCACCCTAAATGAAGACGGGCCCAGATCGGAAGAGCGTCGTGTAGGGAAA
GAGTGTAGATCTCGGTGGTCGCCGTATCATT

Clone 2 SEQ ID NO:17
ATTTAAATTTGGGGAACCATTGATGGATATAAGTGGTATTTAAAGCCACAGGATTAGGCTGGGCACAGTGGCTCATGCCTATAATCCCAG
CCCTTTGGGAGGCTGAGGCAGGTGGATCACTTGAGGCCAGGAGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACCAAAA
AATACAGAAAATTAGCCGTGTGTGGTGGTGCGTGCCTGTAGTCCCAGATACTCAGGAGGGTGAGGCAGGAGAATTGCTTGAATCAGATCG
GAAGAGCTCGTATGCCGTCTTCTGCTTG

Clone 3 SEQ ID NO:18
CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTGGCCCCGCCAGCTCTTGTGGTTGCCTGGAAGCATTATTTCCATCAACACTATCCCC
AGCTCCACGTCGTCCTTTTCACCTCTTTTCCTCGGGACCCCCGCACCCCACAGGATCCTAGTAGTGGTGAGTGGGCAATGAGAGAGGGCA
ACTTGGGAGAGGTGAGTTGGCAGGGGACAAAGGGGAGAACAGAGAGGCTTATTGACAAGGGGGCACCTGGTCTTGGGCCTAAGGGTGGTG
GGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT

Clone 4 SEQ ID NO:19
CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTTTTGCCATTTCCACAGCAGTATGTCCCATCCCTAGAATATCTGGCACCTGGTTAAG
TGTTCAGTACATATTTGTTGAATGGGTAAATGAATGAGAGCTGGAGGGAAATCCAAACTCAGGGGTGCCTGTGCCACAGCAAACACTCTC
CCTCTCACACCACCTGGAATAGAGATCAGCTAGAGCAGAGGCTGCTAAGAGAGGGAACAGAGGCTCCTTGTGACAGGGAGACTAGGATCA
GAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT

Clone 5 SEQ ID NO:20
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGCACAATTTCAGTTCTATCAAGTTTGAAAGCT
GCAAAATTAAGCTGCATTGTTGAGAGATACACAGGTCATAAACTAAAGAGAAATGTAAGGAGTTGATTTCCTTAAAAGGATAAAGCTTGC
ATGTATAGAGGGAGAGGATTATGATCAAGAAGGATGAGTGGCGGCCGGTATGGTGGCTCATCCCTGTAATCCTAGCACTTTGGGAGGCTG
AGGCAGGCGCATTACTTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAACCCTATCTCTACTAAAAATACAAAAAGTTAGC
CAGGTGTGGAGCCGCACGCCTGTGGTCTCAGCTACTCAGGAGGTTGAGGCACGAGAATCGCTTGAACCTGGGAGGATGAGGTTGTAGTGA
GCCAACATCGCACCACTGCACTCCAGCCAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG

Clone 6 SEQ ID NO:21
GATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCCACCCTGCCCTGCAAAACCACCAGCTCCGTGGTC
TCTGGATGGGACTCCCAGGTGCCTGGGGAACCAAAACAAGAAAAAAATGGAGGAGAGTTTTGAGCAAGAACTAAAGCCAAGGAAAGATGG
GGAAGAGGCAAAGACTAGGAATAACAATAATCTTTAGAGCTGCTGGCATTCATTCATTCATCCAGATCGGAAGAGCTCGTATGCCGTCTT
CTGCTTG

Clone 7 SEQ ID NO:22
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTAGAAAATGAGATCGCTGATGGCTTGCAGTTGG
CACGTGAAGGAAAGTGAGGATAAAGAAGGACTTCCAGGTTTTTGTCTTGAGCAACTGGAAATACTGAGATGGGAAGACTGGGGGAGAAGC
AGATTTGAAGGCTTTGGGGAGGAGAGGGGAATCAGAAATGAAATTTCAGATTCTTTTTAAATATCCCTAGTGGAGATGTTGAATAGGCAG
TGGGTAAGTAGTCAGCAGCTTAGGGGAGAGAAGAGGACGGAAATTTGAATTTGGGAAAGATTTAAATTTGGGGAACCATTGATGGATATA
AGTGGTATTTAAAGCCACAGGATTAGGCTGGGCACAGTGGCTCATGCCTATAATCCCAGCCCTTTGGGAGGCTGAGGCAGGTGGATCACT
TGAGGCCAGGAGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACCAAAAAATACAGAAAATTAGCCGTGTGTGGTGGTGC
GTGCCTGTAGTCCCAGATACTCAGGAGGGTGAGGCAGGAGAATTGCTTGAATCCTGGAGGCGGAGGTTACGGTGAGCCAAGATCATACCA
CTGCACTCAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG

Clone 8 SEQ ID NO:23
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGGTGAAAAAAGTAGGTAAGTCTTTTAAGGAGT
TTTCCTGCAAAGCGGACAGAGAACTAGGTCTGGTGGTGGTCGTCAAAGGAGAACTTTTTTCTCCCTCCTTCCCTCCCTTCCTCCTAGATA
CTTGCATGTGTTGTAAAGAGTGAGCACAGTGGTACATCTTTTCTTAGCCACAGTCAGCTGCCCAGGAGCAGTGCTGAATGGGCAGAGCTG
GATTTTACAGGGTTGGGATTTTGCCAGGTGAGTAAGATAGAGGGGAGAAGTGGGACCAGGGAGTTCCAGGTCTGTGAACGGCCCTGGCTG
AGGAGCTGGATCATGAAATCTGAGTCAAGTAAGAAGGAAATTGAGGACACGAGTTGGGTATTGCATAGTGTTACTGTGTTAAGGTCAGGG
GTCAAAGACTTACTGGCATGGAGTAACCAGAGTAAGTGAGCTGGAAAGATGAGTTGTCAGATCGGAAGAGCTCGTATGCCGTCTTCTGCT
TG

Clone 9 SEQ ID NO:24
CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTCAGGGAGGAGGTAGAGAGGAGAAAGAGAGCAGCCCGGGAGCAAGTTCTACAGCCGG
TCAGTGCTGAGTTGTTGGAGCTGGACATCCGGGAGGTGTATCAGCCTGGCTCAGGTGAGTGAGAGCAAGACAGGCATTGGGCTGGGGAAG
GAGTTTGGAAAGGTAAAAGCCGACTGTGAGGAAGGAGGGGTCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTC
GCCGTATCATT

Clone 10 SEQ ID NO:25
CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTTGGAACAGCCGCTCTCACCTCAGTTCATCTGGGGAAGGGGCTACAAAGCAAACAAT
CTTTATTCACAATTGGGGTGGCAGAGGGGAGATACCCCCAGGTCAGTCCAAAAGCAAAGATACTGGGAGGGAAGATGGCGCTGGGCGAGG
AACTCAGCACTCATCCTCACCCAGCAGGGCATAAGGGTTTCGGCCAGCCAGGCTGGACCCTGGAGCCGAGGTTGGGGTCTCCTCATCCCC
TTCTCCCTCCTCATCCGCATCCCGGTCCTCCTCTCCCTCCTCCTCACAGGAGCTGCTCAGCTCTTCCTCTTCCTCCACCTCCTCGTCACC
TGCTGGCCCCACCCTGCCCTGCAAAACCACCAGCTCCGTGGTCTCTGGATGGGACTCCCAGGTGCCTGGGGAACCAAAACAAGAAAAAAA
TGGAGGAGAGTTTTGAGCAAGAACTAAAGCCAAGGAAAGATGGGGAAGAGGCAAAGACTAGGAATAACAATAATCTTTAGAGCTGCTGGC
ATTCATTCATTCATCCATTCATTCAACTTCCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT

IDT Exhibit 1020.3561

Clone 11 SEQ ID NO:26
CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTTGGAGTGCAGTGGCACAATCATAGCTCACTGCAGCCTTGAACTCTTGGGCTCAAGT
GATCCTCCTGCCTCAGCCTCTGAAGTAGCAGAGACTACAGGCACATACCACCACACTTGGCTAGTTTTCTTTATCTTTTGTAAAGATGGG
GTTTCACTATGTTGCCCACACTAGTCTTGAGCTCCTGGTCTCAAGCAATCCTCCCACCTCAGCCTCCCAAAGCGCTGGGACTATATAGGC
ATGAGCCCTCACACATGGCCGTCATCCATTCTTTTACTCAGGTATCAATGTCCTTATTTTTAAAATCAAAGTAACTAAGACTCAGAGTAG
CAAAATCACTTACTCAAGACCTCACAGCTGAGAAGAGGTGGAATTTAACTCAGGCTGTCATGATCCTTCCACTGCAGCAGACGCCTCTTC
TGCCTTGCCCACCGCCACTGGCAGAGATCACCCCTCAGACACCCTGGGGCCTAATGAGACCTGATCGAGATCGGAAGAGCGTCGTGTAGG
GAAAGAGTGTAGACTCGGTGGTCGCCGTATCATT

Clone 12 SEQ ID NO:27
CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTCTCGCCTTTTACCCTTCCCTGTCTCCTTTCCTCTCCTCAGAGATTCTGCCATTGAT
GAAGCTGCTGTCCACTCACCCCTAGATTTTTGTTGCATAGCCAGTAAGATCTCTGGCCTGGAATGTTTTGGGGAAAACTGGAAGGACTGT
GTTATGGGAGTGGGAGTATAACTGGTACCTGGTTAATGCTTTCCCTTTCCATTATTCTTTTCTTCCTCCAGCCTGGGCAGAGAAACGTGG
TTACAAGACAGCCAAGGCGGCTCGGAATGATGTGTACAGAGCAGCCAACAGTCTCTTGCGGCTGGCAGTGGACGGCCGCCTCAGCCTGTG
TTTTCATCCCCCAGGCTACAGTGAACAGAAAGGTCAGAGCCCAGATATTCTTCCCCAGCCCCCTGCTATAGCGTAGGTAAAAGGGTGTGG
GCTTTGTGGCCAGAGAGAGGGTCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT

Clone 13 SEQ ID NO:28
CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTAGGAAAGTGAGGATAAAGAAGGACTTCCAGGTTTTTGTCTTGAGCAACTGGAAATA
CTGAGATGGGAAGACTGGGGGAGAAGCAGATTTGAAGGCTTTGGGGAGGAGAGGGGAATCAGAAATGAAATTTCAGATTCTTTTTAAATA
TCCCTAGTGGAGATGTTGAATAGGCAGTGGGTAAGTAGTCAGCAGCTTAGGGGAGAGAAGAGGACGGAAATTTGAATTTGGGAAAGATTT
AAATTTGGGGAACCATTGATGGATATAAGTGGTATTTAAAGCCACAGGATTAGGCTGGGCACAGTGGCTCATGCCTATAATCCCAGCCCT
TTGGGAGGCTGAGGCAGGTGGATCACTTGAGGCCAGGAGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACCAAAAAATA
CAGAAAATTAGCCGTGTGTGGTGATGCGTGCCTGTAGTCCCAGATACTCAGGAGGGTGAGGCAGGAGATCGGAAGAGCGTCGTGTAGGGA
AAGAGTGTAGATCTCGGTGGTCGCCGTATCATT

Clone 14 SEQ ID NO:29
CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTGGGCTGGTGGGGCGGAAAGTCGTGAGTGTCTCCAGAACCCCGGGCCATACCCGATA
CTTTCAGACCTACTTTCTTACCCCCTCTGTGAAGCTCTGTGACTGCCCAGGCCTCATCTTCCCATCTCTTCTGCCTAGGCAGTTGCAGGT
ATGACGGGGAGGGTGGGTAAGGGAAAGAGAGAAGGTGGGACATTGAGGAAAGTACTGAGTGCTCATTTCCCTCAGATCGGAAGAGCGTCG
TGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT

Clone 15 SEQ ID NO:30
CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTCACGCCTGTGGTCTCAGCTACTCAGGAGGTTGAGGCACGAGAATCGCTTGAACCTG
GGAGGATGAGGTTGTAGTGAGCCAACATCGCACCACTGCACTCCAGCCCGGGTGAGGGAGTGAGACTCTGTCTCAAAAACAAAAACAAAA
AAAAACAAGGACAGATGGAACATGTTGTCACACATTGGGTGGTATGGGGTTCATCAGGCTGCACATGTATGTTCTGTGCATTTTTTTGTA
TGTTGTAGTTTACAGTTACAAAGAAGATAGCAGGAAGAAATGGTGAAAAAAGTAGGTAAGTCTTTTAAGGAGTTTTCCTGCAAAGCGGAG
ATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT

Clone 16 SEQ ID NO:31
GATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTGGAACAGCCGCTCTCACCTCAGTTCATCTGGGGA
AGGGGCTACAAAGCAAACAATCTTTATTCACAATTGGGGTGGCAGAGGGGAGATACCCCCAGGTCAGTCCAAAAGCAAAGATACTGGGAG
GGAAGATGGCGCTGGGCGAGGAACTCAGCACTCATCCTCACCCAGCAGGGCATAAGGGTTTCGGCCAGCCAGGCTGGACCCTGGAGCCGA
GGTTGGGGTCTCCTCATCCCCTTCTCCCTCCTCATCCGCATCCCGGTCCTCCTCTCCCTCCTCCTCACAGGAGCTGCTCAGCTCTTCCTC
TTCCTCCTCCTCCTCGTCACCTGCTGGCCCCACCCTGCCCTGCAAAACCACCAGCTCCGTGGTCTCTGGATGGGACTCCCAGGTGCCTGG
GGAACCAAAACAAGAAAAAAATGGAGGAGAGTTTTGAGCAAGAACTAAAGCCAAGGAAAGATGGAGATCGGAAGAGCTCGTATGCCGTCT
TCTGCTTG

Clone 17 SEQ ID NO:32
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTAGAGAAAGGAAATAGTTGGAGAAGAGGAGGTC
AAGGACTGAGCCTTAGGACAAGCCAGCATTTAGTTGGGCAAAGGACAGTGAGAAGGAGGAAAACCAAGGGAGCGTCCCAGAAGTCAAATG
AAGAAGGTGTTTGAAGAGGAAAGGAGGAATCAGCTGTGTCAACTGTTGCTGACAGGCCAAATGAGAGAACAGAGAGCTGCTCAGCAGGCT
TGGCAATGTGAAGATCCGTGGTTTCAGTGGGGTGGAGAAAGCCAAACTGGAGTAGGCCCATGAGAGAAGGTGCAACAACTTCACATAACA
TTGTGTGAAAAGAGTCTGACCCAATAGCATCCATACTGCACAATTTCAGTTCTATCAAGTTTGAAAGCTGCAAAATTAAGCTGCATTGTT
GAGAGATACACAGGTCATAAACTAAAGAGAAATGTAAGGAGTTGATTTCCTTAAAAGGATAAAGCTTGCATGTATAGAGGGAGAGGATTA
TGATCAAGAAGGATGAGTGGCGGCCGGTATGGTGGCTCATCCCTGTAATCCTAGCACTTTGGGAGGCTGAGGCAGGCGCAAGATCGGAAG
AGCTCGTATGCCGTCTTCTGCTTG

Clone 18 SEQ ID NO:33
CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTGGGATGAGGTCAGCAAGATCACCAGGGCCCACATCATGTAGAGCCCTGAGGCTGTG
ACAGCCATTTTGGATTTTATTCCAAGTCTCATGAGAAGCCAAGGGTAGGTTTTGAACAGGGGAATGATAGGATCTGATTTTGTTTCTTAT
AAGTTTACCTCCTGAGTAGAGAATAAATGATGGGGGGTGGGCAAGAAAGGGAGCAGAGAGAGCAGTTGAGGCTATTTCAGTAATCTAGGA
GAGAAATAAGAGTTGCTTAGAGTAGGATGCTGGAGCTGGAGGTGGTGAGACAGGGCCGGAAGCATGATATATTTTGAAGGTAGAGAAAAT
GAGATCGCTGATGGCTTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT

Clone 19 SEQ ID NO:34
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGGGAGGACTGAACATCGGAAGAAGTTGAGTGG
GATGAAAGGGAGTCAGGCCGATGGGGGAGAGGGTCTTATGGCCCCTGAGAGCTGGCCAGCACTGGCGTCACCGGCCCCTCCCCGCAGGGC
TTCAAGATGGGCTGCGCTCCAGTTCCAACAGCCGCAGCGGGAGCCGGGAGCGGCGAGAGGAACAGACCGACACCTCGGACGGGGAGTCTG
TGACCCATCATATCCGCAGGCTTAACCAGCAGCCTTCTCAGGGGCTGGGTCCACGAGGCTACGACCCAAATCGGTGAGGGTGGGAGGGGG
CGCTGGTCCCGGCTTTCCCGCCTACCCGGAAGTCAGAGCTTTGGGGAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG

IDT Exhibit 1020.3562

Clone 20 SEQ ID NO:35
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCGACCCAAATCGGTGAGGGTGGGAGGGGGCGC
TGGTCCCGGCTTTCCCGCCTACCCGGAAGTCAGAGCTTTGGGGGAAAGCGGGCTGCTACTGGTGAAGACGGTGGGCCTGGGATGCCACAG
TTCTCCGCTAGCCACTCGGCTCCCCACAGCGGGCCACAGTCTTCCTTTCCAGAGGGGCTGGAGAGAGTTGGGCTTTTAGAAGGAGAAGGC
TGAGTATTGCCTGAAAGAAGGACTTGGGGGAAGTCTGACTTGAGAGAGGAGACTTGAACGACTCTGGAGAGAATGGTTTGCGAAGTTGAT
TGTACAAGAGGGGAGAAAATAGGAGTTTGTGGCCACAGGATTGCTCTGGATGTCTCGGTCCCTGTTCCCTTAGATACCGACTGCATTTTG
AGAGAGACAGCAGGGAGGAGGTAGAGAGGAGAAAGAGAGCAGCCCGGGAGCAAGTTCAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTT
G

Example 2

The DNA source used was a purified 10kb region of Human BAC. The DNA was first prepared for ligation to forked adaptors by: fragmentation of the DNA by nebulisation, end repair of the DNA ends to make them blunt-ended and phosphorylation, 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). 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 adaptor at both ends of the fragments.

Materials and Methods

DNA Sample Preparation

Amplified 10kb region of human BAC DNA (140K human chromosome 6 insert in a pTARBAC vector) using two primers (LRPCR primer 3: CGAGGAACTCAGCACTCATC SEQ ID NO:12 and LRPCR primer 4: ATGCCGAGGAAGAAGCCATT SEQ ID NO:13). Used 0.3 µM each primer, 500 µM each dNTP, 0.08 ng/µl BAC DNA, 0.1U/µl pfu ultra hotstart polymerase (Stratagene #600384) diluted in 1X pfu ultra hotstart buffer (Stratagene #600380).

PCR Programme :

92°C 2 mins

IDT Exhibit 1020.3563

92°C 10 secs
58°C 30 secs 10 cycles
68°C 22 mins
92°C 10 secs
58°C 30 secs 25 cycles
68°C 24 mins
Hold at 4°C

Gel purified 10kb product on 0.7% TAE agarose gel, using Qiagen gel extraction kit (Qiagen #28706), according to manufacturer's instructions.

The purified PCR product was 3' tailed with dCTP, using the following: 1.5 µg of purified 10kb product, 800µM dCTP, 250µM $CoCl_2$, (NEB #B0252S), 0.2 U/µl of Tdt (NEB #M0252L), diluted in 1x NEB4 buffer (NEB # B7004S), in a total volume of 50µl. After a 30 minute incubation at 37°C, the reaction was heat inactivated at 70°C for 10 mins. The tailed PCR product was purified from enzymes, buffer, etc. by loading the reaction mix on a Qiagen column (Qiagen #28104), and eluting in 80µl EB (Qiagen).

Nebulization

Materials:

- 10kb region of human BAC, dCTP tailed (from above)
- Buffer [glycerol 53.1 ml, water 42.1 ml, 1 M TrisHCl (pH7.5) 3.7 ml, 0.5 M EDTA 1.1 ml]
- Nebulizer Invitrogen (#K7025-05)
- Qiagen columns PCR purification kit #28104

Mix: 70µl (1.1 micrograms) of DNA
700µl Buffer

IDT Exhibit 1020.3564

Procedure:

Chilled DNA solution was fragmented in the nebulizer on ice for 6 minutes under 32psi of pressure. The recovered volume was purified with a Qiagen PCR purification kit column and eluted in 30µl of EB (Qiagen).

End-repair

Materials:

- T4 DNA Polymerase NEB #M0203S
- 10xNEB 2 buffer NEB #B7002S
- 100x BSA NEB #B9001S
- dNTPs mix (10mM each) NEB #N0447S
- E. coli DNA Pol I large fragment (Klenow) NEB #M0210S
- T4 polynucleotide kinase NEB #M0201L
- T4 PNK buffer NEB #B0201S
- 100 mM ATP Amersham #27-2056-01
- Qiagen columns PCR purification kit #28104

End repair mix was assembled as follows:

DNA	30µl
Water	7.5µl
10xNEB2	5µl
100xBSA	0.5µl
10mM dNTPs	2µl
T4 DNA pol (3U/µl)	5µl
	50µl total

The reaction was incubated for 15 min at room temperature, then 1µl of E. coli DNA Pol I large fragment (Klenow) was added and the reaction incubated for a further

IDT Exhibit 1020.3565

15 min at room temperature. The DNA was purified from enzymes, buffer, etc. by loading the reaction mix on a Qiagen column, and eluting in 30µl EB (Qiagen). The 5′ ends of the DNA were then phosphorylated using polynucleotide kinase as follows:

DNA	30µl
Water	9.5µl
10xPNK buffer	5µl
100mM ATP	0.5µl
T4 PNK (10U/µl)	5µl
	50µl total

The reaction was incubated for 30 min at 37°C, then heat inactivated at 65°C for 20 min. DNA was then purified from enzymes, buffer, etc. by loading the reaction mix on a Qiagen column, and eluting in 30µl EB (Qiagen).

A- tailing reaction

Materials:

•	Taq DNA polymerase	NEB #M0267S
•	10x thermopol buffer	NEB #B9004S
•	1mM dATP	Amersham-Pharmacia #272050
•	Qiagen MinElute column	PCR purification kit #28004

The following reaction mix was assembled:

DNA	30µl
10x thermopol buffer	5µl
1mM dATP	10µl
Taq pol (5U/µl)	3µl
	~50µl total

The reaction was incubated for 30 min at 70°C, then the DNA purified from enzymes, buffer, etc. by loading the

IDT Exhibit 1020.3566

reaction mix on a Qiagen MinElute column, and eluting in 10µl EB (Qiagen).

Anneal forked adaptor

Materials:

- Oligo A: 5'ACACTCTTTCCCTACACGACGCTCTTCCGATCxT SEQ ID NO:3 (x = phosphorothioate bond)
- Oligo B: 5'Phosphate-GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG SEQ ID NO:4
- 50mM Tris/50mM NaCl pH7
- PCR machine

100µM Oligo A	20µl
100µm Oligo B	20µl
Tris/NaCl	10µl
	50µl at 40µM duplex in 10mM Tris/10mM NaCl (pH 7.5)

The adaptor strands were annealed in a PCR machine programmed as 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
Forked adaptor was diluted to 15 µM final concentration in 10mM Tris/10mM NaCl pH 7.5.

Ligation reaction

Materials:

- 15µM forked adaptor (from above)
- A-tailed genomic DNA (from above)
- Quick Ligase NEB #M2200S

IDT Exhibit 1020.3567

- Quick Ligase 2x buffer NEB #B2200S
- PCR machine
- Qiagen columns PCR purification kit #28104

Reaction mix was assembled as follows:

DNA	10µl
2x buffer	25µl
15 µM adaptor	2µl
Water	8ul
Quick Ligase	5µl
	50µl total

The reaction was incubated for 30 min at room temperature then the DNA purified from enzymes, buffer, etc. by loading the reaction mix on a Qiagen column, and eluting in 30µl EB (Qiagen).

Gel purification

Materials:

- Agarose Promega #V3841
- 100 base pair ladder NEB #N3231L
- TAE
- Loading buffer (50 mM Tris (pH8), 40 mM EDTA, 40% w/v sucrose)
- Ethidium bromide
- Gel trays and tank. Electrophoresis unit
- Qiagen gel extraction kit Qiagen #28706

The entire sample from the purified ligation reaction was loaded into one lane of a 4% agarose gel containing ethidium bromide and run at 120V for 90 min. The gel was then viewed on a 'White-light' box and fragments from 120bp

IDT Exhibit 1020.3568

to 170bp excised and purified with a Qiagen column, eluting in 30µl EB (Qiagen).

Exonuclease I Treatment of PCR Primers

Exonuclease I (E. coli) NEB M0293S 20,000 Units/ml

Exonuclease I storage conditions:

100 mM NaCl, 10 mM Tris-HCl (pH 7.5), 0.5 mM EDTA, 5 mM 2-mercaptoethanol, 100 µg/ml BSA and 50 % glycerol

1 x Exonculease Reaction Buffer

67 mM Glycine-KOH

6.7 mM $MgCl_2$

10 mM 2-mercaptoethanol

(pH 9.5 @ 25 °C)

Protocol:

DNA Primers with a phosphorothioate at the n-1 position (5 x 85µl of each Primer (approx 25 µM) were aliquoted into eppendorf tubes. 10 µl of 10 x Exonuclease I Reaction Buffer and 5 µl of Exonuclease I was added to each tube. 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. Then the solutions from the eppendorfs were passed through P6 Bio Rad columns and spun in a centrifuge at 2000 rpm for 2 minutes. An extra 20 µl of H_20 was added 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 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

IDT Exhibit 1020.3569

pellets were dissolved in 50 µl of water, pooled and the UV measurement taken to determine the concentration of the oligonucleotide.

PCR amplification

Materials:

- Gel purified ligated DNA (from above)
- PCR PRIMER 3: exonuclease I treated 5′ AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCxT, where x= phosphorothioate bond SEQ ID NO:9
- PCR PRIMER 4: exonuclease I treated 5′ CAAGCAGAAGACGGCATACGAGCTCTTCCGATCxT, where x= phosphorothioate bond SEQ ID NO:10
- 2X Phusion™ PCR mix NEB/Finnzymes #F531
- PCR machine
- Qiagen MinElute column PCR purification kit #28004

The PCR reaction was prepared as follows:

DNA	0.5µl
2x Phusion™ PCR mix	12.5µl
8 µM PCR primer 3	1.5µl
25 µM PCR primer 4	0.5µl
Water	10µl
	25µl total

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

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

IDT Exhibit 1020.3570

PCR products were purified from enzymes, buffer, etc. on a Qiagen minelute column, eluting in 10µl EB (Qiagen). The resulting DNA library was then ready for amplification on a surface amplification platform.

Validation of Library by Gel Anaylsis

20% of total library was run on a 4-20% TBE PAGE gel (Invitrogen #EC62252) and stained in Vistra green stain according to manufacturer's instructions (Amersham, RPN5786). The stained DNA was visualised using a Typhoon scanner and shown below.

The library was amplified (visualised as a smear from 150-200bp) and no adaptor dimer (forked adaptor that has ligated to itself) was visible by eye using PCR primers 3 and 4 (from above), which have a phosphorothioate modification and were exonuclease I treated. Using the other 3 sets of primers (unmodified and modified with 3' phosphorothioate, but not exonuclease treated), the genomic library was amplified, but so was the adaptor dimer. The exonuclease I treated primer with 3' phosphorothioate modification prevents adaptor dimer amplification with Phusion™ polymerase

Figure 3 shows a 4-20% TBE PAGE gel (Invitrogen, EC62252) of 20% of a total library, stained in Vistra green stain according to manufacturer's instructions (Amersham, RPN5786).

M) Marker lane

a) Amplification of library using unmodified PCR primers (AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGA SEQ ID NO:14 and CAAGCAGAAGACGGCATACGA SEQ ID NO:15).

IDT Exhibit 1020.3571

b) Amplification of library using unmodified PCR primers (AATGATACGGCGACCACCGACACTCTTTCCCTACACGACGCTCTTCCGATCT SEQ ID NO:7 and CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT SEQ ID NO:8).

c) Amplification of library using PCR primers, which have a phosphorothioate bond modification (AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCxT SEQ ID NO:9 and CAAGCAGAAGACGGCATACGAGCTCTTCCGATCxT SEQ ID NO:10, where x is a phosphorothioate bond).

d) Amplification of library using exonuclease I treated PCR primers, which have a phosphorothioate modification (see PCR amplification method above).

Validation of Library by Conventional Sanger Sequencing

4 µl of the library was cloned into a plasmid vector (Zero Blunt TOPO PCR cloning kit, Invitrogen #K2800-20) and plated out on agar, according to the manufacturer's instructions. Colonies were picked, mini-prepped and the cloned inserts sequenced by conventional Sanger sequencing.

These results confirm that the library preparation method produces a library of 'sequenceable' DNA templates containing a mixture of genomic fragments of different sequence, flanked by the two adaptors (AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT SEQ ID NO:7 and AGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG SEQ ID NO:11), required for cluster formation and SBS sequencing. The insert DNA from each of the clones sequenced was found to align to the human BAC reference sequence. The genomic DNA is amplified with low error rates by using Phusion™, a high fidelity polymerase. The adaptor sequences have no mutations/deletions at adaptor-genomic DNA boundaries, which should allow the hybridisation of the SBS sequencing primer and successful

IDT Exhibit 1020.3572

SBS sequencing of inserts. None of the clones were adaptor dimers (lacked genomic insert).

Conventional Sanger sequencing of 20 clones from library.

Clone 1 SEQ ID NO:36
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGCCGTGTGTGGTGGTGCGTGCCTGTAGTCCCA
GATACTCAGGAGGGTGAGGCAGGAGAATTGCTTGAATCCTGGAGGCGGAGGTTACGGTGAGCCAAGATCATACCACTGCACTCCAGCCTG
GGTGACAGAGCAAGACTTCGTCAAAAAAAAAAAAAAAAAAGCCATGGAATTGGATGAGATGTAGGAGAGAAAGGAAATAGTTGGAGAAGA
GGAGGTCAAGGACTGAGCCTTAGGACAAGCCAGCATTTAGTTGGGCAAAGGACAGTGAGAAGGAGGAAAACCAAGGGAGCGTCCCAGAAG
TCAAATGAAGAAGGTGTTTGAAGAGGAAAGGAGGAATCAGCTGTGTCAACTGTTGCTGACAGGCCAAATGAGAGAACAGAGAGCTGCTCA
GCAGGCTTGGCAATGTGAAGATCCGTGGTTTCAGTGGGGTGGAGAAAGCCAAACTGGAGTAGGCCCATGAGAGAAGGTGCAACAACTTCA
CATAACATTGTGTGAAAAGAGTCTGACCCAATAGCATCCATACTGCACAATTTCAGTTCTATCAAGTTTGAAAGCTGCAAAATTAAGCTG
CATTGTTGAGAGATACACAGGTCATAAACTAAAGAGAAATGTAAGGAGTTGATTTCCTTAAAAGGATAAAGCTTGCATGTATAGAGGGAG
AGGATTATGATCAAGAAGGATGAGTGGCGGCCGGTATGGTGGCTCATCCCTGTAATCCTAGATCGGAAGAGCTCGTATGCCGTCTTCTGC
TTG

Clone 2 SEQ ID NO:37
CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTGTGGTGAGTGAGCAGAGTGGCAGTGAGCCCCAAGGCTTCCAGTTTTCTCCTTGGAC
AAGTCATCGTAATTATTATTATTATTTTTTGAGATGGAGTTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCG
CCGTATCATT

Clone 3 SEQ ID NO:38
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTAAAACCTACCCTTGGCTTCTCATGAGACTTGG
AATAAAATCCAAAATGGCTGTCACAGCCTCAGGGCTCTACATAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG

Clone 4 SEQ ID NO:39
CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTGATAGGATCTGATTTTGTTTCTTATAAGTTTACCTCCTGAGTAGAGAATAAATGAT
GGGGGGTGGGCAAGAAAGGGAGCAGAGAGAGCAGTTGAGGCTATTTCAGTAATCTAGGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTG
TAGATCTCGGTGGTCGCCGTATCA

Clone 5 SEQ ID NO:40
CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTCCAGCAGGTGACGAGGAGGAGGAGGAAGAGGAAGAGCTGAGCAGCTCCAGATCGGA
AGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCA

Clone 6 SEQ ID NO:41
CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTGAGAGCAGTTGAGGCTATTTCAGTAATCTAGGAGAGAAATAAGAGTTGCTTAGAGT
AGGATGCTGGAGCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT

Clone 7 SEQ ID NO:42
CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTCTCCCACCCTCACCGATTTGGGTCGTAGCCTCGTGGACCCAGCCCCTGAGAAGGCT
GCTGGTTAAGCCTGCGGATATGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCA

Clone 8 SEQ ID NO:43
CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTAGGTGAGTGAGAGCAAGACAGGCATTGGGCTGGGGAAGGAGTTTGGAAAGGTAAAA
GCCGACTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT

Clone 9 SEQ ID NO:44
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTGCACTCCAGCCCGGGTGAGGGAGTGAGACTC
TGTCTCAAAAACAAAAACAAAAAAAAACAAGGACAGATGGAACATGTTGTCACACATTGGGTGGTATGGGGTTCATCAGGCTGCACATGT
ATGTTCTGTGCATTTTTTTGTATGTTGTAGTTTACAGTTACAAAGAAGATAGCAGGAAGAAATGGTGAAAAAAGTAGGTAAGTCTTTTAA
GGAGTTTTCCTGCAAAGCGGACAGAGAACTAGGTCTGGTGGTGGTCGTCAAAGGAGAACTTTTTTCTCCCTCCTTCCCTCCCTTCCTCCT
AGATACTTGCATGTGTTGTAAAGAGTGAGCACAGTGGTACATCTTTTCTTAGCCACAGTCAGCTGCCCAGGAGCAGTGCTGAATGGGCAG
AGCTGGATTTTACAGGGTTGGGATTTTGCCAGGTGAGTAAGATAGAGGGGAGAAGTGGGACCAGGGAGTTCCAGGTCTGTGAACGGCCCT
GGCTGAGGAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG

Clone 10 SEQ ID NO:45
CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTGAAGCTCAGAGAAACGTGCGTGATTCCAGGGAGGGTAGGGTCAAATGACTTTTGGG
AGATTCTCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT

Clone 11 SEQ ID NO:46
CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTGGAAATTCACAACCTAGGACAGAGTTGATAAGAGGATGGAGCAGTGAAAGTCAACC
CAGAGTTCTCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT

Clone 12 SEQ ID NO:47
CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTAATAAGAGTTGCTTAGAGTAGGATGCTGGAGCTGGAGGTGGTGAGACAGGGCCGGA
AGCATGATATATTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT

IDT Exhibit 1020.3573

Clone 13 SEQ ID NO:48
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTGTGTGGTGGTGCGTGCCTGTAGTCCCAGA
TACTCAGGAGGGTGAGGCAGGAGAATTGCTTGAATCCTAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG

Clone 14 SEQ ID NO:49
CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTAAAGCAATGAAAGGGGATAGAGAGTGACTGAGAGGAGAGGTGCTCAGGGAAAACCT
CCCTGAGGAGAGAACTGGATGATGAGATGAAGTGAGCCATTCAGAACTGTGGGGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGA
TCTCGGTGGTCGCCGTATCATT

Clone 15 SEQ ID NO:50
CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTTAAAAGGATAAAGCTTGCATGTATAGAGGGAGAGGATTATGATCAAGAAGGATGAG
TGGCGGCCGGTATGGTGGCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT

Clone 16 SEQ ID NO:51
CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTCTGAGATGGGAAGACTGGGGGAGAAGCAGATTTGAAGGCTTTGGGGAGGAGAGGGG
AATCAGAAATGAAATTTCAGATTCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT

Clone 17 SEQ ID NO:52
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGAGAGGTGAGTTGGCAGGGGACAAAGGGGAGA
ACAGAGAGGCTTATTGACAAGGGGGCACCTGGTCTTGGGCCTAAGAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG

Clone 18 SEQ ID NO:53
CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTCAAGAAAGGGAGCAGAGAGAGCAGTTGAGGCTATTTCAGTAATCTAGGAGAGAAAT
AAGAGTTGCTTAGAGTAGGATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCG

Clone 19 SEQ ID NO:54
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTAAAGCTTGCATGTATAGAGGGAGAGGATTATG
ATCAAGAAGGATGAGTGGCGGCCAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG

Clone 20 SEQ ID NO:55
CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTTCTACCAAAAAATACAGAAAATTAGCCGTGTGTGGTGGTGCGTGCCTGTAGTCCCA
GATACTCAGGAGGGTGAGGCAGGAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT

Validation of Libraries by SBS Sequencing

Libraries were validated by SBS sequencing. Whole genomic BAC libraries were made using the method detailed in example 2 and sequenced by SBS sequencing. The protocols for preparing clusters on hydrogel surfaces and performing SBS sequencing are detailed in application WO06064199

The adaptor dimer contamination was reduced >300 fold by using PCR primers 3 and 4 (from above), which have a phosphorothioate modification and were exonuclease I treated, compared to using the shorter unmodified primers. The exonuclease I treated primer with 3′ phosphorothioate modification prevents adaptor dimer amplification with Phusion™ polymerase.

SBS sequencing of genomic libraries amplified with

IDT Exhibit 1020.3574

a) unmodified shorter PCR primers (AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGA SEQ ID NO:34 and CAAGCAGAAGACGGCATACGA SEQ ID NO:35)

b) exonuclease I treated longer PCR primers, which have a phosphorothioate bond modification (AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCxT SEQ ID NO:9 and CAAGCAGAAGACGGCATACGAGCTCTTCCGATCxT SEQ ID NO:10, where x is a phosphorothioate bond).

Percentage of adaptor dimers in genomic libraries is reported.

Library	Adaptor Dimer Contamination
a)	9.95% +/- 1.45
b)	0.03% +/- 0.02

While certain of the preferred embodiments of the present invention have been described and specifically exemplified above, it is not intended that the invention be limited to such embodiments. Various modifications may be made thereto without departing from the scope and spirit of the present invention, as set forth in the following claims.

IDT Exhibit 1020.3575

Claims:

1. A method for generating a library of template polynucleotide molecules from one or more primary polynucleotide molecules, said method comprising:

(a) fragmenting said one or more primary polynucleotide molecules to produce target polynucleotide duplexes;

(b) ligating an adaptor polynucleotide construct to both ends of the target polynucleotide duplexes to generate combined ligated adaptor-target-adaptor sequences;

(c) preparing an amplification reaction comprising said combined ligated adaptor-target-adaptor sequences and at least two different primer oligonucleotides, wherein each of said at least two different primer oligonucleotides is complementary to at least a part of the adaptor polynucleotide construct sequence of the combined ligated adaptor-target-adaptor sequences, and complementary to at least one base of either the target polynucleotide duplexes, or an overhang attached to the target duplex;

(d) performing an amplification reaction wherein said at least two different primer oligonucleotides anneal to complementary parts of the adaptor-target-adaptor sequences and are extended by sequential addition of nucleotides to generate a plurality of amplification products complementary to at least one strand of the combined ligated adaptor-target-adaptor sequences, wherein each of said plurality of amplification products has a first common sequence at its 5′ end and a second common sequence at its 3′ end, and wherein

IDT Exhibit 1020.3576

said plurality of amplification products comprises a library of template polynucleotide molecules.

2. A method for generating a library of template polynucleotide molecules from one or more primary polynucleotide molecules, said method comprising:

(a) fragmenting said one or more primary polynucleotide molecules to produce target polynucleotide duplexes;

(b) ligating an adaptor polynucleotide construct to both ends of the target polynucleotide duplexes to generate combined ligated adaptor-target-adaptor sequences, wherein said adaptor polynucleotide construct comprises a nucleic acid strand comprising at least one overhanging base that is modified to render the nucleic acid strand resistant to exonucleolysis;

(c) preparing an amplification reaction comprising said combined ligated adaptor-target-adaptor sequences and at least two different primer oligonucleotides, wherein each of said at least two different primer oligonucleotides is complementary to at least a part of the adaptor polynucleotide construct sequence of the combined ligated adaptor-target-adaptor sequences;

(d) performing an amplification reaction, wherein said at least two different primer oligonucleotides anneal to complementary parts of the combined ligated adaptor-target-adaptor sequences and are extended by sequential addition of nucleotides to generate a plurality of amplification products complementary to at least one strand of the

IDT Exhibit 1020.3577

combined ligated adaptor-target-adaptor sequences, wherein each of said plurality of amplification products has a first common sequence at its 5′ end and a second common sequence at its 3′ end, and wherein said plurality of amplification products comprises a library of template polynucleotide molecules.

3. A method for generating a library of template polynucleotide molecules from one or more primary polynucleotide molecules, said method comprising:

(a) fragmenting said one or more primary polynucleotide molecules to produce target polynucleotide duplexes;

(b) ligating an adaptor polynucleotide construct to both ends of the target polynucleotide duplexes to generate combined ligated adaptor-target-adaptor sequences, wherein said adaptor polynucleotide comprises a nucleic acid strand comprising at least one overhanging base that is modified to render the nucleic acid strand resistant to exonucleolysis;

(c) preparing an amplification reaction comprising said combined ligated adaptor-target-adaptor sequences and at least two different primer oligonucleotides, wherein each of said at least two different primer oligonucleotides is complementary to at least a part of the adaptor polynucleotide construct sequence of the combined ligated adaptor-target-adaptor sequences, and complementary to at least one base of either the target polynucleotide duplexes, or an overhang attached to the target duplex;

IDT Exhibit 1020.3578

(d) performing an amplification reaction, wherein said at least two different primer oligonucleotides anneal to complementary parts of the adaptor-target-adaptor sequences and are extended by sequential addition of nucleotides to generate a plurality of amplification products complementary to at least one strand of the combined ligated adaptor-target-adaptor sequences, wherein each of said plurality of amplification products has a first common sequence at its 5′ end and a second common sequence at its 3′ end and wherein said plurality of amplification products comprises a library of template polynucleotide molecules.

4. The method of any one of claims 1 to 3, wherein fragmentation of the primary polynucleotide molecules is by mechanical fragmentation.

5. The method of claim 4, wherein said mechanical fragmentation of the primary polynucleotide molecules is by sonication, nebulisation or hydrodynamic shearing.

6. The method of any one of claims 1 to 3, wherein fragmentation of the primary polynucleotide molecules is by chemical or enzymatic fragmentation.

7. The method of any one of claims 1 to 6, wherein the one or more primary polynucleotide molecules are DNA molecules.

8. The method of claim 7, wherein the one or more primary polynucleotide molecules are genomic DNA molecules.

9. The method of claim 8, wherein the one or more primary polynucleotide molecules are human genomic DNA molecules.

IDT Exhibit 1020.3579

10. The method of any one of claims 1 to 3, wherein the adaptor polynucleotide construct has a single nucleotide overhang at either the 5′ or 3′ terminus of each duplex strand and the target polynucleotide duplexes have complementary single nucleotide overhangs at the 5′ or 3′ terminus of each duplex strand.

11. The method of claim 10, wherein the adaptor polynucleotide construct has a single nucleotide 'T' overhang at the 3′ terminus of each duplex strand and the target polynucleotide duplexes have complementary single nucleotide 'A' overhangs at the 3′ terminus of each duplex strand.

12. The method of claim 10 or 11, wherein the single nucleotide overhang on the target polynucleotide duplex is added using a DNA polymerase.

13. The method of claim 10, wherein the two primer oligonucleotides of claim 1(c) or claim 3(c) have a 3′ nucleotide that hybridises to the single nucleotide overhang attached to the target DNA.

14. The method of claim 11, wherein the two primer oligonucleotides of claim 1(c) or claim 3(c) have a 3′-'T' nucleotide that hybridises to the 'A' overhang attached to the target DNA.

15. The method of claim 13 or 14, wherein the two primer oligonucleotides have a modification that makes the primers resistant to exonucleolysis.

IDT Exhibit 1020.3580

16. The method of claim 2 or claim 15, wherein said exonuclease resistant modification is a phosphorothioate moiety.

17. The method of claim 2 or claim 15, wherein said exonuclease resistant modification is a methylphosphonate moiety.

18. The method of claim 2 or claim 15, wherein said exonuclease resistant modification is a 2'-O-methyl moiety.

19. The method of claim 2 or claim 15, wherein said exonuclease resistant modification is both a thiophosphate moiety and a 2'-O-methyl moiety.

20. The method of any one of claims 15-19, wherein the primer oligonucleotides are treated with an exonuclease prior to amplification.

21. The method of any one of claims 1 to 3, wherein the sequential addition of nucleotides is carried out by a DNA polymerase.

22. The method of claim 21, wherein said polymerase is a high fidelity DNA polymerase.

23. The method of claim 22, wherein said polymerase possesses exonuclease activity.

24. The method of claim 21, wherein said polymerase is Phusion™ DNA polymerase.

IDT Exhibit 1020.3581

25. The method of claim 1 or 3, wherein at least one of said primer oligonucleotides is longer than the ligated adaptor sequences in both the 3′ and 5′ directions.

26. The method of claim 2, wherein at least one of said primer oligonucleotides is longer than the ligated adaptor sequences in the 5′-direction.

27. The method of any preceding claim further comprising the steps of:

(e) preparing clusters from the library of template polynucleotide molecules;

(f) sequencing the clusters by sequencing by synthesis.

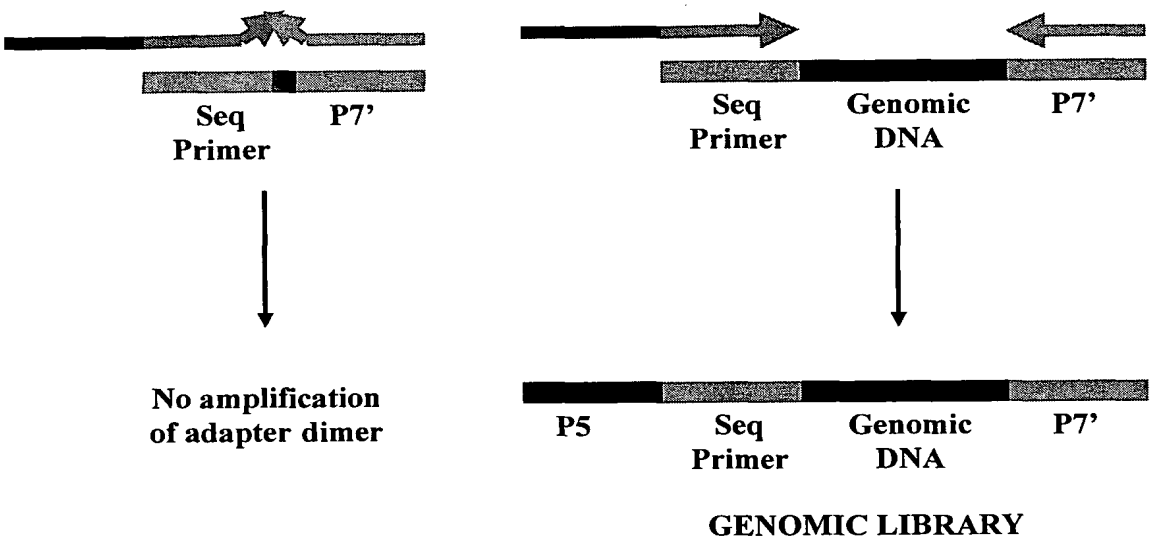
28. A library of template polynucleotide molecules prepared by the method of any one of claims 1 to 26.

29. An array comprising the library of template polynucleotide molecules of claim 28.

30. Use of the library of claim 28 or the array of claim 29 in sequencing.

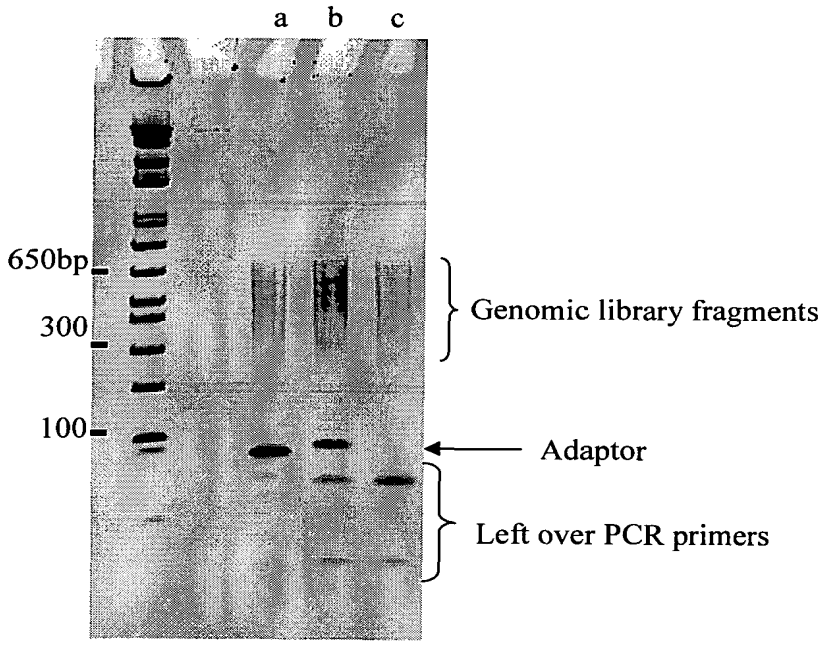
IDT Exhibit 1020.3582

FIG. 1



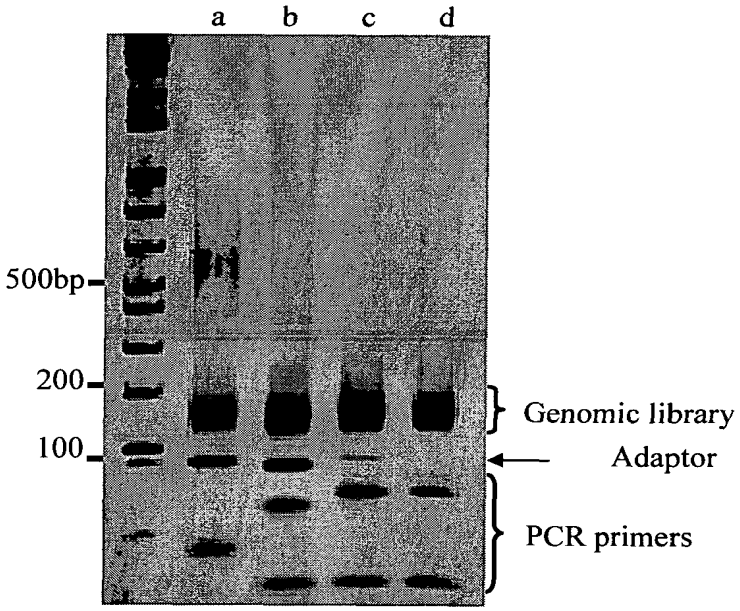
IDT Exhibit 1020.3583

FIG. 2



IDT Exhibit 1020.3584

FIG. 3



IDT Exhibit 1020.3585

Espacenet

Bibliographic data: WO2014150931 (A1) — 2014-09-25

METHODS, COMPOSITIONS AND KITS FOR GENERATION OF STRANDED RNA OR DNA LIBRARIES

Inventor(s): KURN NURITH [US]; LI BIN [US] ± (KURN, NURITH, ; LI, BIN)

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

Classification: **- international:** ***C12P19/34; C40B40/06***
- cooperative: **B01J19/0046**

Application number: WO2014US24581 20140312 Global Dossier

Priority number (s): US201361801510P 20130315 ; US201314030761 20130918

Also published as: CA2903125 (A1) CN105143525 (A) EP2971289 (A1) JP2016511007 (A) SG11201507136S (A) more

Abstract of WO2014150931 (A1)

The invention provides methods and compositions, including kits, for the construction of directional nucleic acid libraries. The invention further provides methods and compositions for the amplification and sequencing of directional cDNA libraries.

IDT Exhibit 1020.3586

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 September 2014 (25.09.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/150931 A1

(51) International Patent Classification:
C40B 40/06 (2006.01) *C12P 19/34* (2006.01)

(21) International Application Number:
PCT/US2014/024581

(22) International Filing Date:
12 March 2014 (12.03.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/801,510 15 March 2013 (15.03.2013) US
14/030,761 18 September 2013 (18.09.2013) US

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

(72) Inventors: **KURN, Nurith**; 2876 Ramona St., Palo Alto, CA 94306 (US). **LI, Bin**; 3723 Heron Way, Palo Alto, CA 94306 (US).

(74) Agents: **HSU, Jennifer, L.** et al.; WILSON SONSINI GOODRICH & ROSATI, 650 Page Mill Road, Palo Alto, CA 94304-1050 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,

[Continued on next page]

(54) Title: METHODS, COMPOSITIONS AND KITS FOR GENERATION OF STRANDED RNA OR DNA LIBRARIES

(57) Abstract: The invention provides methods and compositions, including kits, for the construction of directional nucleic acid libraries. The invention further provides methods and compositions for the amplification and sequencing of directional cDNA libraries.

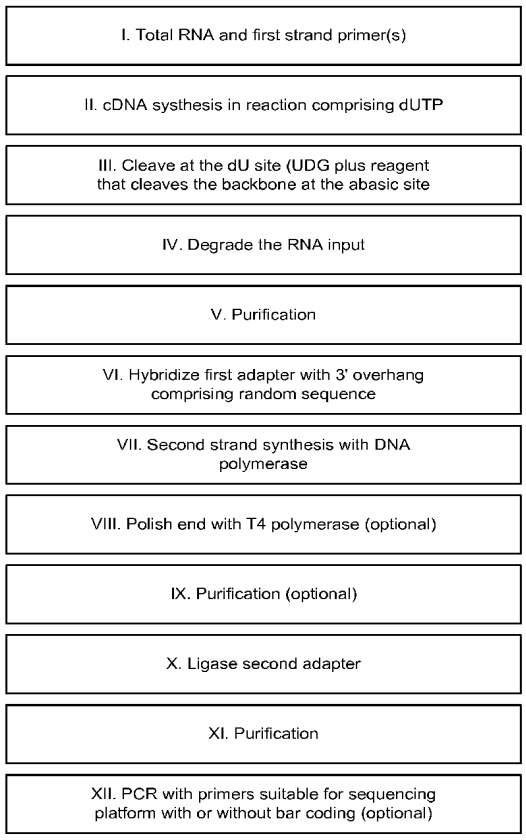


FIG. 3

WO 2014/150931 A1

HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:

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

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

METHODS, COMPOSITIONS AND KITS FOR GENERATION OF STRANDED RNA OR DNA LIBRARIES

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application No. 61/801,510 filed March 15, 2013, and also claims the benefit of U.S. Application Serial. No. 14/030,761, filed September 18, 2013, which applications are incorporated herein by reference in its entirety.

BACKGROUND

[0002] Rapid developments in massively parallel sequencing technologies in recent years have enabled whole genome and whole transcriptome sequencing and analysis, opening new approaches to functional genomics. One of these next generation sequencing methods involves direct sequencing of complementary DNA (cDNA) generated from messenger and structural RNAs (RNA-Seq). RNA-Seq can provide several key advantages over traditional sequencing methods. RNA-Seq can allow for high resolution study of all expressed coding and non-coding transcripts, annotating the 5' and 3' ends and splice junctions of each transcript, quantification of the relative number of transcripts in each cell can provide a way to measure and characterize RNA splicing by measuring the levels of each splice variant. Similarly, massively parallel sequencing technologies can enable whole genome sequencing or sequencing of multiplex targeted genomic sequences of interests at high resolution.

[0003] One potential drawback of performing standard RNA-Seq is the lack of information on the direction of transcription. Standard cDNA libraries constructed for RNA-Seq consist of randomly primed double-stranded cDNA. Non-directional ligation of adaptors containing universal priming sites prior to sequencing can lead to a loss of information as to which strand was present in the original RNA template. Although strand information can be inferred in some cases by subsequent analysis, for example, by using open reading frame (ORF) information in transcripts that encode for a protein or by assessing splice site information in eukaryotic genomes, direct information on the originating strand can be desirable. For example, direct information on which strand was present in the original RNA sample can be used to assign the sense strand to a non-coding RNA, and when resolving overlapping transcripts.

[0004] Several methods have recently been developed for strand-specific RNA-Seq. These methods can be divided into two main classes. The first class can utilize distinct adaptors in a known orientation relative to the 5' and 3' end of the RNA transcript. The end result can be a cDNA library where the 5' and 3' end of the original RNA are flanked by two distinct adaptors. A disadvantage of this method can be that only the ends of the cloned molecules preserve directional information. This situation can be problematic for strand-specific manipulations of long clones, and can lead to loss of directional information when there is fragmentation.

[0005] The second class of strand-specific RNA-Seq methods can mark one strand of either the original RNA (for example, by bisulfite treatment) or the transcribed cDNA (for example, by incorporation of modified nucleotides), followed by degradation of the unmarked strand. Strand marking by bisulfite

IDT Exhibit 1020.3589

treatment of RNA can be labor intensive and can require alignment of the sequencing reads to reference genomes that have all the cytosine bases converted to thymines on one of the two strands. The analysis can further be complicated due to the fact that base conversion efficiency during bisulfite treatment can be imperfect, i.e. less than 100%.

[0006] Strand marking by modification of the second strand of cDNA has become the preferred approach for directional cDNA cloning and sequencing (see e.g., Levin *et al.,* 2010). However, cDNA second strand marking approaches can be insufficient to preserve directionality information when using conventional blunt-end ligation and cDNA library construction strategies with duplex adaptors, where two universal sequencing sites are introduced by two separate adapters.

[0007] A major drawback of the current directional transcriptome or genome sequencing can be the requirement of generating first and second strand copies of the desired input strand, or the RNA transcripts, to generate dsDNA prior to fragmentation and attachment of directional or non-directional adaptors, in so far as random second strand synthesis may introduce unknown distortion to the desired library and add complexity to the sequencing library generation.

[0008] There is a need for improved and simplified methods for directional cDNA libraries for transcriptome or genome sequencing. The methods, compositions, and kits described herein can fulfill this need.

[0009] Provided herein are methods, compositions and kits for the generation of directional sequencing libraries from RNA and dsDNA. The methods, compositions and kits can be used for generation of directional libraries of whole transcriptome, whole genome, targeted or selected transcripts, and can also be applied for the generation of non-directional whole genome sequencing libraries.

SUMMARY

[0010] In one aspect, a method provided herein is the synthesis of complementary DNA strands comprising a non-canonical nucleotide at a defined density to enable fragmentation of the cDNA to a desired size range using an enzyme can that cleave the base portion of the non-canonical nucleotide to generate an abasic site, and further cleavage of the backbone at the abasic site by either enzymatic or chemical or thermal (e.g. heat) means. The DNA fragments produced can comprise a blocked 3'-end. Enzymatic cleavage at the abasic site can produce a 5'-phosphate end, which can be used in a further manipulation for adaptor ligation.

[0011] In another aspect, provided herein is a method of priming second strand synthesis using primers designed to anneal to the 3'-ends of all the fragments of the first strand complementary DNA generated as above.

[0012] First strand complementary DNA synthesis from RNA templates, such as total RNA, can be performed using various priming schemes. First strand primers useful for the performance of the methods provided herein can be random primers, such as random hexamer, which can be capable of priming at multiple sites on the target RNA. In another embodiment, first strand primers can comprise sequences

IDT Exhibit 1020.3590

specific for hybridization to targeted transcripts, or part thereof. In yet another embodiment, the first strand primers can comprise sequences designed to prime on all transcripts other than groups of transcripts which are not desired. For example, the first strand cDNA primers can comprise sequences designed to preferentially prime on all transcripts and not prime on structural RNA, such as all rRNAs.

[0013] Regardless of the design of first strand cDNA primers, first strand synthesis can be carried out by reverse transcriptase in reaction mixtures comprising one or more non-canonical nucleotides in a mixture of the corresponding nucleotides, wherein the ratio of a canonical to non-canonical nucleotide can be selected to result in incorporation of the non-canonical nucleotide at a density that will enable fragmentation to generate fragments within a desired fragment size range. The desired size range of the fragmented products can be selected to fit the desired size range of the inserts in the sequencing libraries, so as to accommodate use on various sequencing platforms of choice, or any other downstream manipulations.

[0014] Generating single stranded cDNA fragments of the desired size range can be beneficial for a fully automated process for the generation of sequencing and other libraries. In some cases, generation of the first strand cDNA fragments does not require any physical methods of fragmentation such as sonication, which can result in loss of product, and can be useful for generation of library from minute amount of template input, such as single cell analysis or analysis of templates from a very small sample.

[0015] The non-canonical nucleotide dUTP can be used in combination with treatment with UNG to generate abasic sites. The fragmentation of the backbone at the abasic site can be carried out in the same reaction mixture by polyamine such as DMED, or combination of enzymes, such as in USER (combination of UNG and endonuclease VIII from NEB). Alternatively, cleavage at the abasic site can be carried out by heating the reaction mixture or by various chemical methods

[0016] Methods provided herein do not require second strand synthesis at random sites, as is commonly used in various library preparation methods. Thus the methods provided herein provide reduce bias of selective priming to generate second strand cDNA.

[0017] The appending of defined and different sequences at the two ends of the cDNA product can be used for generation of stranded libraries, or libraries, which retain strand specificity. The process of appending a defined sequence to the 3'-end of all the fragments generated by a procedure provided herein can be carried out by priming of all fragments with a partial duplex comprising a single stranded DNA at the 3'-end, wherein the single stranded DNA portion comprises a random sequence. The length of the single strand overhang can vary from at least 6 to at least 7, 8, or nine nucleotides. The single strand overhang can hybridize to the 3'-ends of all the generated fragments and can be extended along the fragments by a DNA polymerase. Various structures of the partial duplex primer are anticipated. Some examples are shown in **FIG. 2**. The two strands forming the dsDNA portion can be two oligonucleotides which can further be connected by a loop. The loop, or linker, can comprise an oligonucleotide or can comprise a non-nucleotide linker, or combination thereof. It can also comprise nucleotide analogs.

IDT Exhibit 1020.3591

[0018] Following elongation of the hybridized single stranded DNA portion of the said partial duplex along the fragments by DNA polymerase, the end of the newly synthesized dsDNA can be repaired to generate a blunt end. The second defined sequence at the other end of the synthesized second strand cDNA can be appended by ligation. Various ligation modes are anticipated. Two examples of the ligation of a second adapter are shown in **FIG. 1A** and **1B**. A/T dependent ligation is also possible. The product of the process described thus far can be a second strand cDNA with defined ends at the two ends, which can be suitable for further manipulation, such as amplification, addition of desired sequences suitable for analysis on desired platforms, cloning and the like. The added sequences can comprise one or more barcodes, and/or sequences useful for attachment to a solid surface such as the Illumina sequencing flow cells, and the like. The appended sequences can also comprise random sequences useful for marking all fragments with unique sequence which can enable absolute quantification.

[0019] A workflow of a process for generation of directional sequencing libraries from RNA using methods and compositions described herein is depicted in **FIG. 3**.

[0020] Also provided herein are methods and compositions for generation of libraries from dsDNA templates, such as genomic DNA templates. The libraries can be useful for whole genome amplification and sequencing and can also be useful for library generation from very small samples, without the need for physical fragmentation of the template dsDNA. As shown in **FIG. 4**, initiation of complementary strand synthesis can be carried out without primer annealing to denatured dsDNA templates. DNA synthesis along the template DNA strands can be initiated from a nicked site. The use of various nicking enzymes is well known in the art. Nicking enzymes that are either strand specific or not, can be useful for the methods described herein. Random fragmentation of the complementary DNA generated by extension from the nicking site can be achieved by the random insertion of the non-canonical nucleotide, rather than random nicking. Thus, it is possible to use any desired nicking enzyme regardless of the sequence dependence of the chosen nicking enzyme. Enzymes that nick the dsDNA template to generate large distances between the nicking sites can be desired for maximal coverage and random fragmentation by the methods described herein.

[0021] The process for generation of libraries from dsDNA templates can comprise further steps which are similar to that described for the generation of stranded cDNA sequencing libraries, as is schematically depicted in **FIG. 4**.

[0022] **FIG. 5** describes a process for amplification of fragmented and appended products by Single Primer Isothermal Amplification (SPIA) employing chimeric DNA/RNA primers. The amplification products generated by this process can comprise defined sequences at the 3'- and 5'-portions, thus providing strand retention with respect to the input template.

[0023] In one aspect, described herein is a method for generating a directional cDNA library, the method comprising: a) annealing one or more primers to a template RNA; b) extending the one or more primers in the presence of a reaction mixture comprising dATP, dCTP, dGTP, dTTP, and dUTP, wherein the reaction mixture comprises a ratio of dUTP to dTTP, wherein the ratio permits incorporation of dUTP at a

IDT Exhibit 1020.3592

desired density, thereby generating a one or more first strand complementary DNAs (cDNAs) comprising dUTP incorporated at a desired density; c) selectively cleaving the one or more first strand cDNAs comprising dUTPs incorporated at a desired density with uracil-N-glycosylase (UNG) and an agent capable of cleaving a phosphodiester backbone at an abasic site created by the UNG, wherein the cleaving generates a plurality of first strand cDNA fragments of a desired size comprising a blocked 3' end; d) annealing a first adapter comprising a partial duplex and a 3' overhang to a 3' end of one or more of the plurality of first strand cDNA fragments comprising a blocked 3' end, wherein the first adapter comprises a sequence A, and wherein the annealing comprises hybridizing a random sequence at the 3' overhang to a complementary sequence present at the 3' end of the one or more of the plurality of first strand cDNA fragments comprising a blocked 3' end; e) extending the 3' overhang hybridized to the complementary sequence with a DNA polymerase, wherein one or more double stranded cDNA fragments comprising the sequence A at one end is generated; and f) ligating a second adapter comprising a sequence B to the one or more double stranded cDNA fragments comprising the sequence A at one end, wherein the ligating generates one or more double stranded cDNA fragments comprising the sequence A at one end and the sequence B at an opposite end, thereby generating the directional polynucleotide library. In some embodiments, the one or more primers comprise a random primer. In some embodiments, the one or more primers comprise a sequence specific to a target template RNA or group of RNAs. In some embodiments, the group of RNAs comprises substantially all transcripts. In some embodiments, the group of RNAs does not comprise structural RNA, wherein the structural RNA comprises ribosomal RNA (rRNA). In some embodiments, the method further comprises amplifying the directional cDNA library, thereby generating amplified products. In some embodiments, the method further comprises an additional step of sequencing the amplified products. In some embodiments, the amplification comprises SPIA. In some embodiments, the amplification comprises a use of primers, wherein one or more of the primers comprises one or more barcode sequences. In some embodiments, the sequencing comprises next generation sequencing. In some embodiments, the method further comprises degrading the template RNA following step b.). In some embodiments, the cleaving comprises exposing the template RNA sample to an RNase. In some embodiments, the agent capable of cleaving a phosphodiester backbone comprises an enzyme, chemical agent, and/or heat. In some embodiments, the chemical agent is a polyamine. In some embodiments, the polyamine is N,N-dimethylethylenediamine (DMED). In some embodiments, the enzyme is an endonuclease. In some embodiments, the endonuclease is endonuclease VIII. In some embodiments, the partial duplex comprises a long strand and a short strand, wherein the long strand comprises the sequence A that forms a duplex with the short strand and a 3' overhang. In some embodiments, the short strand further comprises a block at a 3' and/or a 5' end. In some embodiments, the first adapter further comprises a block at a 5' end of the long strand. In some embodiments, the first adapter comprises a plurality of first adapters, wherein the random sequence on each of the plurality of first adapters is different than the random sequence on another of the plurality of first adapters, and wherein each of the plurality of first adapters comprises the sequence A. In some

IDT Exhibit 1020.3593

embodiments, step d) results in substantially all of the plurality of first strand cDNA fragments of a desired size comprising a blocked 3' end generated in step c) further comprising one of the plurality of first adapters annealed the 3' end. In some embodiments, the first adapter further comprises a block at a 5' end of the short strand. In some embodiments, the first adapter further comprises a stem loop, wherein the stem loop links a 5' end of a long strand of the partial duplex with a 3' end of a short strand of the partial duplex, and wherein the long strand comprises the sequence A and the 3' overhang. TIn some embodiments, the 3' overhang comprises at least 6, 7, 8, or 9 nucleotides. In some embodiments, the second adapter comprises a partial duplex, wherein the partial duplex comprises a long strand hybridized to a short strand, wherein the long strand comprises the sequence B and an overhang. TIn some embodiments, the long strand comprises the sequence B and a 3' overhang, and wherein the short strand comprises a block at a 3' end. In some embodiments, the ligating generates the one or more double stranded cDNA fragments comprising the sequence A at one end and the sequence B at an opposite end, wherein the sequence A is at a 5' end on one end and the sequence B is at a 3' end on the opposite end. In some embodiments, the long strand comprises the sequence B and a 5' overhang, and wherein the short strand comprises a block at a 5' end. In some embodiments, the ligating generates the one or more double stranded cDNA fragments comprising the sequence A at one end and the sequence B at an opposite end, wherein the sequence A is at a 5' end on one end and the sequence B is at a 5' end on the opposite end. In some embodiments, a 3' end of the opposite end is extended using the sequence B as a template, thereby generating one or more double stranded cDNA fragments comprising the sequence A at a 5' end on one end and a sequence complementary to the sequence B, B', at a 3' end on the opposite end. In some embodiments, the ligating comprises blunt end ligation, wherein the one or more double stranded cDNA fragments comprising the sequence A at one end generated in step e) are end repaired prior to step f). In some embodiments, the first and/or second adapter further comprises one or more barcodes.

[0024] In one aspect, described herein is a method for whole transcriptome directional sequencing, the method comprising: a) annealing one or more primers to a template RNA; b) extending the primer in the presence of a reaction mixture comprising dATP, dCTP, dGTP, dTTP, and dUTP, wherein the reaction mixture comprises a ratio of dUTP to dTTP, wherein the ratio permits incorporation of dUTP at a desired density, thereby generating one or more first strand complementary DNAs (cDNAs) comprising dUTP incorporated at a desired density; c) selectively cleaving the one or more first strand cDNAs comprising dUTPs incorporated at a desired density with uracil-N-glycosylase (UNG) and an agent capable of cleaving a phosphodiester backbone at an abasic site created by the UNG, wherein the cleaving generates a plurality of first strand cDNA fragments of a desired size comprising a blocked 3' end; d) annealing a first adapter comprising a partial duplex and a 3' overhang to a 3' end of one or more of the plurality of first strand cDNA fragments comprising a blocked 3' end, wherein the first adapter comprises a sequence A, and wherein the annealing comprises hybridizing a random sequence at the 3' overhang to a complementary sequence present at the 3' end of the one or more of the plurality of first strand cDNA fragments comprising a blocked 3' end; e) extending the 3' overhang hybridized to the complementary

IDT Exhibit 1020.3594

sequence with a DNA polymerase, wherein one or more double stranded cDNA fragments comprising the sequence A at one end is generated; f) ligating a second adapter comprising a sequence B to the one or more double stranded cDNA fragments comprising the sequence A at one end, wherein the ligating generates one or more double stranded cDNA fragments comprising the sequence A at one end and the sequence B at an opposite end thereby generating a directional cDNA library; and g) amplifying and/or sequencing the directional cDNA library. 5In some embodiments, the one or more primers comprise a random primer. In some embodiments, the one or more primers comprise a sequence specific to a target template RNA or group of RNAs. In some embodiments, the group of RNAs comprises substantially all transcripts. In some embodiments, the group of RNAs does not comprise structural RNA, wherein the structural RNA comprises ribosomal RNA (rRNA). In some embodiments, the amplification comprises SPIA. In some embodiments, the amplification comprises a use of primers, wherein one or more of the primers comprises a barcode sequence. In some embodiments, the sequencing comprises next generation sequencing. In some embodiments, the method further comprises degrading the template RNA following step b.). In some embodiments, the cleaving comprises exposing the template RNA sample to an RNase. In some embodiments, the agent capable of cleaving a phosphodiester backbone comprises an enzyme, chemical agent, and/or heat. In some embodiments, the chemical agent is a polyamine. In some embodiments, the polyamine is N,N-dimethylethylenediamine (DMED). In some embodiments, the enzyme is an endonuclease. In some embodiments, the endonuclease is endonuclease VIII. In some embodiments, the partial duplex comprises a long strand and a short strand, wherein the long strand comprises the sequence A that forms a duplex with the short strand and a 3' overhang. In some embodiments, the short strand further comprises a block at a 3' and/or a 5' end. In some embodiments, the first adapter further comprises a block at a 5' end of the long strand. In some embodiments, the first adapter comprises a plurality of first adapters, wherein the random sequence on each of the plurality of first adapters is different than the random sequence on another of the plurality of first adapters, and wherein each of the plurality of first adapters comprises the sequence A. In some embodiments, step d) results in substantially all of the plurality of first strand cDNA fragments of a desired size comprising a blocked 3' end generated in step c) further comprising one of the plurality of first adapters annealed the 3' end. In some embodiments, the first adapter further comprises a block at a 5' end of the short strand. In some embodiments, the first adapter further comprises a stem loop, wherein the stem loop links a 5' end of a long strand of the partial duplex with a 3' end of a short strand of the partial duplex, and wherein the long strand comprises the sequence A and the 3' overhang. In some embodiments, the 3' overhang comprises at least 6, 7, 8, or 9 nucleotides. In some embodiments, the second adapter comprises a partial duplex, wherein the partial duplex comprises a long strand hybridized to a short strand, wherein the long strand comprises the sequence B and an overhang. In some embodiments, the long strand comprises the sequence B and a 3' overhang, and wherein the short strand comprises a block at a 3' end. In some embodiments, the ligating generates the one or more double stranded cDNA fragments comprising the sequence A at one end and the sequence B at an opposite end, wherein the sequence A is at a 5' end on

IDT Exhibit 1020.3595

one end and the sequence B is at a 3' end on the opposite end. In some embodiments, the long strand comprises the sequence B and a 5' overhang, and wherein the short strand comprises a block at a 5' end. In some embodiments, the ligating generates the one or more double stranded cDNA fragments comprising the sequence A at one end and the sequence B at an opposite end, wherein the sequence A is at a 5' end on one end and the sequence B is at a 5' end on the opposite end.. In some embodiments, a 3' end of the opposite end is extended using the sequence B as a template, thereby generating one or more double stranded cDNA fragments comprising the sequence A at a 5' end on one end and a sequence complementary to the sequence B, B', at a 3' end on the opposite end. In some embodiments, the ligating comprises blunt end ligation, wherein the one or more double stranded cDNA fragments comprising the sequence A at one end generated in step e) are end repaired prior to step f). In some embodiments, the first and/or second adapter further comprises one or more barcodes.

[0025] In one aspect, described herein is a method for generating a directional cDNA library, the method comprising: a) treating a template dsDNA with a nicking enzyme, wherein the treating generates one or more breaks in a phosphodiester backbone of one strand of the template dsDNA, wherein the break produces one or more 3' hydroxyls in the one strand; b) extending the one or more 3' hydroxyls, wherein the extending is performed in the presence of a reaction mixture comprising dATP, dCTP, dGTP, dTTP, and dUTP, wherein the reaction mixture comprises a ratio of dUTP to dTTP, wherein the ratio permits incorporation of dUTP at a desired density, thereby generating one or more first strand complementary DNAs (cDNAs) comprising dUTP incorporated at a desired density; c) selectively cleaving the one or more first strand cDNAs comprising dUTPs incorporated at a desired density with uracil-N-glycosylase (UNG) and an agent capable of cleaving a phosphodiester backbone at an abasic site created by the UNG, wherein the cleaving generates a plurality of first strand cDNA fragments of a desired size comprising a blocked 3' end; d) annealing a first adapter comprising a partial duplex and a 3' overhang to a 3' end of one or more of the plurality of first strand cDNA fragments comprising a blocked 3' end, wherein the first adapter comprises a sequence A, and wherein the annealing comprises hybridizing a random sequence at the 3' overhang to a complementary sequence present at the 3' end of the one or more of the plurality of first strand cDNA fragments comprising a blocked 3' end; e) extending the 3' overhang hybridized to the complementary sequence with a DNA polymerase, wherein one or more double stranded cDNA fragments comprising the sequence A at one end is generated; and f) ligating a second adapter comprising a sequence B to the one or more double stranded cDNA fragments comprising the sequence A at one end, wherein the ligating generates one or more double stranded cDNA fragments comprising the sequence A at one end and the sequence B at an opposite end thereby generating a directional cDNA library. In some embodiments, the method further comprises amplifying the directional cDNA library, thereby generating amplified products. In some embodiments, the method further comprises an additional step of sequencing the amplified products. In some embodiments, the amplification comprises SPIA. In some embodiments, the amplification comprises a use of primers, wherein one or more of the primers comprises one or more barcode sequences. In some embodiments, the sequencing comprises next generation sequencing. In

IDT Exhibit 1020.3596

some embodiments, the nicking enzyme comprises a strand specific nicking enzyme. In some embodiments, the extending the one or more 3' hydroxyls in step b) is performed with a DNA polymerase comprising strand displacement activity. In some embodiments, the agent capable of cleaving a phosphodiester backbone comprises an enzyme, chemical agent, and/or heat. In some embodiments, the chemical agent is a polyamine. In some embodiments, the polyamine is N,N-dimethylethylenediamine (DMED). In some embodiments, the enzyme is an endonuclease. In some embodiments, the endonuclease is endonuclease VIII. In some embodiments, the partial duplex comprises a long strand and a short strand, wherein the long strand comprises the sequence A that forms a duplex with the short strand and a 3' overhang. In some embodiments, the short strand further comprises a block at a 3' and/or a 5' end. In some embodiments, the first adapter further comprises a block at a 5' end of the long strand. In some embodiments, the first adapter comprises a plurality of first adapters, wherein the random sequence on each of the plurality of first adapters is different than the random sequence on another of the plurality of first adapters, and wherein each of the plurality of first adapters comprises the sequence A. In some embodiments, step d) results in substantially all of the plurality of first strand cDNA fragments of a desired size comprising a blocked 3' end generated in step c) further comprising one of the plurality of first adapters annealed the 3' end. In some embodiments, the first adapter further comprises a block at a 5' end of the short strand. In some embodiments, the first adapter further comprises a stem loop, wherein the stem loop links a 5' end of a long strand of the partial duplex with a 3' end of a short strand of the partial duplex, and wherein the long strand comprises the sequence A and the 3' overhang. In some embodiments, the 3' overhang comprises at least 6, 7, 8, or 9 nucleotides. In some embodiments, the second adapter comprises a partial duplex, wherein the partial duplex comprises a long strand hybridized to a short strand, wherein the long strand comprises the sequence B and an overhang. In some embodiments, the long strand comprises the sequence B and a 3' overhang, and wherein the short strand comprises a block at a 3' end. TIn some embodiments, the ligating generates the one or more double stranded cDNA fragments comprising the sequence A at one end and the sequence B at an opposite end, wherein the sequence A is at a 5' end on one end and the sequence B is at a 3' end on the opposite end. In some embodiments, the long strand comprises the sequence B and a 5' overhang, and wherein the short strand comprises a block at a 5' end. In some embodiments, the ligating generates the one or more double stranded cDNA fragments comprising the sequence A at one end and the sequence B at an opposite end, wherein the sequence A is at a 5' end on one end and the sequence B is at a 5' end on the opposite end. In some embodiments, a 3' end of the opposite end is extended using the sequence B as a template, thereby generating one or more double stranded cDNA fragments comprising the sequence A at a 5' end on one end and a sequence complementary to the sequence B, B', at a 3' end on the opposite end. In some embodiments, the ligating comprises blunt end ligation, wherein the one or more double stranded cDNA fragments comprising the sequence A at one end generated in step e) are end repaired prior to step f). In some embodiments, the first and/or second adapter further comprises one or more barcodes.

IDT Exhibit 1020.3597

[0026] In one aspect, described herein is a method for whole genome sequencing, the method comprising: a) treating genomic DNA with a nicking enzyme, wherein the treating generates one or more breaks in a phosphodiester backbone of a one strand of the genomic DNA, wherein the breaks produce one or more 3' hydroxyls in the one strand; b) extending the one or more 3' hydroxyls, wherein the extending is performed in the presence of a reaction mixture comprising dATP, dCTP, dGTP, dTTP, and dUTP, wherein the reaction mixture comprises a ratio of dUTP to dTTP, wherein the ratio permits incorporation of dUTP at a desired density, thereby generating one or more first strand complementary DNAs (cDNAs) comprising dUTP incorporated at a defined frequency; c) selectively cleaving the one or more first strand cDNA comprising dUTPs incorporated at a desired density with uracil-N-glycosylase (UNG) and an agent capable of cleaving a phosphodiester backbone at an abasic site created by the UNG, wherein the cleaving generates a plurality of first strand cDNA fragments of a desired size comprising a blocked 3' end; d) annealing a first adapter comprising a partial duplex and a 3' overhang to a 3' end of one or more of the plurality of first strand cDNA fragments comprising a blocked 3' end, wherein the first adapter comprises a sequence A, and wherein the annealing comprises hybridizing a random sequence at the 3' overhang to a complementary sequence present at the 3' end of the one or more of the plurality of first strand cDNA fragments comprising a blocked 3' end; e) extending the 3' overhang hybridized to the complementary sequence with a DNA polymerase, wherein one or more double stranded cDNA fragments comprising the sequence A at one end is generated; f) ligating a second adapter comprising a sequence B to the one or more double stranded cDNA fragments comprising the sequence A at one end, wherein the ligating generates one or more double stranded cDNA fragments comprising the sequence A at one end and the sequence B at an opposite end thereby generating a directional cDNA library; and g) amplifying and/or sequencing the directional cDNA library. In some embodiments, the amplification comprises SPIA. In some embodiments, the amplification comprises a use of primers, wherein one or more of the primers comprises a barcode sequence. In some embodiments, the sequencing comprises next generation sequencing. In some embodiments, the nicking enzyme comprises a strand specific nicking enzyme. In some embodiments, the extending the one or more 3' hydroxyls in step b) is performed with a DNA polymerase comprising strand displacement activity. In some embodiments, the agent capable of cleaving a phosphodiester backbone comprises an enzyme, chemical agent, and/or heat. In some embodiments, the chemical agent is a polyamine. In some embodiments, the polyamine is N,N-dimethylethylenediamine (DMED). In some embodiments, the enzyme is an endonuclease. In some embodiments, the endonuclease is endonuclease VIII. In some embodiments, the partial duplex comprises a long strand and a short strand, wherein the long strand comprises the sequence A that forms a duplex with the short strand and a 3' overhang. In some embodiments, the short strand further comprises a block at a 3' and/or a 5' end. In some embodiments, the first adapter further comprises a block at a 5' end of the long strand. In some embodiments, the first adapter comprises a plurality of first adapters, wherein the random sequence on each of the plurality of first adapters is different than the random sequence on another of the plurality of first adapters, and wherein each of the plurality of first adapters

IDT Exhibit 1020.3598

comprises the sequence A. In some embodiments, step d) results in substantially all of the plurality of first strand cDNA fragments of a desired size comprising a blocked 3' end generated in step c) further comprising one of the plurality of first adapters annealed the 3' end. In some embodiments, the first adapter further comprises a block at a 5' end of the short strand. In some embodiments, the first adapter further comprises a stem loop, wherein the stem loop links a 5' end of a long strand of the partial duplex with a 3' end of a short strand of the partial duplex, and wherein the long strand comprises the sequence A and the 3' overhang. In some embodiments, the 3' overhang comprises at least 6, 7, 8, or 9 nucleotides. In some embodiments, the second adapter comprises a partial duplex, wherein the partial duplex comprises a long strand hybridized to a short strand, wherein the long strand comprises the sequence B and an overhang. In some embodiments, the long strand comprises the sequence B and a 3' overhang, and wherein the short strand comprises a block at a 3' end. In some embodiments, the ligating generates the one or more double stranded cDNA fragments comprising the sequence A at one end and the sequence B at an opposite end, wherein the sequence A is at a 5' end on one end and the sequence B is at a 3' end on the opposite end. In some embodiments, the long strand comprises the sequence B and a 5' overhang, and wherein the short strand comprises a block at a 5' end. In some embodiments, the ligating generates the one or more double stranded cDNA fragments comprising the sequence A at one end and the sequence B at an opposite end, wherein the sequence A is at a 5' end on one end and the sequence B is at a 5' end on the opposite end. In some embodiments, a 3' end of the opposite end is extended using the sequence B as a template, thereby generating one or more double stranded cDNA fragments comprising the sequence A at a 5' end on one end and a sequence complementary to the sequence B, B', at a 3' end on the opposite end. In some embodiments, the ligating comprises blunt end ligation, wherein the one or more double stranded cDNA fragments comprising the sequence A at one end generated in step e) are end repaired prior to step f). In some embodiments, the first and/or second adapter further comprises one or more barcodes.

[0027] In one aspect, described herein is a method for generating a directional polynucleotide library, the method comprising: a) reverse transcribing a template RNA in the presence of one or more primers, reverse transcriptase, and a reaction mixture comprising a non-canonical nucleotide, wherein the reaction mixture comprises a ratio of the non-canonical nucleotide suitable to permit incorporation of the non-canonical nucleotide at a desired density, thereby generating a one or more first strand complementary DNAs (cDNAs) comprising the non-canonical nucleotide incorporated at a desired density; b) selectively cleaving the one or more first strand cDNAs comprising the non-canonical nucleotide incorporated at a desired density with a cleavage agent, wherein the cleaving with the cleavage agent generates a plurality of first strand cDNA fragments of a desired size comprising a blocked 3' end; c) annealing a first adapter comprising a partial duplex and a 3' overhang to a 3' end of one or more of the plurality of first strand cDNA fragments comprising a blocked 3' end, wherein the first adapter comprises a sequence A, and wherein the annealing comprises hybridizing a random sequence at the 3' overhang to a complementary sequence present at the 3' end of the one or more of the plurality of first strand cDNA fragments

IDT Exhibit 1020.3599

comprising a blocked 3' end; d) extending the 3' overhang hybridized to the complementary sequence with a DNA polymerase, wherein one or more double stranded cDNA fragments comprising the sequence A at one end is generated; and e) ligating a second adapter comprising a sequence B to the one or more double stranded cDNA fragments comprising the sequence A at one end, wherein the ligating generates one or more double stranded cDNA fragments comprising the sequence A at one end and the sequence B at an opposite end thereby generating the directional polynucleotide library. In some embodiments, the template RNA comprises mRNA. In some embodiments, the one or more primers comprise a random primer. In some embodiments, the one or more primers comprise a sequence specific to a target RNA or group of RNAs. In some embodiments, the group of RNAs comprises substantially all transcripts. In some embodiments, the group of RNAs does not comprise structural RNA, wherein the structural RNA comprises ribosomal RNA (rRNA). In some embodiments, the method further comprises degrading the template RNA following step a). In some embodiments, the non-canonical dNTP comprises dUTP. In some embodiments, the cleavage agent comprises a glycosylase and a polyamine, heat, or an enzyme. In some embodiments, the glycoslyase is uracil-N-glycosylase (UNG). In some embodiments, the polyamine is N,N-dimethylethylenediamine (DMED). In some embodiments, the enzyme comprises an endonuclease. In some embodiments, the endonuclease is endonuclease VIII. In some embodiments, the first adapter comprises a plurality of first adapters, wherein the random sequence on each of the plurality of first adapters is different than the random sequence on another of the plurality of first adapters, and wherein each of the plurality of first adapters comprises the sequence A. In some embodiments, the annealing results in substantially all of the plurality of first strand cDNA fragments of a desired size comprising a blocked 3' end further comprising one of the plurality of first adapters annealed the 3' end. In some embodiments, the partial duplex comprises a long strand and a short strand, wherein the long strand comprises the sequence A that forms a duplex with the short strand and a 3' overhang. In some embodiments, the short strand further comprises a block at a 3' and/or a 5' end. In some embodiments, the first adapter further comprises a stem loop, wherein the stem loop links a 5' end of a long strand of the partial duplex with a 3' end of a short strand of the partial duplex, and wherein the long strand comprises the sequence A and the 3' overhang. In some embodiments, the first adapter further comprises a block at a 5' end of the long strand. In some embodiments, the first adapter further comprises a block at a 5' end of the short strand. In some embodiments, the 3' overhang comprises at least 6, 7, 8, or 9 nucleotides. In some embodiments, the second adapter comprises a duplex, partial duplex, or single strand comprising a duplex portion connected by a stem loop. In some embodiments, the first and/or second adapter further comprises one or more barcodes. In some embodiments, the second adapter comprises a partial duplex, wherein the partial duplex comprises a long strand hybridized to a short strand, wherein the long strand comprises the sequence B and an overhang. In some embodiments, the long strand comprises the sequence B and a 3' overhang, and wherein the short strand comprises a block at a 3' end. In some embodiments, the ligating generates the one or more double stranded cDNA fragments comprising the sequence A at one end and the sequence B at an opposite end, wherein the sequence A is at a 5' end on

IDT Exhibit 1020.3600

one end and the sequence B is at a 3' end on the opposite end. In some embodiments, the long strand comprises the sequence B and a 5' overhang, and wherein the short strand comprises a block at a 5' end. In some embodiments, the ligating generates the one or more double stranded cDNA fragments comprising the sequence A at one end and the sequence B at an opposite end, wherein the sequence A is at a 5' end on one end and the sequence B is at a 5' end on the opposite end. In some embodiments, a 3' end of the opposite end is extended using the sequence B as a template, thereby generating one or more double stranded cDNA fragments comprising the sequence A at a 5' end on one end and a sequence complementary to the sequence B, B', at a 3' end on the opposite end. In some embodiments, the method further comprises amplifying the directional cDNA library, thereby generating amplified products. further comprising an additional step of sequencing the amplified products. In some embodiments, the amplification comprises SPIA. In some embodiments, the amplification comprises a use of primers, wherein one or more of the primers comprises a barcode sequence. In some embodiments, the sequencing comprises next generation sequencing. In some embodiments, the ligating comprises blunt end ligation, wherein the one or more double stranded cDNA fragments comprising the sequence A at one end generated in step e) are end repaired prior to step f).

[0028] In one aspect, described herein is a method for generating a directional polynucleotide library, the method comprising: a) treating a template DNA with a nicking enzyme, wherein the treating generates one or more breaks in a phosphodiester backbone of one strand of the template DNA, wherein the one or more breaks produce one or more 3' hydroxyls in the one strand; b) extending the one or more 3' hydroxyls, wherein the extending is performed in the presence of a reaction mixture comprising a non-canonical nucleotide, wherein the reaction mixture comprises a ratio of the non-canonical nucleotide suitable to permit incorporation of the non-canonical nucleotide at a desired density, thereby generating a one or more first strand complementary DNAs (cDNAs) comprising the non-canonical nucleotide incorporated at a desired density; c) selectively cleaving the one or more first strand cDNAs comprising the non-canonical nucleotide incorporated at a desired density with a cleavage agent, wherein the cleaving with the cleavage agent generates a plurality of first strand cDNA fragments of a desired size comprising a blocked 3' end; d) annealing afirst adapter comprising a partial duplex and a 3' overhang to a 3' end of one or more of the plurality of first strand cDNA fragments comprising a blocked 3' end, wherein the first adapter comprises a sequence A, and wherein the annealing comprises hybridizing a random sequence at the 3' overhang to a complementary sequence present at the 3' end of the one or more of the plurality of first strand cDNA fragments comprising a blocked 3' end; e) extending the 3' overhang hybridized to the complementary sequence with a DNA polymerase, wherein one or more double stranded cDNA fragments comprising the sequence A at one end is generated; and f) ligating a second adapter comprising a sequence B to the one or more double stranded cDNA fragments comprising the sequence A at one end, wherein the ligating generates one or more double stranded cDNA fragments comprising the sequence A at one end and the sequence B at an opposite end thereby generating the directional polynucleotide library. In some embodiments, the template DNA comprises double stranded DNA (dsDNA). In some

IDT Exhibit 1020.3601

embodiments, the template DNA comprises genomic DNA. In some embodiments, the nicking enzyme comprises a strand specific nicking enzyme. In some embodiments, the extending the 3' hydroxyl in step b) is performed with a DNA polymerase comprising strand displacement activity. In some embodiments, the non-canonical dNTP comprises dUTP. In some embodiments, the cleavage agent comprises a glycosylase and a polyamine, heat, or an enzyme. In some embodiments, the glycoslyase is uracil-N-glycosylase (UNG). In some embodiments, the polyamine is N,N-dimethylethylenediamine (DMED). In some embodiments, the enzyme comprises an endonuclease. In some embodiments, the endonuclease is endonuclease VIII. In some embodiments, the first adapter comprises a plurality of first adapters, wherein the random sequence on each of the plurality of first adapters is different than the random sequence on another of the plurality of first adapters, and wherein each of the plurality of first adapters comprises the sequence A. In some embodiments, the annealing results in substantially all of the plurality of first strand cDNA fragments of a desired size comprising a blocked 3' end further comprising one of the plurality of first adapters annealed the 3' end. In some embodiments, the partial duplex comprises a long strand and a short strand, wherein the long strand comprises the sequence A that forms a duplex with the short strand and a 3' overhang. In some embodiments, the short strand further comprises a block at a 3' and/or a 5' end. In some embodiments, the first adapter further comprises a stem loop, wherein the stem loop links a 5' end of a long strand of the partial duplex with a 3' end of a short strand of the partial duplex, and wherein the long strand comprises the sequence A and the 3' overhang. In some embodiments, the first adapter further comprises a block at a 5' end of the long strand. In some embodiments, the first adapter further comprises a block at a 5' end of the short strand. In some embodiments, the 3' overhang comprises at least 6, 7, 8, or 9 nucleotides. In some embodiments, the second adapter comprises a duplex, partial duplex, or single strand comprising a duplex portion connected by a stem loop. In some embodiments, the first and/or second adapter further comprises one or more barcodes. In some embodiments, the second adapter comprises a partial duplex, wherein the partial duplex comprises a long strand hybridized to a short strand, wherein the long strand comprises the sequence B and an overhang. In some embodiments, the long strand comprises the sequence B and a 3' overhang, and wherein the short strand comprises a block at a 3' end. In some embodiments, the ligating generates the one or more double stranded cDNA fragments comprising the sequence A at one end and the sequence B at an opposite end, wherein the sequence A is at a 5' end on one end and the sequence B is at a 3' end on the opposite end. In some embodiments, the long strand comprises the sequence B and a 5' overhang, and wherein the short strand comprises a block at a 5' end. In some embodiments, the ligating generates the one or more double stranded cDNA fragments comprising the sequence A at one end and the sequence B at an opposite end, wherein the sequence A is at a 5' end on one end and the sequence B is at a 5' end on the opposite end. In some embodiments, a 3' end of the opposite end is extended using the sequence B as a template, thereby generating one or more double stranded cDNA fragments comprising the sequence A at a 5' end on one end and a sequence complementary to the sequence B, B', at a 3' end on the opposite end. In some embodiments, the method further comprises

IDT Exhibit 1020.3602

amplifying the directional cDNA library, thereby generating amplified products. In some embodiments, the method further comprises an additional step of sequencing the amplified products. In some embodiments, the amplification comprises SPIA. In some embodiments, the amplification comprises a use of primers, wherein one or more of the primers comprise a barcode sequence. In some embodiments, the sequencing comprises next generation sequencing. In some embodiments, the ligating comprises blunt end ligation, wherein the one or more double stranded cDNA fragments comprising the sequence A at one end generated in step e) are end repaired prior to step f).

[0029] In one aspect, described herein is a method for generating a directional polynucleotide library, the method comprising: a) chemically cleaving a phosphodiester backbone of one or more polynucleotides comprising one or more abasic sitse at the one or more abasic sites, whereby one or more polynucleotides within a desired size range and comprising a blocked 3' end are generated; b) appending a first adapter to a 3' end of the one or more polynucleotides comprising a blocked 3' end, wherein the first adapter comprises a sequence A, wherein the sequence A is non-hybridizable to the one or more polynucleotides comprising a blocked 3' end; c) extending a 3' end of the first adapter appended to the 3' end of the one or more polynucleotides comprising a blocked 3' end using the one or more polynucleotides comprising a blocked 3' end as template, wherein one or more double stranded polynucleotides comprising the sequence A at one end is generated; and d) appending a second adapter comprising a sequence B to the one or more double stranded polynucleotide comprising the sequence A at one end, wherein the sequence B is different than the sequence A and the appending generates one or more double stranded polynucleotides comprising the sequence A at one end and the sequence B at an opposite end, thereby generating the directional polynucleotide library. In some embodiments, the phosphodiester backbone is cleaved with a polyamine to generate one or more polynucleotides within a desired size range and with a blocked 3' end. In some embodiments, the polyamine is N, N'-dimethylethylenediamine (DMED). In some embodiments, the one or more polynucleotides comprising one or more abasic sites are generated by cleaving a base portion of a non-canonical nucleotide in one or more polynucleotides with an enzyme capable of cleaving the base portion of the non-canonical nucleotide, whereby an abasic site is generated. In some embodiments, the non-canonical nucleotide is selected from the group consisting of dUTP, dITP, and 5-OH-Me-dCTP. In some embodiments, the enzyme capable of cleaving the base portion of the non-canonical nucleotide is an N-glycosylase. In some embodiments, the N-glycosylase is selected from the group consisting of Uracil N-Glycosylase (UNG), hypoxanthine-N-Glycosylase, and hydroxy-methyl cytosine-N-glycosylase. In some embodiments, the non-canonical nucleotide is dUTP and the enzyme capable of cleaving the base portion of the non-canonical nucleotide is UNG. In some embodiments, the non-canonical nucleotide is dUTP, the enzyme capable of cleaving the base portion of the non-canonical nucleotide is UNG, and the phosphodiester backbone is cleaved with DMED. In some embodiments, the one or more polynucleotides comprising one or more non-canonical nucleotides are synthesized in the presence of two or more different non-canonical nucleotides, whereby one or more polynucleotides comprising two or more different non-canonical nucleotides are synthesized. In some embodiments,the

IDT Exhibit 1020.3603

one or more polynucleotides comprising one or more abasic sites are synthesized from a template nucleic acid comprising DNA or RNA. In some embodiments, the template nucleic acid is selected from the group consisting of mRNA, cDNA, and genomic DNA. TIn some embodiments, the one or more polynucleotides comprising one or more abasic sites are single stranded or double stranded. In some embodiments, the one or more polynucleotides comprising one or more abasic sites are synthesized by an amplification method selected from the group consisting of polymerase chain reaction (PCR), strand displacement amplification (SDA), multiple displacement amplification (MDA), rolling circle amplification (RCA), single primer isothermal amplification (SPIA), and Ribo-SPIA. In some embodiments, the one or more polynucleotide comprising one or more abasic sites are synthesized by a method selected from the group consisting of reverse transcription, primer extension, limited primer extension, replication, and nick translation. In some embodiments, the first adapter further comprises a partial duplex and a 3' overhang. In some embodiments, the first adapter comprises a plurality of first adapters, wherein the random sequence on each of the plurality of first adapters is different than the random sequence on another of the plurality of first adapters, and wherein each of the plurality of first adapters comprises the sequence A. In some embodiments, the annealing results in substantially all of the plurality of first strand cDNA fragments of a desired size comprising a blocked 3' end further comprising one of the plurality of first adapters annealed the 3' end. In some embodiments, the appending comprises annealing the 3' overhang of the first adapter to the 3' end of the polynucleotide comprising a blocked 3' end, wherein the annealing comprises hybridizing a random sequence at the 3' overhang to a complementary sequence present at the 3' end of the polynucleotide comprising a blocked 3' end. In some embodiments, the partial duplex comprises a long strand and a short strand, wherein the long strand comprises the sequence A that forms a duplex with the short strand and the 3' overhang. In some embodiments, the short strand further comprises a block at a 3' and/or a 5' end of the short strand. In some embodiments, the first adapter further comprises a stem loop, wherein the stem loop links a 5' end of a long strand of the partial duplex with a 3' end of a short strand of the partial duplex, and wherein the long strand comprises the sequence A and the 3' overhang. In some embodiments, the first adapter further comprises a block at a 5' end of the long strand. In some embodiments, the first adapter further comprises a block at a 5' end of the short strand. In some embodiments, the 3' overhang comprises at least 6, 7, 8, or 9 nucleotides. In some embodiments, step d) comprises ligating the second adapter. In some embodiments, the ligating comprises blunt end ligation. In some embodiments, the polynucleotide comprising the sequence A at one end generated in step c) is end repaired prior to step d). In some embodiments, the second adapter comprises a duplex, partial duplex, or single strand comprising a duplex portion connected by a stem loop. In some embodiments, the first and/or second adapter further comprises one or more barcodes. In some embodiments, the second adapter comprises a partial duplex, wherein the partial duplex comprises a long strand hybridized to a short strand, wherein the long strand comprises the sequence B and an overhang. In some embodiments, the long strand comprises the sequence B and a 3' overhang, and wherein the short strand comprises a block at a 3' end. In some

IDT Exhibit 1020.3604

embodiments, the appending of the second adapter generates the one or more double stranded polynucleotides comprising the sequence A at one end and the sequence B at an opposite end, wherein the sequence A is at a 5' end on one end and the sequence B is at a 3' end on the opposite end. In some embodiments, the long strand comprises the sequence B and a 5' overhang, and wherein the short strand comprises a block at a 5' end. In some embodiments, the appending of the second adapter generates the one or more double stranded polynucleotides comprising the sequence A at one end and the sequence B at an opposite end, wherein the sequence A is at a 5' end on one end and the sequence B is at a 5' end on the opposite end. In some embodiments, a 3' end of the opposite end is extended using the sequence B as a template, thereby generating one or more double stranded polynucleotides comprising the sequence A at a 5' end on one end and a sequence complementary to the sequence B, B', at a 3' end on the opposite end. In some embodiments, the method further comprises amplifying the directional cDNA library, thereby generating amplified products. In some embodiments, the method further comprises an additional step of sequencing the amplified products. In some embodiments, the amplification comprises SPIA. In some embodiments, the amplification comprises a use of primers, wherein one or more of the primers comprises a barcode sequence. In some embodiments, the sequencing comprises next generation sequencing.

[0030] In one aspect, described herein is a method for generating a directional polynucleotide library, the method comprising: a) synthesizing one or morepolynucleotides from a template nucleic acid in the presence of a non-canonical nucleotide, whereby one or more polynucleotides comprising the non-canonical nucleotide are generated; b) cleaving a base portion of the non-canonical nucleotide from the one or more synthesized polynucleotides with an enzyme capable of cleaving the base portion of the non-canonical nucleotide, whereby an abasic site is generated; c) cleaving a phosphodiester backbone of the one or more polynucleotides comprising the abasic site at the abasic site, whereby one or more polynucleotides within a desired size range comprising a blocked 3' end are generated; d) appending a first adapter to a 3' end of the one or more polynucleotides comprising a blocked 3' end, wherein the first adapter comprises a sequence A, wherein the sequence A is non-hybridizable to the one or more polynucleotides comprising a blocked 3' end; ;e) extending a 3' end of the first adapter appended to the 3' end of the one or more polynucleotides comprising a blocked 3' end using the one or more polynucleotides comprising a blocked 3' end as template, wherein one or more double stranded polynucleotides comprising the sequence A at one end are generated; and f) appending a second adapter comprising a sequence B to the one or more double stranded polynucleotides comprising the sequence A at one end, wherein the sequence B is different than the sequence A and the appending generates one or more double stranded polynucleotides comprising the sequence A at one end and the sequence B at an opposite end, thereby generating the directional polynucleotide library. In some embodiments, steps (b) and (c) are performed simultaneously in the same reaction mixture. In some embodiments, the method comprises synthesizing the one or more polynucleotides from the template nucleic acid in the presence of all four canonical nucleotides and a non-canonical nucleotide, wherein the non-canonical nucleotide is

provided at a ratio suitable for generating fragments within the desired size range. In some embodiments, the one or more polynucleotides comprising the non-canonical nucleotide are synthesized by an amplification method selected from the group consisting of polymerase chain reaction (PCR), strand displacement amplification (SDA), multiple displacement amplification (MDA), rolling circle amplification (RCA), single primer isothermal amplification (SPIA), and Ribo-SPIA. In some embodiments, the one or more polynucleotides comprising the non-canonical nucleotide are synthesized by a method selected from the group consisting of reverse transcription, primer extension, limited primer extension, replication, and nick translation. In some embodiments, the first adapter further comprises a partial duplex and a 3' overhang. In some embodiments, the first adapter comprises a plurality of first adapters, wherein the random sequence on each of the plurality of first adapters is different than the random sequence on another of the plurality of first adapters, and wherein each of the plurality of first adapters comprises the sequence A. In some embodiments, the annealing results in substantially all of the plurality of first strand cDNA fragments of a desired size comprising a blocked 3' end further comprising one of the plurality of first adapters annealed the 3' end. In some embodiments, the appending comprises annealing the 3' overhang of the first adapter to the 3' end of the one or more polynucleotides comprising a blocked 3' end, wherein the annealing comprises hybridizing a random sequence at the 3' overhang to a complementary sequence present at the 3' end of the one or more polynucleotides comprising a blocked 3' end. In some embodiments, the partial duplex comprises a long strand and a short strand, wherein the long strand comprises the sequence A that forms a duplex with the short strand and the 3' overhang. In some embodiments, the short strand further comprises a block at a 3' and/or a 5' end. In some embodiments, the long strand further comprises a block at the 5' end. In some embodiments, the first adapter further comprises a stem loop, wherein the stem loop links a 5' end of a long strand of the partial duplex with a 3' end of a short strand of the partial duplex, and wherein the long strand comprises the sequence A and the 3' overhang. In some embodiments, the first adapter further comprises a block at a 5' end of the short strand. In some embodiments, the 3' overhang comprises at least 6, 7, 8, or 9 nucleotides. In some embodiments, step f) comprises ligating the second adapter. In some embodiments, the ligating comprises blunt end ligation. In some embodiments, the one or more polynucleotides comprising the sequence A at one end generated in step e) are end repaired prior to step f). In some embodiments, the second adapter comprises a duplex, partial duplex, or single strand comprising a duplex portion connected by a stem loop. In some embodiments, the first and/or second adapter further comprises one or more barcodes. In some embodiments, the second adapter comprises a partial duplex, wherein the partial duplex comprises a long strand hybridized to a short strand, wherein the long strand comprises the sequence B and an overhang. In some embodiments, the long strand comprises the sequence B and a 3' overhang, and wherein the short strand comprises a block at a 3' end. In some embodiments, the appending of the second adapter generates the one or more double stranded polynucleotides comprising the sequence A at one end and the sequence B at an opposite end, wherein the sequence A is at a 5' end on one end and the sequence B is at a 3' end on the opposite end. In some

IDT Exhibit 1020.3606

embodiments, the long strand comprises the sequence B and a 5' overhang, and wherein the short strand comprises a block at a 5' end. In some embodiments, the appending of the second adapter generates the one or more double stranded polynucleotides comprising the sequence A at one end and the sequence B at an opposite end, wherein the sequence A is at a 5' end on one end and the sequence B is at a 5' end on the opposite end. In some embodiments, a 3' end of the opposite end is extended using the sequence B as a template, thereby generating one or more double stranded polynucleotides comprising the sequence A at a 5' end on one end and a sequence complementary to the sequence B, B', at a 3' end on the opposite end. In some embodiments, the method further comprises amplifying the directional polynucleotide library, thereby generating amplified products. In some embodiments, the method further comprises an additional step of sequencing the amplified products. In some embodiments, the amplification comprises SPIA. In some embodiments, the amplification comprises a use of primers, wherein one or more of the primers comprise a barcode sequence. In some embodiments, the sequencing comprises next generation sequencing.

[0031] Also provided herein is a method for generating a directional cDNA library, the method comprising: (a) annealing one or more primers to a template RNA; (b) extending the one or more primers in the presence of a reaction mixture comprising dATP, dCTP, dGTP, dTTP, and dUTP, wherein the reaction mixture comprises a ratio of dUTP to dTTP, wherein the ratio permits incorporation of dUTP at a desired density, thereby generating a one or more first strand complementary DNAs (cDNAs) comprising dUTP incorporated at a desired density; (c) selectively cleaving the one or more first strand cDNAs comprising dUTPs incorporated at a desired density with uracil-N-glycosylase (UNG) and an agent capable of cleaving a phosphodiester backbone at an abasic site created by the UNG, wherein the cleaving generates a plurality of first strand cDNA fragments of a desired size comprising a blocked 3' end; (d) annealing a first adapter comprising a partial duplex and a 3' overhang to a 3' end of one or more of the plurality of first strand cDNA fragments comprising a blocked 3' end, wherein the first adapter comprises a sequence A, and wherein the annealing comprises hybridizing a random sequence at the 3' overhang to a complementary sequence present at the 3' end of the one or more of the plurality of first strand cDNA fragments comprising a blocked 3' end; (e) extending the 3' overhang hybridized to the complementary sequence with a DNA polymerase, wherein one or more double stranded cDNA fragments comprising the sequence A at one end is generated; (f) ligating a second adapter comprising a sequence B to the one or more double stranded cDNA fragments comprising the sequence A at one end, wherein the ligating generates one or more double stranded cDNA fragments comprising the sequence A at one end and the sequence B at an opposite end, thereby generating the directional polynucleotide library; and (g) optionally, amplifying and/or sequencing the directional cDNA library.

[0032] Also provided herein is a method for generating a directional cDNA library, the method comprising: (a) treating a template dsDNA with a nicking enzyme, wherein the treating generates one or more breaks in a phosphodiester backbone of one strand of the template dsDNA, wherein the break produces one or more 3' hydroxyls in the one strand; (b) extending the one or more 3' hydroxyls, wherein

IDT Exhibit 1020.3607

the extending is performed in the presence of a reaction mixture comprising dATP, dCTP, dGTP, dTTP, and dUTP, wherein the reaction mixture comprises a ratio of dUTP to dTTP, wherein the ratio permits incorporation of dUTP at a desired density, thereby generating one or more first strand complementary DNAs (cDNAs) comprising dUTP incorporated at a desired density; (c) selectively cleaving the one or more first strand cDNAs comprising dUTPs incorporated at a desired density with uracil-N-glycosylase (UNG) and an agent capable of cleaving a phosphodiester backbone at an abasic site created by the UNG, wherein the cleaving generates a plurality of first strand cDNA fragments of a desired size comprising a blocked 3' end; (d) annealing a first adapter comprising a partial duplex and a 3' overhang to a 3' end of one or more of the plurality of first strand cDNA fragments comprising a blocked 3' end, wherein the first adapter comprises a sequence A, and wherein the annealing comprises hybridizing a random sequence at the 3' overhang to a complementary sequence present at the 3' end of the one or more of the plurality of first strand cDNA fragments comprising a blocked 3' end; (e) extending the 3' overhang hybridized to the complementary sequence with a DNA polymerase, wherein one or more double stranded cDNA fragments comprising the sequence A at one end is generated; (f) ligating a second adapter comprising a sequence B to the one or more double stranded cDNA fragments comprising the sequence A at one end, wherein the ligating generates one or more double stranded cDNA fragments comprising the sequence A at one end and the sequence B at an opposite end thereby generating a directional cDNA library; and (g) optionally, amplifying and/or sequencing the directional cDNA library.

[0033] Also provided herein is a method for generating a whole genome library, the method comprising: (a) denaturing nicked and/or fragmented dsDNA template nucleic acid; (b) annealing a first adapter comprising a partial duplex and a 3' overhang to a 3' end of one or more of the plurality of single-stranded DNA fragments, wherein the first adapter comprises a sequence A, and wherein the annealing comprises hybridizing a random sequence at the 3' overhang to a complementary sequence present at the 3' end of the one or more of the plurality of single-stranded DNA fragments; (c) extending the 3' overhang hybridized to the complementary sequence with a DNA polymerase, wherein one or more double stranded cDNA fragments comprising the sequence A at one end is generated; (e) ligating a second adapter comprising a sequence B to the one or more double stranded cDNA fragments comprising the sequence A at one end, wherein the ligating generates one or more double stranded cDNA fragments comprising the sequence A at one end and the sequence B at an opposite end thereby generating a directional cDNA library; and (f) optionally, amplifying and/or sequencing the directional cDNA library.

[0034] In some embodiments of any of the foregoing methods, the one or more primers comprise a random primer. In some embodiments, the one or more primers comprise a sequence specific to a group of RNAs comprising substantially all transcripts. In some embodiments, the one or more primers comprise a sequence specific to a group of RNAs which does not comprise structural RNA, wherein the structural RNA comprises ribosomal RNA (rRNA). In some embodiments, the agent capable of cleaving a phosphodiester backbone comprises an enzyme, chemical agent, and/or heat. In some embodiments, the chemical agent is a polyamine. In some embodiments, the polyamine is N,N-dimethylethylenediamine

IDT Exhibit 1020.3608

(DMED). In some embodiments, the first adaptor comprises a long strand and a short strand, wherein the long strand comprises the sequence A that forms a duplex with the short strand and a 3' overhang. In some embodiments, the first adapter comprises a plurality of first adapters, wherein the random sequence on each of the plurality of first adapters is different than the random sequence on another of the plurality of first adapters, and wherein each of the plurality of first adapters comprises the sequence A. In some embodiments, the first adapter further comprises a stem loop, wherein the stem loop links a 5' end of a long strand of the partial duplex with a 3' end of a short strand of the partial duplex, and wherein the long strand comprises the sequence A and the 3' overhang. In some embodiments, the 3' overhang comprises at least 6, 7, 8, or 9 nucleotides. In some embodiments, the second adapter comprises a partial duplex, wherein the partial duplex comprises a long strand hybridized to a short strand, wherein the long strand comprises the sequence B and an overhang. In some embodiments, the long strand comprises the sequence B and a 3' overhang, and wherein the short strand comprises a block at a 3' end. In some embodiments, the ligating generates the one or more double stranded cDNA fragments comprising the sequence A at one end and the sequence B at an opposite end, wherein the sequence A is at a 5' end on one end and the sequence B is at a 3' end on the opposite end. In some embodiments, the long strand comprises the sequence B and a 5' overhang, and wherein the short strand comprises a block at a 5' end. In some embodiments, the ligating generates the one or more double stranded cDNA fragments comprising the sequence A at one end and the sequence B at an opposite end, wherein the sequence A is at a 5' end on one end and the sequence B is at a 5' end on the opposite end. In some embodiments, a 3' end of the opposite end is extended using the sequence B as a template, thereby generating one or more double stranded cDNA fragments comprising the sequence A at a 5' end on one end and a sequence complementary to the sequence B, B', at a 3' end on the opposite end. In some embodiments, the nicking enzyme comprises a strand specific nicking enzyme. In some embodiments, the extending the one or more 3' hydroxyls in step b) is performed with a DNA polymerase comprising strand displacement activity. In some embodiments, the ligating comprises blunt end ligation, wherein the one or more double stranded cDNA fragments comprising the sequence A at one end generated in step e) are end repaired prior to step f). In some embodiments, the first and/or second adapter further comprises one or more barcodes.

INCORPORATION BY REFERENCE

[0035] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

IDT Exhibit 1020.3609

BRIEF DESCRIPTION OF THE DRAWINGS

[0036] The novel features are set forth with particularity in the appended claims. A better understanding of features and advantages will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of methods, compositions, and kits provided herein are utilized, and the accompanying drawings of which:

[0037] **FIGs. 1A** and **1B** depicts methods for the generation of directional cDNA libraries from RNA templates. **FIG. 1A** depicts the generation of a directional cDNA library from an RNA template comprising strand specific products with defined sequences A and B at the 5' and 3' ends of the product, respectively. **FIG. 1B** depicts the generation of a directional cDNA library from an RNA template comprising strand specific products with defined sequences A and B' at the 5' and 3' ends of the product, respectively.

[0038] **FIG. 2** depicts first adapters comprising a 3' overhang comprising random sequence for use in the methods depicted in **FIGs. 1A** and **1B. I** depicts a first adapter comprising a 3' overhang comprising a long strand and a short single strand complementary to the 5' portion of the longer strand with blocking groups (x) at both ends. A block can also be present at the 5' end of the long strand. Any or all of the blocking groups can be optional. The ends of the oligonucleotides can be furthered protected by phosphothioate bonds. **II** depicts a first adapter comprising a 3' overhang and a stem loop oligonucleotide. The loop portion of the stem loop can comprise DNA or RNA or combinations thereof, nonnucleotide linker, nucleotide analogs, or a mixture thereof. The 5' end can also comprise a blocking group. The ends can be furthered protected by phosphothioate bonds.

[0039] **FIG. 3** depicts a workflow for generation of stranded cDNA library from an RNA template.

[0040] **FIG. 4** depicts library generation from a double stranded DNA (e.g., genomic DNA) template employing nicking enzyme(s) and a DNA polymerase in combination with the methods depicted in **FIGs. 1A** and **1B.**

[0041] **FIG. 5** depicts single primer isothermal amplification of a cDNA product generated by the methods depicted in **FIGs. 1A** and **1B.**

[0042] **FIG. 6** depicts a Bioanalyzer (Agilent) trace of a size distribution of a directional sequencing library produced from 100ng Universal Human Reference (UHR) total RNA, as described in Example 1.

[0043] **FIG. 7** depicts transcriptome sequencing data of directional sequencing libraries (s4_L2DR14; s4_L2DR15) from UHR total RNA (100ng) generated as described in Example 1.

[0044] **FIG. 8** depicts the correlation of reads per kilobase of transcript per million (RPKM) value of the transcriptome sequencing data of two directional sequencing libraries (s4_L2DR14; s4_L2DR15) from UHR total RNA (100ng) generated as described in Example 1.

[0045] **FIG. 9** depicts a summary of sequencing data obtained from three directional sequencing library generated from UHR total RNA as described in Examples 1 and 2.

[0046] **FIG. 10** depicts transcriptome sequencing data from directional sequencing libraries from UHR total RNA (1 ng) generated as described in Example 2.

IDT Exhibit 1020.3610

DETAILED DESCRIPTION

I. Overview

[0047] Provided herein are methods, compositions, and kits for the construction of directional nucleic acid sequencing libraries from nucleic acid (e.g., RNA and DNA) templates. In one aspect, provided herein are methods, compositions, and kits for generating nucleic acid libraries from RNA and DNA templates that are compatible with high throughput sequencing methods and simultaneously maintain the directional (strandedness) information of the original nucleic acid sample. The methods can be used to generate libraries representing the whole transcriptome as well as the whole genome without the need for physical fragmentation of the template genomic dsDNA. The methods can also be used to generate libraries from very small samples, including single cells.

II. Strand-Specific Selection

[0048] The compositions, methods, and kits provided herein can be used for retaining directional information for a template nucleic acid. The template nucleic acid can be a RNA or DNA. The template nucleic acid can be single-stranded or double-stranded. The terms "strand specific," "directional," or "strandedness" can refer to the ability to differentiate in a double-stranded polynucleotide between the two strands that are complementary to one another. The terms "stranded library", "stranded cDNA library", "directional library" or "directional cDNA library" can be used interchangeably. The term "strand marking" can refer to any method for distinguishing between the two strands of a double-stranded polynucleotide. The term "selection" can refer to any method for selecting between the two strands of a double-stranded polynucleotide.

[0049] Based on the methods described herein, the retention of the directionality and strand information of the nucleic acid template can be determined with greater than 50% efficiency. The efficiency of retention of directionality and strand orientation using the methods described herein can be > 50%, >55%, > 60%, >65%, >70%, >75%, > 80%, >85%, > 90%, or > 95%. The efficiency of retention of directionality and strand orientation can be > 70%, > 80%, > 90% or > 99%. The methods described herein can be used to generate directional polynucleotide libraries wherein greater than 50% of the polynucleotides in the polynucleotide library comprise a specific strand orientation. The retention of a specific strand orientation using the methods described herein can be > 50%, >55%, > 60%, >65%, >70%, >75%, > 80%, >85%, > 90%, or >95%. The retention of specific strand orientation of polynucleotides in the directional polynucleotide library can be > 99%.

III. Polynucleotides, samples, and nucleotides

[0050] The directional nucleic acid library can be generated from a nucleic acid template obtained from any source of nucleic acid. The nucleic acid can be RNA or DNA. The nucleic acid can be single-stranded or double stranded. In some cases, the nucleic acid is DNA. The DNA can be obtained and

IDT Exhibit 1020.3611

purified using standard techniques in the art and include DNA in purified or unpurified form. The DNA can be mitochondrial DNA, cell-free DNA, complementary DNA (cDNA), or genomic DNA. In some cases, the nucleic acid is genomic DNA. The DNA can be plasmid DNA, cosmid DNA, bacterial artificial chromosome (BAC), or yeast artificial chromosome (YAC). The DNA can be derived from one or more chromosomes. For example, if the DNA is from a human, the DNA can derived from one or more of chromosome 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, X, or Y. In some cases, the DNA is double-stranded DNA. In some cases, the double-stranded DNA is genomic DNA. In some cases, the DNA is cDNA. In some cases, the cDNA is double-stranded cDNA. In some cases, the cDNA is derived from RNA, wherein the RNA is subjected to first strand synthesis followed by second strand synthesis. The RNA can be obtained and purified using standard techniques in the art and include RNAs in purified or unpurified form, which include, but are not limited to, mRNAs, tRNAs, snRNAs, rRNAs, retroviruses, small non-coding RNAs, microRNAs, polysomal RNAs, pre-mRNAs, intronic RNA, viral RNA, cell free RNA and fragments thereof. The non-coding RNA, or ncRNA can include snoRNAs, microRNAs, siRNAs, piRNAs and long nc RNAs.

[0051] The source of nucleic acid for use in the methods described herein can be a sample comprising the nucleic acid. The nucleic acid can be isolated from the sample and purified by any of the methods known in the art for purifying the nucleic acid from the sample. The sample can be derived from a non-cellular entity comprising polynucleotides (e.g., a virus) or from a cell-based organism (e.g., member of archaea, bacteria, or eukarya domains). In some cases, the sample is obtained from a swab of a surface, such as a door or bench top.

[0052] The sample can be from a subject, e.g., a plant, fungi, eubacteria, archeabacteria, protest, or animal. The subject can be an organism, either a single-celled or multi-cellular organism. The subject can be cultured cells, which can be primary cells or cells from an established cell line, among others. The sample can be isolated initially from a multi-cellular organism in any suitable form. The animal can be a fish, e.g., a zebrafish. The animal can be a mammal. The mammal can be, e.g., a dog, cat, horse, cow, mouse, rat, or pig. The mammal can be a primate, e.g., a human, chimpanzee, orangutan, or gorilla. The human can be a male or female. The sample can be from a human embryo or human fetus. The human can be an infant, child, teenager, adult, or elderly person. The female can be pregnant, suspected of being pregnant, or planning to become pregnant. In some cases, the sample is a single or individual cell from a subject and the polynucleotides are derived from the single or individual cell. In some cases, the sample is an individual micro-organism, or a population of micro-organisms, or a mixture of micro-organisms and host cellular or cell free nucleic acids.

[0053] The sample can be from a subject (e.g., human subject) who is healthy. In some cases, the sample is taken from a subject (e.g., an expectant mother) at at least 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 weeks of gestation. In some cases, the subject is affected by a genetic disease, a carrier for a genetic disease or at risk for developing or passing down a genetic disease, where a genetic disease is any disease that can be linked to a genetic variation such as mutations,

IDT Exhibit 1020.3612

insertions, additions, deletions, translocation, point mutation, trinucleotide repeat disorders and/or single nucleotide polymorphisms (SNPs).

[0054] The sample can be from a subject who has a specific disease, disorder, or condition, or is suspected of having (or at risk of having) a specific disease, disorder or condition. For example, the sample can be from a cancer patient, a patient suspected of having cancer, or a patient at risk of having cancer. The cancer can be, e.g., acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), adrenocortical carcinoma, Kaposi Sarcoma, anal cancer, basal cell carcinoma, bile duct cancer, bladder cancer, bone cancer, osteosarcoma, malignant fibrous histiocytoma, brain stem glioma, brain cancer, craniopharyngioma, ependymoblastoma, ependymoma, medulloblastoma, medulloeptithelioma, pineal parenchymal tumor, breast cancer, bronchial tumor, Burkitt lymphoma, Non-Hodgkin lymphoma, carcinoid tumor, cervical cancer, chordoma, chronic lymphocytic leukemia (CLL), chromic myelogenous leukemia (CML), colon cancer, colorectal cancer, cutaneous T-cell lymphoma, ductal carcinoma in situ, endometrial cancer, esophageal cancer, Ewing Sarcoma, eye cancer, intraocular melanoma, retinoblastoma, fibrous histiocytoma, gallbladder cancer, gastric cancer, glioma, hairy cell leukemia, head and neck cancer, heart cancer, hepatocellular (liver) cancer, Hodgkin lymphoma, hypopharyngeal cancer, kidney cancer, laryngeal cancer, lip cancer, oral cavity cancer, lung cancer, non-small cell carcinoma, small cell carcinoma, melanoma, mouth cancer, myelodysplastic syndromes, multiple myeloma, medulloblastoma, nasal cavity cancer, paranasal sinus cancer, neuroblastoma, nasopharyngeal cancer, oral cancer, oropharyngeal cancer, osteosarcoma, ovarian cancer, pancreatic cancer, papillomatosis, paraganglioma, parathyroid cancer, penile cancer, pharyngeal cancer, pituitary tumor, plasma cell neoplasm, prostate cancer, rectal cancer, renal cell cancer, rhabdomyosarcoma, salivary gland cancer, Sezary syndrome, skin cancer, nonmelanoma, small intestine cancer, soft tissue sarcoma, squamous cell carcinoma, testicular cancer, throat cancer, thymoma, thyroid cancer, urethral cancer, uterine cancer, uterine sarcoma, vaginal cancer, vulvar cancer, Waldenstrom Macroglobulinemia, or Wilms Tumor. The sample can be from the cancer and/or normal tissue from the cancer patient.

[0055] The sample can be aqueous humour, vitreous humour, bile, whole blood, blood serum, blood plasma, breast milk, cerebrospinal fluid, cerumen, enolymph, perilymph, gastric juice, mucus, peritoneal fluid, saliva, sebum, semen, sweat, tears, vaginal secretion, vomit, feces, or urine. The sample can be obtained from a hospital, laboratory, clinical or medical laboratory. The sample can be taken from a subject.

[0056] The sample can be an environmental sample comprising medium such as water, soil, air, and the like. The sample can be a forensic sample (e.g., hair, blood, semen, saliva, etc.). The sample can comprise an agent used in a bioterrorist attack (e.g., influenza, anthrax, smallpox).

[0057] The sample can comprise nucleic acid. The nucleic acid can be, e.g., mitochondrial DNA, genomic DNA, mRNA, siRNA, miRNA, cRNA, single-stranded DNA, double-stranded DNA, single-stranded RNA, double-stranded RNA, tRNA, rRNA, or cDNA. The sample can comprise cell-free nucleic acid. The sample can be a cell line, genomic DNA, cell-free plasma, formalin fixed paraffin

IDT Exhibit 1020.3613

embedded (FFPE) sample, or flash frozen sample. A formalin fixed paraffin embedded sample can be deparaffinized before nucleic acid is extracted. The sample can be from an organ, e.g., heart, skin, liver, lung, breast, stomach, pancreas, bladder, colon, gall bladder, brain, etc. Nucleic acids can be extracted from a sample by means available to one of ordinary skill in the art.

[0058] The sample can be processed to render it competent for fragmentation, ligation, denaturation, and/or amplification or any of the methods provided herein. Exemplary sample processing can include lysing cells of the sample to release nucleic acid, purifying the sample (e.g., to isolate nucleic acid from other sample components, which can inhibit enzymatic reactions), diluting/concentrating the sample, and/or combining the sample with reagents for further nucleic acid processing. In some examples, the sample can be combined with a restriction enzyme, reverse transcriptase, or any other enzyme of nucleic acid processing.

[0059] The methods described herein can be used for analyzing or detecting one or more target nucleic acids. The term polynucleotide, or grammatical equivalents, can refer to at least two nucleotides covalently linked together. A polynucleotide described herein can contain phosphodiester bonds, although in some cases, as outlined below (for example in the construction of primers and probes such as label probes), nucleic acid analogs are included that can have alternate backbones, comprising, for example, phosphoramide (Beaucage et al., Tetrahedron 49(10):1925 (1993) and references therein; Letsinger, J. Org. Chem. 35:3800 (1970); Sprinzl et al., Eur. J. Biochem. 81:579 (1977); Letsinger et al., Nucl. Acids Res. 14:3487 (1986); Sawai et al, Chem. Lett. 805 (1984), Letsinger et al., J. Am. Chem. Soc. 110:4470 (1988); and Pauwels et al., Chemica Scripta 26:141 91986)), phosphorothioate (Mag et al., Nucleic Acids Res. 19:1437 (1991); and U.S. Pat. No. 5,644,048), phosphorodithioate (Briu et al., J. Am. Chem. Soc. 111:2321 (1989), O-methylphosphoroamidite linkages (see Eckstein, Oligonucleotides and Analogues: A Practical Approach, Oxford University Press), and peptide nucleic acid (also referred to herein as "PNA") backbones and linkages (see Egholm, J. Am. Chem. Soc. 114:1895 (1992); Meier et al., Chem. Int. Ed. Engl. 31:1008 (1992); Nielsen, Nature, 365:566 (1993); Carlsson et al., Nature 380:207 (1996), all of which are incorporated by reference). Other analog nucleic acids include those with bicyclic structures including locked nucleic acids (also referred to herein as "LNA"), Koshkin et al., J. Am. Chem. Soc. 120.13252 3 (1998); positive backbones (Denpcy et al., Proc. Natl. Acad. Sci. USA 92:6097 (1995); non-ionic backbones (U.S. Pat. Nos. 5,386,023, 5,637,684, 5,602,240, 5,216,141 and 4,469,863; Kiedrowshi et al., Angew. Chem. Intl. Ed. English 30:423 (1991); Letsinger et al., J. Am. Chem. Soc. 110:4470 (1988); Letsinger et al., Nucleoside & Nucleotide 13:1597 (1994); Chapters 2 and 3, ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research", Ed. Y. S. Sanghui and P. Dan Cook; Mesmaeker et al., Bioorganic & Medicinal Chem. Lett. 4:395 (1994); Jeffs et al., J. Biomolecular NMR 34:17 (1994); Tetrahedron Lett. 37:743 (1996)) and non-ribose backbones, including those described in U.S. Pat. Nos. 5,235,033 and 5,034,506, and Chapters 6 and 7, ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research", Ed. Y. S. Sanghui and P. Dan Cook. Nucleic acids containing one or more carbocyclic sugars are also included within the definition of nucleic acids

IDT Exhibit 1020.3614

(see Jenkins et al., Chem. Soc. Rev. (1995) pp 169 176). Several nucleic acid analogs are described in Rawls, C & E News Jun. 2, 1997 page 35. "Locked nucleic acids" are also included within the definition of nucleic acid analogs. LNAs are a class of nucleic acid analogues in which the ribose ring is "locked" by a methylene bridge connecting the 2'-O atom with the 4'-C atom. All of these references are hereby expressly incorporated by reference. These modifications of the ribose-phosphate backbone can be done to increase the stability and half-life of such molecules in physiological environments. For example, PNA:DNA and LNA-DNA hybrids can exhibit higher stability and thus can be used in some cases. The nucleic acids can be single stranded or double stranded, as specified, or contain portions of both double stranded or single stranded sequence. Depending on the application, the nucleic acids can be DNA (including, e.g., genomic DNA, mitochondrial DNA, and cDNA), RNA (including, e.g., mRNA and rRNA) or a hybrid, where the nucleic acid contains any combination of deoxyribo- and ribo-nucleotides, and any combination of bases, including uracil, adenine, thymine, cytosine, guanine, inosine, xathanine hypoxathanine, isocytosine, isoguanine, etc.

[0060] The term “unmodified nucleotide” or “unmodified dNTP” or “classic dNTP” can refer to the four deoxyribonucleotide triphosphates dATP (deoxyadenosine triphosphate), dCTP (deoxycytidine triphosphate), dGTP (deoxyguanosine triphosphate) and dTTP (deoxythymidine triphosphate) that can normally used as building blocks in the synthesis of DNA.

[0061] The term “canonical dNTP” or “canonical nucleotide” can be used to refer to the four deoxyribonucleotide triphosphates dATP, dCTP, dGTP and dTTP that are normally found in DNA.

[0062] The term “modified nucleotide,” “modified dNTP,” or “nucleotide analog,” can refer to any molecule suitable for substituting one corresponding unmodified nucleotide or classic dNTP. Such modified nucleotides must be able to undergo a base pair matching identical or similar to the classic or unmodified dNTP it replaces. The modified nucleotide or dNTP must be suitable for specific degradation or cleavage in which it is selectively degraded or cleaved by a suitable degrading or cleavage agent. The modified nucleotide must mark the DNA strand containing the modified nucleotide eligible for selective removal or cleavage or facilitate separation of the polynucleotide strands. Such a removal or cleavage or separation can be achieved by molecules, particles or enzymes interacting selectively with the modified nucleotide, thus selectively removing or marking for removal or cleaving only one polynucleotide strand.

[0063] The term “non-canonical” can refer to nucleic acid bases in DNA other than the four canonical bases in DNA, or their deoxyribonucleotide or deoxyribonucleoside analogs. Although uracil is a common nucleic acid base in RNA, uracil is a non-canonical base in DNA. In some cases, the non-canonical dNTP is dUTP.

[0064] The term "barcode" can refer to a known nucleic acid sequence that allows some feature of a nucleic acid with which the barcode is associated to be identified. In some cases, the feature of the nucleic acid to be identified is the sample from which the nucleic acid is derived. In some cases, barcodes are at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or more nucleotides in length. In some cases, barcodes are shorter than 10, 9, 8, 7, 6, 5, or 4 nucleotides in length. An oligonucleotide (e.g., primer or

adapter) can comprise about, more than, less than, or at least 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 different barcodes. Barcodes can be associated (e.g., via annealing or ligation) with template nucleic acids derived from a sample comprising the template nucleic acids. In some cases, barcodes associated with template nucleic acids derived from one sample are different than barcodes associated with template nucleic acids derived from another sample. The barcodes associated with template nucleic acids derived from a first sample can be of different length than barcodes associated with template nucleic acids derived from a second sample. Barcodes can be of sufficient length and comprise sequences that can be sufficiently different to allow the identification of samples based on barcodes with which they are associated. In some cases, a barcode, and the sample source with which it is associated, can be identified accurately after the mutation, insertion, or deletion of one or more nucleotides in the barcode sequence, such as the mutation, insertion, or deletion of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more nucleotides. In some cases, each barcode in a plurality of barcodes differ from every other barcode in the plurality at at least three nucleotide positions, such as at least 3, 4, 5, 6, 7, 8, 9, 10, or more positions. In some cases, an adapter comprises at least one of a plurality of barcode sequences. In some cases, barcodes for a second adapter oligonucleotide are selected independently from barcodes for a first adapter/primer oligonucleotide. In some cases, first adapter/primer oligonucleotides and second adapter oligonucleotides having barcodes are paired, such that adapters of the pair comprise the same or different one or more barcodes. In some cases, the methods described herein further comprise identifying the sample from which a template nucleic acid is derived based on a barcode sequence to which the target nucleic acid is joined. A barcode can comprise a polynucleotide sequence that when joined to a template nucleic acid serves as an identifier of the sample from which the template nucleic acid was derived.

[0065] In some cases, the barcodes comprise a random sequence that is useful for uniquely marking each individual fragment within a sample comprising a plurality of nucleic acid fragments. The uniquely appended barcode provides a means of quantification of the unique fragments during downstream quantification procedures such as massively parallel next generation sequencing. The barcodes can be part of any adapter and/or primer useful in the methods described herein and thereby be appended to an individual fragment or plurality of fragments by the methods provided herein. In these cases, the barcodes are appended at random and are unique for the fragments to which they are appended rather than the sample. These barcodes can be combined with barcodes that are specific for the sample, or the source of the nucleic acid.

[0066] Conditions that "allow" or "permit" an event to occur or conditions that are "suitable" for an event to occur, such as polynucleotide synthesis, cleavage of a base portion of a non-canonical nucleotide, cleavage of a phosphodiester backbone at an abasic site, and the like, or "suitable" conditions are conditions that do not prevent such events from occurring. Thus, these conditions permit, enhance, facilitate, and/or are conducive to the event. Such conditions, known in the art and described herein, depend upon, for example, the nature of the polynucleotide sequence, temperature, and buffer conditions.

IDT Exhibit 1020.3616

These conditions also depend on what event is desired, such as polynucleotide synthesis, cleavage of a base portion of a non-canonical nucleotide, cleavage of a phosphodiester backbone at an abasic site, etc.

IV. Synthesis of polynucleotides comprising a non-canonical nucleotide

[0067] A polynucleotide comprising a non-canonical nucleotide can be produced by synthesizing a polynucleotide from a template nucleic acid in the presence of at least one non-canonical nucleotide, whereby a polynucleotide comprising a non-canonical nucleotide is generated. The frequency of incorporation of non-canonical nucleotides into the polynucleotide (e.g., first strand cDNA) relates to the size of fragment produced using the methods provided herein because the spacing between non-canonical nucleotides in the polynucleotide comprising a non-canonical nucleotide, along with the reaction conditions used, can determine the approximate size of the fragments resulting from generation of an abasic site from the non-canonical nucleotide and cleavage of the backbone at the abasic site, as described herein. The desired size range of the fragments can be varied according to the requirements of downstream applications, such as generation of sequencing library suitable for massively parallel sequencing.

[0068] The polynucleotides generated by the methods provided herein can be DNA or complementary DNA (cDNA), wherein the cDNA is complementary to a template nucleic acid, though, as noted herein, a polynucleotide can comprise altered and/or modified nucleotides, internucleotide linkages, ribonucleotides, etc..

[0069] Methods for synthesizing polynucleotides (e.g., single and double stranded DNA) from a template nucleic acid are well known in the art, and include, but is not limited to, single primer isothermal amplification (SPIA™), Ribo-SPIA™, PCR, reverse transcription, primer extension, limited primer extension, replication (including rolling circle replication), strand displacement amplification (SDA), nick translation, multiple displacement amplification (MDA), rolling circle amplification (RCA) and, *e.g.*, any method that results in synthesis of the complement of a template nucleic acid sequence such that at least one non-canonical nucleotide can be incorporated into a polynucleotide. *See, e.g.,* Kurn, U.S. Patent No. 6,251,639; Kurn, WO 02/00938; Kurn, U.S. Patent No. 6,946,251, Kurn, U.S. Patent No. 6,692,918 ; Mullis, U.S. Patent No. 4,582,877; Wallace, U.S. patent No. 6,027,923; U.S. Patent No. 5,508,178; 5,888,819; 6,004,744; 5,882,867; 5,710,028; 6,027,889; 6,004,745; 5,763,178; 5,011,769; *see also* Sambrook (1989) "Molecular Cloning: A Laboratory Manual", second edition; Ausebel (1987, and updates) "Current Protocols in Molecular Biology", Mullis, (1994) "PCR: The Polymerase Chain Reaction". One or more methods known in the art can be used to generate a polynucleotide comprising a non-canonical nucleotide. It is understood that the polynucleotide comprising a non-canonical nucleotide can be single stranded or double stranded or partially double stranded, and that one or both strands of a double stranded polynucleotide can comprise a non-canonical nucleotide. For convenience, "DNA" can be used herein to describe (and exemplify) a polynucleotide. A DNA, and, thus, a polynucleotide can be a complementary DNA (cDNA) generated by producing a nucleotide strand complementary to a template

IDT Exhibit 1020.3617

nucleic acid (e.g., a cDNA produced by first and/or second strand synthesis from an RNA template or a cDNA produced from an extension or replication reaction using a template DNA). Suitable methods include methods that result in one single- or double-stranded polynucleotide comprising a non-canonical nucleotide (for example, reverse transcription, production of double stranded cDNA, a single round of DNA replication), as well as methods that result in multiple single stranded or double stranded copies or copies of the complement of a template (for example, single primer isothermal amplification or Ribo-SPIA™ or PCR). In some cases, a single-stranded polynucleotide comprising a non-canonical nucleotide is synthesized using single primer isothermal amplification. *See* Kurn, U.S. Patent Nos. 6,251,639 and 6,692,918.

[0070] A polynucleotide comprising a non-canonical nucleotide can be generated from a template in the presence of all four canonical nucleotides and at least one non-canonical nucleotide under reaction conditions suitable for synthesis of polynucleotides, including suitable enzymes and primers, if necessary. Reaction conditions and reagents, including primers, for synthesizing a polynucleotide comprising a non-canonical nucleotide are known in the art, and further discussed herein. Suitable non-canonical nucleotides are well-known in the art, and include: deoxyuridine triphosphate (dUTP), deoxyinosine triphosphate (dITP), 5-hydroxymethyl deoxycytidine triphosphate (5-OH-Me-dCTP). *See, e.g.,* Jendrisak, U.S. Patent No. 6,190,865 B1; *Mol. Cell Probes* (1992) 251-6. Two or more different non-canonical nucleotides can be incorporated into the polynucleotide synthesized from the template nucleic acid by a DNA polymerase as provided herein, whereby a polynucleotide comprising at least two different non-canonical nucleotides can be generated.

[0071] In some cases, a polynucleotide comprising a non-canonical nucleotide is generated by reverse transcription from a template nucleic acid or a plurality of template nucleic acids in the presence of a non-canonical nucleotide as provide herein, wherein the template nucleic acid is RNA. In some cases, a polynucleotide comprising a non-canonical nucleotide is generated by a second strand synthesis reaction in the presence of a non-canonical nucleotide as provide herein using a first strand cDNA generated by reverse transcription from a template nucleic acid, wherein the template nucleic acid is RNA. In some cases, a primer used for reverse transcription comprises a random primer, wherein the random primer comprises random sequence directed against one or more RNA templates. In some cases, a primer used for reverse transcription comprises a sequence specific to a target RNA or group of RNAs. The group of RNAs can comprise substantially all transcripts. The group of RNAs targeted can be all RNAs except structural RNA, e.g. ribosomal RNA (rRNA). In some cases, a primer used for second strand synthesis comprises a random primer, wherein the random primer comprises random sequence directed against one or more RNA templates used for first strand cDNA synthesis. In some cases, a primer used for second strand synthesis comprises a sequence specific to a target RNA or group of RNAs used for first strand cDNA synthesis. The group of RNAs can comprise substantially all transcripts. The group of RNAs targeted can be all RNAs except structural RNA, e.g., ribosomal RNA (rRNA). In some cases, the primer

IDT Exhibit 1020.3618

or primers used for synthesis of either first or second strand cDNA, or both, can be designed to hybridize to specific targets on the polynucleotide template or templates.

[0072] In some cases, a polynucleotide comprising a non-canonical nucleotide is generated by a primer extension reaction from a template nucleic acid in the presence of a non-canonical nucleotide as provided herein, wherein the template nucleic acid is DNA. The DNA can be a dsDNA. The dsDNA can be denatured by any method known in the art prior to the primer extension reaction. The primer can comprise random sequence or sequence directed against a specific target sequence or groups of sequences. In some cases, the polynucleotide comprising a non canonical nucleotide is generated by extension from a nick or break in the phosphodiester backbone of one strand in a dsDNA. It is understood that while a single template nucleic acid is used for simplicity, the primer extension reaction can be performed on one or more template nucleic acids or a mixture thereof, thereby generating a one or more products from the primer extension reaction.

[0073] In some cases, a polynucleotide comprising a non-canonical nucleotide is generated by a strand displacement amplification reaction from a template nucleic acid, or a plurality of template nucleic acids, in the presence of non-canonical nucleotides as provide herein, wherein the template nucleic acid is DNA. The DNA can be a dsDNA generated by any of the methods described herein or genomic DNA. The dsDNA can be treated with a nicking enzyme or endonuclease. The nicking enzyme can produce a break in the phosphodiester backbone of one strand in a dsDNA template (e.g. genomic DNA), thereby generating a free 3' hydroxyl (OH). The free 3' OH can be extended using a DNA dependent DNA polymerase comprising strand displacement activity as provided herein, wherein the other strand of the dsDNA template can be used as template. The nicking enzyme can be strand specific or non-strand specific. The nicking enzyme or endonuclease for use in the methods provided herein can include any nicking enzyme known in the art, including those provided by New England Biolabs. Examples of nicking endonucleases include, but are not limited to, top strand cleaving Nt.AlwI, Nt.BbvCI, Nt.BstNBI, Nt.SapI, or Nt.CviPII, or bottom strand cleaving Nb.BbvCI, Nb.BsmI, or Nb.BsrDI. A nicking endonuclease can be, e.g., Nt.BspQI, Nt.BsmAI, or Nb.Mva1269I.

[0074] **FIG. 4** depicts an exemplary method using strand displacement amplification to generate a polynucleotide comprising a non-canonical nucleotide from a genomic DNA template. Double stranded DNA (genomic DNA) is treated with a nicking enzyme to produce nicks (e.g., one or more) in one strand of the dsDNA template. The nicks in the one strand of the dsDNA following treatment with a nicking enzyme can thereby produce one or more 3' hydroxyls (OHs). Optionally, the nicking enzyme can be sense selective, thereby maintaining the strandedness of the template DNA. The dsDNA comprising nicks (e.g. one or more) in one strand can then be treated with a DNA polymerase comprising strand displacement activity in the presence of a reaction mixture comprising all four dNTPs (e.g. dATP, dTTP, dCTP, and dGTP), and a non-canonical nucleotide (e.g., dUTP), wherein the DNA polymerase can use the one or more 3' OHs produced by the nicking enzyme to perform an extension reaction using the other, or non-nicked, strand of the dsDNA as template, thereby generating single stranded products or

IDT Exhibit 1020.3619

polynucleotides (e.g., one or more or a plurality) comprising uracil bases. The single stranded products or polynucleotides comprising uracil bases can then be treated with UDG in combination with heat or a polyamine (DMED) as provided herein to generate multiple or a plurality of single stranded polynucleotides comprising a block at the 3' end. The frequency of incorporation of dUTP into the single stranded products comprising uracil bases can be controlled as provided herein in order that multiple fragments comprising 3' end blocks are generated following treatment with a cleavage agent (e.g., UDG and heat or DMED).

[0075] Conditions for limited and/or controlled incorporation of a non-canonical nucleotide are known in the art. *See, e.g.,* Jendrisak, U.S. Patent No. 6,190,865 B1; *Mol. Cell Probes* (1992) 251-6; *Anal. Biochem.* (1993) 211:164-9; *see also* Sambrook (1989) "Molecular Cloning: A Laboratory Manual", second edition; Ausebel (1987, and updates) "Current Protocols in Molecular Biology". The frequency (or spacing) of non-canonical nucleotides in the resulting polynucleotide comprising a non-canonical nucleotide, and thus the average size of fragments generated using the methods provided herein (i.e., following cleavage of a base portion of a non-canonical nucleotide, and cleavage of a phosphodiester backbone at a non-canonical nucleotide), can be controlled by variables known in the art, including: frequency of nucleotide(s) corresponding to the non-canonical nucleotide(s) in the template (or other measures of nucleotide content of a sequence, such as average G-C content), ratio of canonical to non-canonical nucleotide present in the reaction mixture; ability of the polymerase to incorporate the non-canonical nucleotide, relative efficiency of incorporation of non-canonical nucleotide verses canonical nucleotide, and the like. The average fragmentation size can also relate to the reaction conditions used during fragmentation, as provided herein. The reaction conditions can be empirically determined, for example, by assessing average fragment size generated using the methods provided herein.

[0076] The methods for generating polynucleotides comprising a non-canonical nucleotide as provided herein can be used to incorporate a non-canonical nucleotide exactly, more than, less than, at least, at most, or about every 5, 10, 15, 20, 25, 30, 40, 50, 65, 75, 85, 100, 123, 150, 175, 200, 225, 250, 300, 350, 400, 450, 500, 550, 600, or 650 nucleotides apart in the resulting polynucleotide comprising a non-canonical nucleotide. The non-canonical nucleotide can be incorporated about every 200 nucleotides, about every 100 nucleotide, or about every 50 nucleotides. The non-canonical nucleotide can be incorporated about every 50 to about 200 nucleotides. In some cases, a 1:5 ratio of dUTP and dTTP is used in the reaction mixture. Other exemplary ratios can be exactly, about, more than, less than, at least, or at most 1:1, 1:2, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9, 1:10, 1:15, 1:20, or 1:50 dUTP to dTTP.

[0077] A template nucleic acid (along which a polynucleotide comprising a non-canonical nucleotide is synthesized) can be any template nucleic acid from any source. A template nucleic acid includes double-stranded, partially double-stranded, and single-stranded nucleic acids from any source in purified or unpurified form, which can be DNA (dsDNA and ssDNA) or RNA, including tRNA, mRNA, rRNA, mitochondrial DNA and RNA, chloroplast DNA and RNA, DNA-RNA hybrids, or mixtures thereof, genes, chromosomes, plasmids, the genomes of biological material such as microorganisms, e.g., bacteria,

IDT Exhibit 1020.3620

yeasts, viruses, viroids, molds, fungi, plants, animals, humans, and fragments thereof. Obtaining and purifying nucleic acids use standard techniques in the art. RNAs can be obtained and purified using standard techniques in the art. A DNA template (including genomic DNA template) can be transcribed into RNA form, which can be achieved using methods disclosed in Kurn, U.S. Patent No. 6,251,639 B1, and by other techniques (such as expression systems) known in the art. RNA copies of genomic DNA would generally include untranscribed sequences generally not found in mRNA, such as introns, regulatory and control elements, etc. DNA copies of an RNA template can be synthesized using methods described in Kurn, U.S. Patent No. 6,946,251 or other techniques known in the art). Synthesis of polynucleotide comprising a non-canonical nucleotide from a DNA-RNA hybrid can be accomplished by denaturation of the hybrid to obtain a ssDNA and/or RNA, cleavage with an agent capable of cleaving RNA from an RNA/DNA hybrid, and other methods known in the art. In some cases, the template RNA is cleaved simultaneously with the fragmentation of the synthesized polynucleotide comprising the non-canonical nucleotide. The template can be only a minor fraction of a complex mixture such as a biological sample and can be obtained from various biological material by procedures well known in the art. The template can be known or unknown and can contain more than one desired specific nucleic acid sequence of interest, each of which can be the same or different from each other. Therefore, the methods provided herein can be useful not only for producing one specific polynucleotide comprising a non-canonical nucleotide, but also for producing simultaneously a plurality of different specific polynucleotides comprising a non-canonical nucleotide. The template DNA can be a sub-population of nucleic acids, for example, a subtractive hybridization probe, total genomic DNA, restriction fragments, a cDNA library, cDNA prepared from total mRNA, a cloned library, or amplification products of any of the templates described herein. In some cases, the initial step of the synthesis of the complement of a portion of a template nucleic acid sequence is template denaturation. The denaturation step can be thermal denaturation or any other method known in the art, such as alkali treatment. In other cases, the initial step of the synthesis of the complement or a portion of a template nucleic acid sequence is a nicking step. Nicking of a double stranded template can be carried out by an enzymatic reaction or by physical or chemical means.

[0078] A polynucleotide, or first strand cDNA, comprising a non-canonical nucleotide (e.g., dUTP) is described as a single nucleic acid. It is understood that the polynucleotide can be a single polynucleotide, or a population of polynucleotides (from a few to a multiplicity to a very large multiplicity of polynucleotides). It is further understood that a polynucleotide comprising a non-canonical nucleotide can be a multiplicity or plurality (from small to very large) of different polynucleotide molecules. Such populations can be related in sequence (e.g., member of a gene family or superfamily) or extremely diverse in sequence (e.g., generated from all mRNA, generated from all genomic DNA, etc.). Polynucleotides can also correspond to a single sequence (which can be part or all of a known gene, for example a coding region, genomic portion, etc.). Methods, reagents, and reaction conditions for

IDT Exhibit 1020.3621

generating specific polynucleotide sequences and multiplicities or pluralities of polynucleotide sequences are known in the art.

[0079] Suitable methods of synthesis of a polynucleotide comprising a non-canonical nucleotide can be template-dependent (in the sense that polynucleotide comprising a non-canonical nucleotide is synthesized along a nucleic acid template, as generally described herein). It is understood that non-canonical nucleotides can be incorporated into a polynucleotide as a result of template-independent methods. For example, one or more primer(s) can be designed to comprise one or more non-canonical nucleotides. *See, e.g.,* Richards, U.S. Patent Nos. 6,037,152, 5,427,929, and 5,876,976. Inclusion of a non-canonical nucleotide in a primer may be particularly suitable for methods such as single primer isothermal amplification. *See* Kurn, U.S. Patent No. 6,251,639 B1; Kurn, WO 02/00938; Kurn, U.S. Patent Publication No. 2003/0087251 A1. Non-canonical nucleotide(s) can also be added to a polynucleotide by template-independent methods such as tailing or ligation of a second polynucleotide comprising a non-canonical nucleotide. Methods for tailing and ligation are well-known in the art.

V. Generating directional libraries from first strand cDNA

Cleaving a base portion of a non-canonical nucleotide to create an abasic site

[0080] In some cases, a polynucleotide comprising a non-canonical nucleotide is treated with an agent, such as an enzyme, capable of generally, specifically, or selectively cleaving a base portion of the non-canonical nucleotide to create an abasic site. As used herein, "abasic site" encompasses any chemical structure remaining following removal of a base portion (including the entire base) with an agent capable of cleaving a base portion of a nucleotide, *e.g.*, by treatment of a non-canonical nucleotide (present in a polynucleotide chain) with an agent (*e.g.*, an enzyme, acidic conditions, or a chemical reagent) capable of effecting cleavage of a base portion of a non-canonical nucleotide. In some embodiments, the agent (such as an enzyme) catalyzes hydrolysis of the bond between the base portion of the non-canonical nucleotide and a sugar in the non-canonical nucleotide to generate an abasic site comprising a hemiacetal ring and lacking the base (interchangeably called "AP" site), though other cleavage products are contemplated for use in the methods provided herein. Suitable agents and reaction conditions for cleavage of base portions of non-canonical nucleotides are known in the art, and include: N-glycosylases (also called "DNA glycosylases" or "glycosidases") including Uracil N-Glycosylase ("UNG"; specifically cleaves dUTP) (interchangeably termed "uracil DNA glyosylase"), hypoxanthine-N-Glycosylase, and hydroxy-methyl cytosine-N-glycosylase; 3-methyladenine DNA glycosylase, 3- or 7- methylguanine DNA glycosylase, hydroxymethyluracil DNA glycosylase; T4 endonuclease V. *See, e.g.*, Lindahl, *PNAS* (1974) 71(9):3649-3653; Jendrisak, U.S. Patent No. 6,190,865 B1. In some cases, UNG is used to cleave a base portion of the dUTP incorporation in polynucleotides generated by the methods provided herein.

[0081] The cleavage of base portions of non-canonical nucleotides present in polynucleotides comprising non-canonical nucleotides generated by the methods provided herein can be general, specific or selective cleavage, in the sense that the agent (such as an enzyme) capable of cleaving a base portion of

IDT Exhibit 1020.3622

a non-canonical nucleotide generally, specifically or selectively cleaves the base portion of a particular non-canonical nucleotide, whereby greater than about 98%, about 95%, about 90%, about 85%, or about 80% of the base portions cleaved are base portions of non-canonical nucleotides. However, the extent of cleavage can be less. Thus, reference to specific cleavage is exemplary. The general, specific or selective cleavage can be desirable for control of the fragment size in the methods provided herein for generating polynucleotide fragments comprising a block at the 3' end (i.e., the fragments generated by cleavage of the backbone at an abasic site). The reaction conditions can be selected such that the reaction in which the abasic site(s) are created can run to completion.

[0082] A polynucleotide comprising a non-canonical nucleotide as generated by the methods provided herein can be purified following synthesis of the polynucleotide with the non-canonical nucleotide (to eliminate, for example, residual free non-canonical nucleotides that can be present in the reaction mixture). In some cases, there is no intermediate purification between the synthesis of the polynucleotide comprising the non-canonical nucleotide and subsequent steps (such as cleavage of a base portion of the non-canonical nucleotide and cleavage of a phosphodiester backbone at the abasic site).

[0083] As noted herein, for convenience, cleavage of a base portion of a non-canonical nucleotide (whereby an abasic site is generated) has been described as a separate step. It is understood that this step can be performed simultaneously with synthesis of the polynucleotide comprising a non-canonical nucleotide (as provided herein), and cleavage of the backbone at an abasic site (fragmentation). It is further understood that the step of synthesis of a polynucleotide comprising a non-canonical nucleotide and the cleavage of the non-canonical nucleotide to generate an abasic site can be done simultaneously, while the cleavage of the backbone at the abasic site can be performed in a follow-up step. The cleavage of the backbone at the abasic site can be performed simultaneously with a step comprising degradation of the template nucleic acid or the two steps can be carried out sequentially.

[0084] It is understood that the choice of non-canonical nucleotide can dictate the choice of enzyme to be used to cleave the base portion of that non-canonical nucleotide, to the extent that particular non-canonical nucleotides are recognized by particular enzymes that are capable of cleaving a base portion of the non-canonical nucleotide. The choice of the at least one non-canonical nucleotide can be further dictated by the efficiency of incorporation into the synthesized polynucleotide comprising the non-canonical nucleotide by the DNA polymerase used.

Cleaving the backbone at or near the abasic site to generate a polynucleotide fragment

[0085] The backbone of the polynucleotide comprising an abasic site as generated by the methods provided herein can be cleaved at or near the abasic site with an agent that generates a polynucleotide fragment with a blocked 3' end. It is understood that cleavage of the base portion of a nucleotide to create an abasic site and cleavage of the polynucleotide backbone can be performed simultaneously. For convenience, however, these reactions are described as separate steps.

IDT Exhibit 1020.3623

[0086] Following generation of an abasic site by cleavage of the base portion of a nucleotide, for example, a non-canonical nucleotide present in the polynucleotide as generated herein, the backbone of the polynucleotide can be cleaved at or near the abasic site, for example, the site of incorporation of a non-canonical nucleotide (also termed the abasic site, following cleavage of the base portion of the non-canonical nucleotide), with an agent capable of effecting cleavage of the backbone at the abasic site to generate a polynucleotide fragment comprising a blocked 3' end. Cleavage of the polynucleotide backbone (also termed "fragmentation") can result in at least two fragments (depending on the number of abasic sites present in the polynucleotide comprising an abasic site, and the extent of cleavage), one of which does not comprise a blocked 3' end.

[0087] Suitable agents (for example, an enzyme, a chemical and/or reaction conditions such as heat) capable of cleavage of the backbone at an abasic site to generate a polynucleotide fragment with a blocked 3' end are well known in the art, and include: heat treatment and/or chemical treatment (including basic conditions, acidic conditions, alkylating conditions, or amine mediated cleavage of abasic sites, (*see e.g.,* McHugh and Knowland, *Nucl. Acids Res.* (1995) 23(10):1664-1670; *Bioorgan. Med. Chem* (1991) 7:2351; Sugiyama, *Chem. Res. Toxicol.* (1994) 7: 673-83; Horn, *Nucl. Acids. Res.*, (1988) 16:11559-71). As used herein, "agent" or "cleavage agent" encompasses reaction conditions such as heat. In some cases, cleavage is with a polyamine, such as N, N'-dimethylethylenediamine (DMED). *See, e.g.* McHugh and Knowland, *supra*. In some cases cleavage is with a combination of enzymes. An example of a combination of enzymes for use in the methods provided herein is USER (combination of UNG and endonuclease VIII from New England Biolabs).

[0088] The cleavage can be between the nucleotide immediately 3' to the abasic residue and the abasic residue. As is well known in the art, cleavage can be 3' to the abasic site (e.g., cleavage between the deoxyribose ring and 3' -phosphate group of the abasic residue and the deoxyribose ring of the adjacent nucleotide, generating a free 5' phosphate group on the deoxyribose ring of the adjacent nucleotide), such that an abasic site is located at the 3' end of the resulting fragment. Treatment under basic conditions or with amines (such as N, N'-dimethylethylenediamine) can result in cleavage of the phosphodiester backbone immediately 3' to the abasic site to produce a polynucleotide fragment with a blocked 3' end. In addition, more complex forms of cleavage are also possible, for example, cleavage such that cleavage of the phosphodiester backbone and cleavage of (a portion of) the abasic nucleotide results. For example, under certain conditions, cleavage using chemical treatment and/or thermal treatment can comprise a β-elimination step which results in cleavage of a bond between the abasic site deoxyribose ring and its 3' phosphate, generating a reactive α,β-unsaturated aldehyde which can be labeled or can undergo further cleavage and cyclization reactions. *See, e.g.* Sugiyama, *Chem. Res. Toxicol.* (1994) 7: 673-83; Horn, *Nucl. Acids. Res.*, (1988) 16:11559-71. It is understood that more than one method of cleavage can be used, including two or more different methods which result in multiple, different types of cleavage products comprising blocked 3' ends.

IDT Exhibit 1020.3624

[0089] The cleavage of the backbone at an abasic site can be general, specific or selective cleavage, whereby greater than about 98%, about 95%, about 90%, about 85%, or about 80% of the cleavage is at an abasic site. However, extent of cleavage can be less. Thus, reference to specific cleavage is exemplary. General, specific or selective cleavage can be desirable for control of the fragment size in the methods of generating polynucleotide fragments comprising blocked 3' ends for the generation of directional polynucleotide libraries as provided herein. The reaction conditions can be selected such that the cleavage reaction is performed in the presence of a large excess of reagents and allowed to run to completion with minimal concern about excessive cleavage of the polynucleotide (i.e., while retaining a desired fragment size, which can be determined by spacing of incorporated non-canonical nucleotides, during the synthesis step, above). The extent of cleavage can be less, such that polynucleotide fragments can be generated comprising an abasic site at an end and an abasic site(s) within or internal to the polynucleotide fragment (i.e., not at an end).

[0090] As noted herein, in embodiments in which an abasic site is generated by cleavage of a base portion of a non-canonical nucleotide in a polynucleotide synthesized in the presence of a non-canonical nucleotide, the frequency of incorporation of non-canonical nucleotides into the polynucleotide relates to the size of fragment produced using the methods provided herein because the spacing between non-canonical nucleotides in the polynucleotide comprising a non-canonical nucleotide, as well as the reaction conditions selected, determines the approximate size of the resulting fragments (following cleavage of a base portion of a non-canonical nucleotide, whereby an abasic site is generated, and cleavage of the backbone at the abasic site as described herein). It is generally desired to affect complete cleavage of the backbone at the abasic site(s) so as to generate fragments that are devoid of abasic sites when the fragments serve as a template for second strand synthesis so as to enable polymerase activity along the entire fragment target with high efficiency and fidelity.

[0091] For the methods provided herein for generating directional polynucleotide libraries, suitable fragment sizes can be exactly, greater than, less than, at least, at most, or about 5, 10, 15, 20, 25, 30, 40, 50, 65, 75, 85, 100, 123, 150, 175, 200, 225, 250, 300, 350, 400, 450, 500, 550, 600, 650 nucleotides in length. In some cases, the fragment can be about 200 nucleotides, about 100 nucleotides, or about 50 nucleotides in length. In other cases, the size of a population of fragments can be about 50 to 200 nucleotides. It is understood that the fragment size is approximate, particularly when populations of fragments are generated, because the incorporation of a non-canonical nucleotide (which relates to the fragment size following cleavage) can vary from template to template, and also between copies of the same template. Thus, fragments generated from same starting material (such as a single polynucleotide template) may have different (and/or overlapping) sequence, while still having the same approximate size or size range.

[0092] Following cleavage of the polynucleotide backbone at the abasic site, every fragment can comprise one abasic site (if cleavage is completely efficient), except for the 3'-most fragment, which can lack an abasic site. All other fragments can comprise a 3' abasic site (a blocked 3' end). In some cases,

IDT Exhibit 1020.3625

fragmentation of the backbone of the first strand cDNA or polynucleotide at the abasic site as generated by the methods provided herein can generate fragments comprising a blocked 3'-end, and a phosphate at the 5'-end.

Polymerase extension of an adapter appended to a polynucleotide fragment.

[0093] In some cases, an oligonucleotide is appended to a 3' end of a polynucleotide comprising a blocked 3' end, and optionally a 5' phosphate, prepared by the methods provided herein. The oligonucleotide can be appended by annealing single stranded DNA present at a 3' end of the oligonucleotide to the 3' end of the polynucleotide comprising a blocked 3' end. In some cases, a polynucleotide with a blocked 3' end, and optionally a 5' phosphate, prepared by the methods provided herein is hybridized to an oligonucleotide comprising an overhang with a 3' hydroxyl (OH) group and extended from the 3' OH group of the oligonucleotide with a template dependent polymerase, wherein the overhang with a 3' OH anneals to the 3' end of the polynucleotide fragments. The oligonucleotide can be an adapter or primer. The oligonucleotide can comprise DNA, RNA, or a combination thereof. The oligonucleotide can be about, less than about, or more than about 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100, or 200 nucleotides in length. The oligonucleotide can comprise a partial duplex or be single stranded. In some cases, the oligonucleotide comprises a partial duplex adapter, wherein the partial duplex comprises a long strand and a short strand. In some cases, the oligonucleotide comprising a partial duplex adapter has overhangs of about, more than, less than, or at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides. The overhang can be a 3' overhang. In some cases, the overhang is a 3' overhang, wherein the overhang comprises at least 6, 7, 8, or 9 nucleotides. In some cases, a 3' overhang of the oligonucleotide hybridizes to sequence present at the 3' end of a polynucleotide comprising a blocked 3' end as generated by the methods described herein. In some cases, the oligonucleotide comprises duplexed sequence. In some cases, the oligonucleotide comprises about, more than, less than, or at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100, 200, or more of base paired or duplexed sequence. In some cases, a partial duplex present in an oligonucleotide comprising the partial duplex and a 3' overhang serves to prevent hybridization of the oligonucleotide to an internal sequence present in a polynucleotide comprising a 3' end block as generated by the methods provided herein. The duplex portion of a oligonucleotide comprising a partial duplex and a 3' overhang as described herein can permit preferential hybridization of the 3' overhang of the oligonucleotide to a 3' end of a polynucleotide comprising a block at the 3'end rather than hybridization to internal sequences present in the polynucleotide comprising a block at the 3'end. The preferential hybridization can be due to steric hindrance and stacking effects caused by the duplex portion of the oligonucleotide. In some cases, the oligonucleotide is single stranded. In some cases, a single-stranded adapter comprises about, more than, less than, or at least 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100, or 200 nucleotides in length. In some cases, the oligonucleotide is a single stranded tailed primer comprising a 3' portion

IDT Exhibit 1020.3626

that is hybridizable to a sequence at the 3' end of a polynucleotide comprising a blocked 3' end as generated by the methods provided herein, and a 5' portion that is non-hybridizable. The non-hybridizable portion can further comprise an identifier sequence (e.g., barcode, TruSeq sequence, etc.). In some cases, the single-stranded oligonucleotide forms a stem-loop or hairpin structure comprising a 3' overhang, wherein the 3' overhang hybridizes to sequence present at the 3' end of a polynucleotide comprising a blocked 3' end as generated by the methods described herein. In some cases, the stem of the hairpin is about, less than about, or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, or more nucleotides in length. In some cases, the loop sequence of a hairpin is about, less than about, or more than about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, or more nucleotides in length. In some cases, the oligonucleotide comprising a stem loop structure has a 3' overhang of about, more than, less than, or at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides. In some cases, the oligonucleotide comprises one or more barcodes. In some cases, one or more barcodes are in a stem and/or a loop of the oligonucleotide. An oligonucleotide comprising a stem loop can further comprise a restriction endonuclease site within the loop. An oligonucleotide comprising a stem loop can further comprise a restriction endonuclease site within the stem. The oligonucleotide comprising a 3' overhang directed against sequence present at the 3' end of a polynucleotide comprising a block at the 3' end can further comprise a block at any and/or all other ends except the 3' end of the 3' overhang. The oligonucleotide can further comprise known or universal sequence (e.g., sequence A) and, thus, allow generation and/or use of sequence specific primers for the universal or known sequence. Some examples of adapters or primers for this step are shown in **FIG. 2**. The two strands forming the dsDNA portion can be two oligonucleotides which can further be connected by a loop. The loop, or linker, can comprise an oligonucleotide, a non-nucleotide linker, or combination thereof. It can also comprise nucleotide analogs. In some cases, an oligonucleotide comprises a partial duplex comprising a first end comprising a blunt end and a second end comprising a 3' overhang, wherein the partial duplex is formed between a long strand and a short strand, wherein the long strand comprises a known or universal sequence (e.g. sequence A) that forms a duplex with the short strand and a 3' overhang. The short strand can have a block at the 3' and/or 5' end. The long strand can have a block at the 5' end. The 3' or 5' blocks can comprise any block or blocking group provided herein. The 3' overhang can comprise sequence complementary to sequence present at the 3' blocked end of a polynucleotide comprising a non-canonical nucleotide as generated by the methods provided herein. The single stranded 3' overhang can comprise a random sequence. In some cases, a pool or plurality of oligonucleotides comprising 3' overhangs comprising random sequence are annealed to a 3' end of a plurality of polynucleotides comprising a blocked 3' end as generated by any of the methods provided herein. In some cases, the random sequence of each of the pool or plurality of oligonucleotides comprises a different random sequence. In some cases, the random sequence of each of the pool or plurality of oligonucleotides comprises a same random sequence. In some cases, the pool or plurality of oligonucleotides comprises a same universal or known sequence (e.g., sequence A). In some cases, the pool or plurality of

IDT Exhibit 1020.3627

oligonucleotides comprises a different universal or known sequence. In some cases, a single strand 3' overhang of an oligonucleotide (e.g., first adapter) hybridizes to the 3'-ends of substantially all the polynucleotides comprising a 3' blocked end as generated by the methods provide herein. In some cases, a pool or plurality of single strand 3' overhangs provided by a pool or plurality of oligonucleotides (e.g., first adapters), wherein each oligonucleotide (e.g., first adapter) of the pool or plurality of oligonucleotides (e.g., first adapters) comprises a 3' overhang comprising a different random sequence, hybridize to the 3'-ends of substantially all the polynucleotides comprising a 3' blocked end as generated by any of the methods provide herein. A single strand 3' overhang of an oligonucleotide (e.g., first adapter) can hybridize to more than, less than, at least, at most, or about 1%, 2%, 3%,4%,5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5 % or 100% of the polynucleotides comprising a 3' blocked end as generated by the methods provide herein. In some cases, the single strand 3' overhang hybridizes to the 3'-ends of between 1-10%, 10-20%, 20-30%, 30-40%, 40-50%, 50-60%, 60-70%, 70-80%, 80-90%, 90-95%, 95-99% or 90-100% of the polynucleotides comprising a 3' blocked end as generated by the methods provide herein. In some cases, the single strand 3' overhang hybridizes to the 3'-ends of about 1 to about 10%, about 10 to about 20%, about 20 to about 30%, about 30 to about 40%, about 40 to about 50%, about 50 to about 60%, about 60 to about 70%, about 70 to about 80%, about 80 to about 90%, or about 90 to about 100% of the polynucleotides comprising a 3' blocked end as generated by the methods provide herein. A pool or plurality of single strand 3' overhangs provided by a pool or plurality of oligonucleotides (e.g., first adapters), wherein each oligonucleotide (e.g., first adapter) of the pool or plurality of oligonucleotides (e.g., first adapters) comprises a 3' overhang comprising a different random sequence, can hybridize to more than, less than, at least, at most, or about 1%, 2%, 3%,4%,5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5 % or 100% of the polynucleotides comprising a 3' blocked end as generated by the methods provide herein. In some cases, the pool or plurality of single strand 3' overhangs provided by a pool or plurality of oligonucleotides (e.g., first adapters), wherein each oligonucleotide (e.g., first adapter) of the pool or plurality of oligonucleotides (e.g., first adapters) comprises a 3' overhang comprising a different random sequence, hybridizes to the 3'-ends of between 1-10%, 10-20%, 20-30%, 30-40%, 40-50%, 50-60%, 60-70%, 70-80%, 80-90%, 90-95%, 95-99% or 90-100% of the polynucleotides

IDT Exhibit 1020.3628

comprising a 3' blocked end as generated by the methods provide herein. In some cases, the pool or plurality of single strand 3' overhangs provided by a pool or plurality of oligonucleotides (e.g., first adapters), wherein each oligonucleotide (e.g., first adapter) of the pool or plurality of oligonucleotides (e.g., first adapters) comprises a 3' overhang comprising a different random sequence, hybridizes to the 3'-ends of about 1 to about 10%, about 10 to about 20%, about 20 to about 30%, about 30 to about 40%, about 40 to about 50%, about 50 to about 60%, about 60 to about 70%, about 70 to about 80%, about 80 to about 90%, or about 90 to about 100% of the polynucleotides comprising a 3' blocked end as generated by the methods provide herein. In some cases, the oligonucleotide comprises one or more barcodes. In some cases, the one or more barcodes are in a stem and/or a loop. In some cases the barcodes comprise a random sequence that is useful for uniquely marking an individual polynucleotide generated by the methods described herein to which the barcode is appended. In some cases, the barcodes are appended at random and are unique for the fragment to which it was appended. These barcodes can be combined with barcodes that are specific for a sample of a template nucleic acid.

[0094] In some cases, the method can further comprise performing an extension reaction. The extension reaction can be performed using any number of methods known in the art including, but not limited to, the use of a DNA dependent DNA polymerase with strand displacement activity and all four dNTPs (i.e. dATP, dTTP, dCTP, and dGTP), wherein the dNTPs are unmodified. In some cases, the extension reaction is performed with a DNA polymerase and unmodified dNTPs (i.e. dATP, dTTP, dCTP, and dGTP). In some cases, the extension reaction extends the 3' overhang annealed to the complementary sequence found at the 3' blocked end of the polynucleotide comprising a blocked 3' end, thereby generating a double stranded polynucleotide comprising non complementary ends, wherein the polynucleotide comprising the 3' block serves as the template polynucleotide. The double stranded polynucleotide comprising non-complementary ends can comprise a known or universal sequence (e.g., sequence A) from the oligonucleotide at one end and a sequence complementary to the 5' end of the polynucleotide comprising a blocked 3' end that served as template for the extension reaction at the opposite end of the polynucleotide. The double stranded polynucleotide generated by the extension reaction can comprise a first strand comprising a fragment of the template polynucleotide, and a second strand comprising sequence complementary to the fragment of the template polynucleotide and the known or universal sequence (e.g., sequence A), wherein the known sequence is present at the 5' end of the second strand, and wherein the 3' end of the first strand comprises a gap in the phosphodiester backbone between the sequence complementary to the known or universal sequence (e.g., sequence A), and the 3' block from the template polynucleotide. The known or universal sequence (e.g., sequence A) can serve to mark the strand comprising the known or universal sequence (e.g., sequence A). In cases where the non-canonical nucleotide is incorporated during first strand cDNA synthesis, generation of the marked strand by the methods provided herein produces a marked strand representing the sequence of the template nucleic acid. In cases where the non-canonical nucleotide is incorporated during second strand cDNA

IDT Exhibit 1020.3629

synthesis, generation of the marked strand by the methods provided herein produces a marked strand representing the sequence complementary to the template nucleic acid.

[0095] In some cases, a double stranded polynucleotide comprising non-complementary ends wherein one end comprises a known or universal sequence (e.g., sequence A) at one end is end repaired following an extension reaction. End repair can include the generation of blunt ends, non-blunt ends (i.e. sticky or cohesive ends), or single base overhangs such as the addition of a single dA nucleotide to the 3′-end of the double-stranded nucleic acid product by a polymerase lacking 3′-exonuclease activity. In some cases, end repair is performed on the double stranded polynucleotide comprising known or universal sequence (e.g., sequence A) at one end to produce a blunt end on the end opposite the one end comprising the known sequence, wherein one end comprises a known or universal sequence (e.g., sequence A) and an opposite end comprises a blunt end with a 3' OH. End repair can be performed using any number of enzymes and/or methods known in the art. An overhang can comprise about, more than, less than, or at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides.

[0096] The method can further comprise appending an adapter to the double-stranded polynucleotide comprising sequence A at one end and a 3' OH at the opposite end. In some cases, the adapter annealed to polynucleotide comprising a 3' block as generated by the methods provided herein is a first adapter, while the adapter appended to an opposite end of the double-stranded polynucleotide comprising first adapter sequence at one end is a second adapter. Ligation can be blunt end ligation or sticky or cohesive end ligation. Appending the second adapter can be through ligation. The ligation can be performed with any of the enzymes known in the art for performing ligation (e.g., T4 DNA ligase). The second adapter can be any type of adapter known in the art including, but not limited to, a conventional duplex or double stranded adapter. The adapter can comprise DNA, RNA, or a combination thereof. The second adapter can be about, less than about, or more than about 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100, or 200 nucleotides in length. The second adapter can be a duplex adapter, partial duplex adapter, or single stranded adapter. In some cases, the second adapter is a duplex adapter. In some cases, the duplex adapter can be about, less than about, or more than about 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100, or 200 nucleotides in length. In some cases, the second adapter is a partial duplex adapter, wherein the adapter comprises a long strand and a short strand. In some cases, the second adapter comprising a partial duplex adapter has overhangs of about, more than, less than, or at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides. In some cases, the overhang is a 5' overhang. In some cases, the overhang is a 3' overhang. In some cases, the partial duplex of the second adapter comprises about, more than, less than, or at least 5, 6, 7, 8, 9, 10, 12, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100, 200, or more of base paired or duplexed sequence. In some cases, the adapter comprises a single stranded adapter. In some cases, a single-stranded adapter comprises about, more than, less than, or at least 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100, or 200 nucleotides in length. In some cases, the single-stranded adapter forms a stem-loop or hairpin structure. In some cases, the stem of the hairpin adapter is

IDT Exhibit 1020.3630

about, less than about, or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, or more nucleotides in length. In some cases, the loop sequence of a hairpin adapter is about, less than about, or more than about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, or more nucleotides in length. The second adapter can further comprise known or universal sequence (e.g., sequence B) and, thus, allow generation and/or use of sequence specific primers for the universal or known sequence. A second adapter comprising a stem loop can further comprise a restriction endonuclease site within the loop. A second adapter comprising a stem loop can further comprise a restriction endonuclease site within the stem. In the methods provided herein, a known or universal sequence of a second adapter as provided herein can be the same or different from a known or universal sequence of a first adapter as provided herein. In some cases, a first adapter comprises sequence A and a second adapter comprises sequence B, wherein sequence B is different or non-complementary to sequence A. In some cases, a second adapter comprises one or more barcodes. In some cases, one or more barcodes are in a stem and/or a loop.

[0097] In some cases, appending of the second adapter to the double-stranded polynucleotide comprising known or universal sequence (e.g., sequence A) at one end and a 3' OH at the opposite end is by blunt end ligation. In some cases, appending of the second adapter is by cohesive or sticky end ligation, wherein an overhang in the second adapter hybridizes to an overhang in the double stranded polynucleotide comprising complementary sequence to the overhang. In some cases, the second adapter comprises a ligation strand or first strand capable of ligation to a 5'end of the double-stranded polynucleotide comprising known or universal sequence (e.g., sequence A) at one end and a 3' OH at the opposite end and a non-ligation strand or second strand incapable of ligation to either end of the double-stranded polynucleotide comprising known or universal sequence (e.g., sequence A) at one end and a 3' OH at the opposite end. In some cases, the second adapter comprises a ligation strand or first strand capable of ligation to a 3'end of the double-stranded polynucleotide comprising known or universal sequence (e.g., sequence A) at one end and a 3' OH at the opposite end and a non-ligation strand or second strand incapable of ligation to either end of the double-stranded polynucleotide comprising known or universal sequence (e.g., sequence A) at one end and a 3' OH at the opposite end. In some cases, the second adapter is a partial duplex adapter, wherein the adapter comprises a long strand and a short strand, and wherein the long strand is the ligation strand or first strand, while the short strand is the non-ligation strand or second strand. The short strand can have a block at the 3' and/or 5' end. The long strand can have a block at the 3' or 5' end. The 3' or 5' blocks can comprise any block or blocking group provided herein. In some cases, the partial duplex has strands of unequal length. In some cases, the partial duplex comprises an overhang at one end of the adapter and a blunt end at another end of the adapter. The overhang can be at the 3' end or the 5' end. In some cases, the partial duplex comprises an overhang at each end of the adapter. The overhang can be of equal length or unequal length. In some cases, the 5' end of the ligation strand does not comprise a 5' phosphate group. In some cases, the 5' end of the ligation strand does comprise a 5' phosphate, wherein the 3' end of the polynucleotide lacks a free 3'

IDT Exhibit 1020.3631

hydroxyl. In some cases, the second adapter comprises a long strand comprising a 3' overhang and a known sequence (e.g., sequence B) that forms a partial duplex with a short strand, wherein the short strand comprises a block at a 3' end, and wherein the long strand is ligated to the 3' OH at the opposite end of the double-stranded polynucleotide comprising known or universal sequence (e.g., sequence A) at one end and a 3' OH at the opposite end, thereby generating a double stranded polynucleotide comprising known or universal sequence at both ends. Further to these cases, the double stranded polynucleotide comprising known or universal sequence at both ends comprises one strand comprising known or universal sequence derived from the oligonucleotide annealed to the polynucleotide comprising a blocked 3' end and extended as described herein at the 5' end and the known or universal sequence derived from ligation of the second adapter. In some cases, the one strand comprises sequence A at a 5' end and sequence B at a 3' end. In some cases, the second adapter comprises a long strand comprising a 5' overhang and a known sequence (e.g., sequence B) that forms a partial duplex with a short strand, wherein the short strand comprises a block at a 5' end, and wherein the long strand is ligated to the 5' phosphate at the opposite end of the double-stranded polynucleotide comprising known or universal sequence (e.g., sequence A) at one end and a 3' OH at the opposite end, thereby generating a double stranded polynucleotide comprising known or universal sequence at both ends. Further to these cases, the ligating of the second adapter to the double-stranded polynucleotide comprising known or universal sequence (e.g., sequence A) at one end and a 3' OH at the opposite end generates a double stranded polynucleotide comprising known or universal sequence (e.g., sequence A) derived from the oligonucleotide annealed to the polynucleotide comprising a blocked 3' end and extended as described herein at one end and the known or universal sequence (e.g., sequence B) derived from the second adapter at an opposite end, wherein the known or universal sequence (e.g., sequence A) derived from the oligonucleotide annealed to the polynucleotide comprising a blocked 3' end and extended as described herein is at a 5' end on one end and the known or universal sequence (e.g., sequence B) derived from the second adapter is at a 5' end on the opposite end. In some cases, the one strand comprises sequence A at a 5' end of one strand and sequence B at a 5' end on another strand, wherein the 3' end of the strand comprising sequence A is extended using the sequence B as a template, thereby generating one or more double stranded polynucleotides comprising the sequence A at a 5' end on one end and a sequence complementary to sequence B, B', at a 3' end on the opposite end.

[0098] In some cases, the method further comprises a denaturing step, a double stranded polynucleotide comprising non complementary known or universal sequences on opposite ends generated by the methods provided herein are denatured. Denaturation can be achieved using any of the methods known in the art which can include, but are not limited to, heat denaturation, and/or chemical denaturation. Heat dentauration can be performed by raising the temperature of the reaction mixture to be above the melting temperature of the polynucleotide comprising non complementary known or universal sequences on opposite ends generated by the methods provided herein. The melting temperature can be about, more than, less than, or at least 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47,

IDT Exhibit 1020.3632

48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, or 95 degrees C. The temperature can be raised above the melting temperature by about, more than, less than, or at least 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 degrees C. Chemical denaturation can be performed using bases (i.e. NaOH), and/or competitive denaturants (i.e. urea, or formaldehyde). In some cases, denaturation generates single stranded polynucleotides comprising non-complementary known or universal sequences on opposite ends generated by the methods provided herein.

[0099] Following denaturation, a single stranded polynucleotide comprising non complementary known or universal sequences on opposite ends generated by the methods provided herein are amplified, thereby generating directional polynucleotide libraries. The known or universal sequence on one or a first end can be derived from the first adapter, while the known or universal sequence on the other or a second end can be derived from the second adapter as described herein. The amplification can be performed using primer pairs directed against the non-complementary known or universal sequences present on the opposite ends. The amplification can be performed using amplification method known in the art, which can include, but is not limited to, PCR or single primer isothermal amplification (SPIA). In some cases, a single-stranded polynucleotide comprising sequence A at a 5' end and sequence B at a 3' end is amplified using a primer pair, wherein a first primer of the primer pair comprises sequence complementary to a portion of sequence B and a second primer of the primer pair comprising sequence complementary to a portion of the complement of sequence A, sequence A'. In some cases, single stranded polynucleotide comprising sequence A at a 5' end of a one strand and sequence B' at a 3' end is amplified using a primer pair, wherein a first primer of the primer pair comprises sequence complementary to a portion of sequence B' and a second primer of the primer pair comprising sequence complementary to a portion of the complement of sequence A, sequence A'. In some cases, the first and/or second primer further comprises one or more identifier sequences. In some cases, the identifier sequences comprise a non-hybridizable tail on the first and/or second primer. The identifier sequence can be a barcode sequence, a flow cell sequence, an index sequence, or a combination thereof. In some cases, the index sequence is a Truseq primer sequence compatible with the next generation sequencing platform produced by Illumina. In some cases, the first and/or second primer can bind to a solid surface. The solid surface can be a planar surface or a bead. The planar surface can be the surface of a chip, microarray, well, or flow cell. In some cases, the first and/or second primer comprises one or more sequence elements products of the amplification reaction (i.e. amplification products) to a solid surface, wherein the one or more sequences are complementary to one or more capture probes attached to a solid surface. Other sequence elements known in the art that can be compatible with other massively parallel next generation sequencing platforms can be incorporated in the tail sequences.

[00100] Sequencing can be any method of sequencing, including any of the next generation sequencing (NGS) methods described herein. In some cases, the NGS method comprises sequencing by synthesis. In some embodiments, sequencing is performed with primers directed against known or universal sequence

IDT Exhibit 1020.3633

introduced into the polynucleotides generated by the methods provided herein by the adapters appended to the polynucleotides. In some cases, sequencing is performed with primers directed against identifier sequence introduced into the polynucleotides by the first and/or second primer used to amplify the single-stranded polynucleotide comprising non-complementary known or universal sequence at opposite ends. The identifier sequence can be a barcode sequence, a flow cell sequence, and/or index sequence. In some cases, the index sequence is a Truseq primer sequence compatible with the next generation sequencing platform produced by Illumina.

[00101] A schematic depicting an exemplary workflow using the methods described herein for generating a directional polynucleotde library from an RNA sample is shown in **FIG. 3**. Step **I** starts with isolating total RNA from a sample and annealing first strand primers to the total RNA. The first strand primers can comprise random sequence or sequence specific to a specific transcript or group of transcripts. The first strand primers can be designed to prime all transcripts except certain transcripts (e.g., rRNA and/or mitochondrial RNA). In step **II**, first strand cDNA synthesis is performed on the total RNA isolated in step **I** using the first strand primers from step **I**. The first strand cDNA synthesis reaction is performed in the presence of a reaction mixture comprising all four dNTPs and the non-canonical dNTP, dUTP. Step **III** entails cleaving the first strand cDNA comprising dU using UDG to generate abasic sites, and a cleavage agent capable of cleaving the phsophodiester backbone at the abasic site generated by UDG. The cleavage agent can be DMED or heat. Step **III** generates polynucleotides comprising a block at the 3' end, and, optionally, a 5' phosphate. The incorporation of dUTP during step **II** can be controlled by controlling the amount or a ratio of dUTP to the other dNTPs within the reaction mixture such that step **II** produces first strand cDNA comprising uracil bases at a desired density, whereby step **III** generates polynucleotides comprising a block at the 3; end of a desired size. The desired size can be determined by a downstream application, like, for example, a specific next generation sequencing platform. The template total RNA from step **I** is degraded in step **IV** and the polynuleotides generated in step **III** are purified in step **V**. Degradation of the template RNA can be performed using an RNase (e.g., RNaseH or RNase I) or by heat treatment. Following purification, a first adapter comprising a 3' overhang comprising random sequence is annealed to sequence present at the 3' end of the polynucleotides generated in step **III**. The first adapter can be single stranded and comprise a hairpin structure in addition to the 3' overhang. The first adapter can be a plurality of first adapters, wherein each of the plurality of first adapters comprises a different random sequence and each of the plurality comprises a same universal sequence. The first adapter can comprise two oligonucleotides that form a partial duplex wherein one strand is longer than the other strand at the 3' end and thereby comprises a 3' overhang. The first adapter can further comprise a first universal sequence. Once annealed, the 3' end of the overhang annealed to the 3' end of the polynucleotides generated in step **III** is extended with a DNA polymerase to produce a second strand cDNA. The end of the newly generated second strand can be polished using T4 polymerase in step **VIII**, and then purified in step **IX**. Ultimately, a second adapter is ligated to the double stranded polynucleotide product of step **VII**. The second adapter can comprise a

IDT Exhibit 1020.3634

second universal sequence. The product of step **X** can comprise a double stranded polynucleotide comprising one strand with a first universal sequence on one end and a second universal sequence on a second, opposite end with an insert comprising sequence representing a portion of the original RNA template between the first and second ends. The product of step **X** is then purified in step **XI** and subjected to PCR with primers directed against the first and second universal sequences appended to the product of step **X** in step **XII**. The primers can be suitable for any of the next generation sequencing platforms known in the art and can further comprise barcodes and/or any other identifier sequence known in the art.

[00102] A schematic exemplary of an embodiment of the methods described herein for generating a directional polynucleotide library from an RNA template is shown in **FIG. 1A**. As illustrated in step **I** of **FIG. 1A**, a primer is hybridized to a template RNA. As provided herein, the primer can comprise random sequence, transcript specific sequence, and/or an oligo dT. In step **II**, the primer is extended in the presence of dUTP to produce a first strand cDNA or polynucleotide extension product. The extension can be performed using an RNA dependent DNA polymerase as provided herein. In step **III**, following degradation of the template RNA, the polynucleotide comprising uracil bases is degraded using UNG and heat or a polyamine (DMED), thereby producing multiple fragments comprising a 3' blocked end. The degradation of the template RNA can be performed using an RNase (e.g. RNase H or RNase I). Alternatively, the RNA template polynucleotide can be degraded by other methods that include, but are not limited to, heat or alkaline pH treatment, or combination of various methods. Heat treatment for the degradation of the RNA template can also be used for the cleavage of the backbone of the complementary DNA comprising the abasic sites, thus achieving fragmentation of the complementary DNA and the RNA template in a single step. In step **IV**, a first adapter is annealed to sequence present at the 3' blocked end of the polynucleotides generated in step **III**. The first adapter comprises a 3' overhang comprising random sequence at the 3' end, whereby the 3' overhang binds a complementary sequence at the 3' blocked end of the polynucleotides generated in step **III**. The first adapter can be a plurality of first adapters, wherein each of the plurality of first adapters comprises a different random sequence, wherein the random sequence on one of the plurality of first adapters can anneal to complementary sequence present at the 3' end on one or more of the polynucleotides generated in step **III**. Each of the plurality can comprise sequence A.The 3' end of the annealed 3' overhang of the first adapter is extended along the polynucleotide comprising the blocked 3' end in step **V**, thereby generating double stranded polynucleotides with sequence A appended to the 5' end of one strand of the double stranded polynucleotide. The sequence complementary to sequence A, A', is not appended to the other strand of the double stranded polynucleotide generated in step **V** due to the 3' block generated in step **III**. In step **VI**, a second adapter is ligated to the end of the double stranded polynucleotide generated in step **V,** opposite the end comprising sequence A. The second adapter comprises a partial duplex, formed between a long strand comprising a sequence B and a short strand comprising a portion of the complement of sequence B, B'. The long strand further comprises a 3' overhang, while the short strand further

IDT Exhibit 1020.3635

comprises a block at the 3' end. The block can be any block or blocking group as provided herein. In step **VI**, the long strand serves as a ligation strand, while the short strand serves as a non-ligation strand, whereby the 5' end of the long strand is ligated to the 3' end of the strand of the double stranded polynucleotide produced in step **V** comprising sequence A at its 5' end, thereby generating a double stranded polynucleotide comprising non-complementary ends. The ligation can be performed using any of the methods provided herein including, but not limited to, generating a blunt end at the end of the double stranded polynucleotide generated in step **V** and performing blunt end ligation. One strand of the double stranded polynucleotide generated in step **VI** comprises a strand specific polynucleotide comprising sequence A at a 5; end and sequence B at a 3' end. The strand specific polynucleotide can be amplified using any of the amplification methods provided herein. In some cases, the amplification comprises performed an amplification reaction using a first primer directed against sequence B, and a second primer directed against the complement of sequence A, A'. Either or both of the first or second primer can further comprise a non-hybridizable tail, wherein the tail comprises a reverse flow cell sequence, a TruSeq primer sequence,a barcode sequence and/or any other desired sequence useful for downstream applications as described herein. Following amplification with the first and second primers, an amplification product comprising double stranded polynucleotide sequence appended with non-complementary adapter sequence at each end derived from the ligated adapter and flow cell sequences are generated. The amplification products can be compatible with any of the next generation sequencing platform as provided herein.

[00103] **FIG. 1B** shows a schematic exemplary of an embodiment of the methods described herein for generating a directional polynucleotide library from an RNA template. Steps **I** through **V** of FIG. **1B** are identical to steps **I** through **V** of FIG. **1A**. Similar to FIG. 1A, the second adapter of step **VI** of FIG. **1B** comprises a partial duplex, formed between a long strand comprising a sequence B and a short strand comprising a portion of the complement of sequence B, B'. In contrast to **FIG. 1A**, the long strand of second adapter of step **VI** of **FIG. 1B** comprises a 5' overhang, while the short strand further comprises a block at the 5' end. The block can be any block or blocking group as provided herein. In step **VI**, the long strand serves as a ligation strand, while the short strand serves as a non-ligation strand, whereby the 5' end of the long strand is ligated to the 5' end of the oppoiste strand of the double stranded polynucleotide produced in step **V** comprising sequence A at its 5' end, thereby generating a double stranded polynucleotide comprising non-complementary ends. The ligation can be performed using any of the methods provided herein including, but not limited to, generating a blunt end at the end of the double stranded polynucleotide generated in step **V** and performing blunt end ligation. Due to the block at the 5' end, the short strand is not ligated to the strand of the double stranded polynucleotide generated in step **V** comprising sequence A at a 5' end, whereby a gap exists. In step **VII**, the double stranded polynucleotide generated in step **VI** is subjected to a fill in reaction, whereby the 3' end of the strand comprising sequence A at its 5' end is extended using a DNA polymerase comprising strand displacement activity as provided herein using sequence B as a template. Alternatively, the non ligated strand may be

IDT Exhibit 1020.3636

removed by an exonuclease activity of the polymerase. Step **VII** generates a double stranded polynucleotide comprising one strand of the double stranded polynucleotide comprising a strand specific polynucleotide comprising sequence A at a 5; end and sequence B' at a 3' end. In some cases, the second adapter of step IV comprises a double stranded adapter, wherein a first strand comprise sequence B and a second strand comprising sequence B', wherein the first strand comprises a block at both ends, while the second strand comprises a blocking group at the 3' end. In these cases, ligation of the second adapter generates a double stranded polynucleotide comprising one strand of the double stranded polynucleotide comprising a strand specific polynucleotide comprising sequence A at a 5; end and sequence B' at a 3' end without requiring step VII. The strand specific polynucleotide can be amplified using any of the amplification methods provided herein. In some cases, the amplification comprises an amplification reaction using a first primer directed against sequence B', and a second primer directed against the complement of sequence A, A'. Either or both of the first or second primer can further comprise a non-hybridizable tail, wherein the tail comprises a reverse flow cell sequence, a TruSeq primer sequence and/or a barcode sequence. Following amplification with the first and second primers, an amplification product comprising double stranded polynucleotide sequence appended with non-complementary adapter sequence at each end derived from the ligated adapter and flow cell sequences are generated. The amplification products can be compatible with the next generation sequencing platform as provided herein.

[00104] A schematic exemplary of an embodiment of the methods described herein for amplifying a polynucleotide generated by the methods provided herein using SPIA is shown in **FIG. 5**. In step **I**, a chimeric amplification primer is hybridized to a polynucleotide comprising sequence A at the 5' end and sequence B at the 3' end generated by the methods provided herein. The chimeric amplification primer can comprise a 3' DNA portion comprising sequence C and a 5' RNA portion comprising sequence D, wherein sequence C comprises sequence complementary to a portion of sequence B, and wherein sequence D comprises sequence non-hybridizable to the polynucleotide. In step **II**, an extension reaction is performed using a DNA polymerase comprising RNA dependent DNA polymerase activity, wherein the 3' end of sequence C is extended using the polynucleotide as template, and wherein the 3' end of sequence B of the polynucleotide is extended using sequence D as the template, thereby generating a double stranded polynucleotide comprising sequence A and its complement A' at one end and a heteroduplex comprising RNA sequence D and its DNA complement D' at the other end. In step **III**, sequence D is cleaved using RNaseH, wherein a double stranded polynucleotide comprising sequence A and its complement A' at one end and a 3' single stranded DNA overhang comprising sequence C on the other end is generated. In step **IV**, an amplification chimeric primer comprising a 5' RNA portion complementary to sequence D' is annealed to sequence D' and extended using a strand displacement DNA polymerase, wherein the DNA polymerase displaces a single stranded amplification product comprising sequence A' at the 3' end and sequence C at the 5' end, wherein a double stranded polynucleotide comprising sequence A and its complement A' at one end and a heteroduplex comprising

IDT Exhibit 1020.3637

RNA sequence D and its DNA complement D' at the other end is newly generated. Steps **III** and **IV** is then repeated to generate a pool of amplification products.

VI. Oligonucleotides

[00105] The term "oligonucleotide" can refer to a polynucleotide chain, typically less than 200 residues long, e.g., between 15 and 100 nucleotides long, but also intended to encompass longer polynucleotide chains. Oligonucleotides can be single-or double-stranded. The terms "primer" and "oligonucleotide primer" can refer to an oligonucleotide capable of hybridizing to a complementary nucleotide sequence. The term "oligonucleotide" can be used interchangeably with the terms "primer," "adapter," and "probe."

[00106] The term "hybridization"/ "hybridizing" and "annealing" can be used interchangeably and can refer to the pairing of complementary nucleic acids.

[00107] The term "primer" can refer to an oligonucleotide, generally with a free 3' hydroxyl group, that is capable of hybridizing with a template (such as a target polynucleotide, target DNA, target RNA or a primer extension product) and is also capable of promoting polymerization of a polynucleotide complementary to the template. A primer can contain a non-hybridizing sequence that constitutes a tail of the primer. A primer can still be hybridizing to a target even though its sequences may not fully complementary to the target.

[00108] Primers can be oligonucleotides that can be employed in an extension reaction by a polymerase along a polynucleotide template, such as in PCR or cDNA synthesis, for example. The oligonucleotide primer can be a synthetic polynucleotide that is single stranded, containing a sequence at its 3′-end that is capable of hybridizing with a sequence of the target polynucleotide. Normally, the 3′ region of the primer that hybridizes with the target nucleic acid has at least 80%, 90%, 95%, or 100%, complementarity to a sequence or primer binding site.

[00109] Primers can be designed according to known parameters for avoiding secondary structures and self-hybridization. Different primer pairs can anneal and melt at about the same temperatures, for example, within about 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 °C. of another primer pair. In some cases, greater than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200, 500, 1000, 5000, 10,000 or more primers are initially used. Such primers may be able to hybridize to the genetic targets described herein. In some cases, about 2 to about 10,000, about 2 to about 5,000, about 2 to about 2,500, about 2 to about 1,000, about 2 to about 500, about 2 to about 100, about 2 to about 50, about 2 to about 20, about 2 to about 10, or about 2 to about 6 primers are used.

[00110] Primers can be prepared by a variety of methods including but not limited to cloning of appropriate sequences and direct chemical synthesis using methods well known in the art (Narang et al., Methods Enzymol. 68:90 (1979); Brown et al., Methods Enzymol. 68:109 (1979)). Primers can also be obtained from commercial sources such as Integrated DNA Technologies, Operon Technologies, Amersham Pharmacia Biotech, Sigma, and Life Technologies. The primers can have an identical melting temperature. The melting temperature of a primer can be about, more than, less than, or at least 30, 31,

IDT Exhibit 1020.3638

32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 81, 82, 83, 84, or 85° C. In some cases, the melting temperature of the primer is about 30 to about 85°C., about 30 to about 80°C., about 30 to about 75°C., about 30 to about 70°C., about 30 to about 65°C., about 30 to about 60°C., about 30 to about 55°C., about 30 to about 50°C., about 40 to about 85°C., about 40 to about 80°C., about 40 to about 75°C., about 40 to about 70°C., about 40 to about 65°C., about 40 to about 60°C., about 40 to about 55°C., about 40 to about 50°C., about 50 to about 85°C., about 50 to about 80°C., about 50 to about 75°C., about 50 to about 70°C., about 50 to about 65°C., about 50 to about 60°C., about 50 to about 55°C., about 52 to about 60°C., about 52 to about 58°C., about 52 to about 56°C., or about 52 to about 54°C.

[00111] The lengths of the primers can be extended or shortened at the 5' end or the 3' end to produce primers with desired melting temperatures. One of the primers of a primer pair can be longer than the other primer. The 3' annealing lengths of the primers, within a primer pair, can differ. Also, the annealing position of each primer pair can be designed such that the sequence and length of the primer pairs yield the desired melting temperature. An equation for determining the melting temperature of primers smaller than 25 base pairs is the Wallace Rule (Td=2(A+T)+4(G+C)). Computer programs can also be used to design primers, including but not limited to Array Designer Software (Arrayit Inc.), Oligonucleotide Probe Sequence Design Software for Genetic Analysis (Olympus Optical Co.), NetPrimer, and DNAsis from Hitachi Software Engineering. The T_M (melting or annealing temperature) of each primer can be calculated using software programs such as Net Primer (free web based program at http://www.premierbiosoft.com/netprimer/index.html). The annealing temperature of the primers can be recalculated and increased after any cycle of amplification, including but not limited to about cycle 1, 2, 3, 4, 5, about cycle 6 to about cycle 10, about cycle 10 to about cycle 15, about cycle 15 to about cycle 20, about cycle 20 to about cycle 25, about cycle 25 to about cycle 30, about cycle 30 to about cycle 35, or about cycle 35 to about cycle 40. After the initial cycles of amplification, the 5' half of the primers can be incorporated into the products from each loci of interest; thus the T_M can be recalculated based on both the sequences of the 5' half and the 3' half of each primer.

[00112] The annealing temperature of the primers can be recalculated and increased after any cycle of amplification, including but not limited to about cycle 1, 2, 3, 4, 5, about cycle 6 to about cycle 10, about cycle 10 to about cycle 15, about cycle 15 to about cycle 20, about cycle 20 to about cycle 25, about cycle 25 to about cycle 30, about cycle 30 to about 35, or about cycle 35 to about cycle 40. After the initial cycles of amplification, the 5' half of the primers can be incorporated into the products from each loci of interest, thus the TM can be recalculated based on both the sequences of the 5' half and the 3' half of each primer.

[00113] "Complementary" can refer to complementarity to all or only to a portion of a sequence. The number of nucleotides in the hybridizable sequence of a specific oligonucleotide primer should be such that stringency conditions used to hybridize the oligonucleotide primer will prevent excessive random

IDT Exhibit 1020.3639

non-specific hybridization. Usually, the number of nucleotides in the hybridizing portion of the oligonucleotide primer will be at least as great as the defined sequence on the target polynucleotide that the oligonucleotide primer hybridizes to, namely, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least about 20, and generally from about 6 to about 10 or 6 to about 12 of 12 to about 200 nucleotides, usually about 10 to about 50 nucleotides. A target polynucleotide can be larger than an oligonucleotide primer or primers as described previously.

[00114] In some cases, the identity of the target polynucleotide sequence is known, and hybridizable primers can be synthesized precisely according to the antisense sequence of the aforesaid target polynucleotide sequence. In other cases, when the target polynucleotide sequence is unknown, the hybridizable sequence of an oligonucleotide primer can be a random sequence. Oligonucleotide primers comprising random sequences can be referred to as "random primers", as described below. In yet other cases, an oligonucleotide primer such as a first primer or a second primer comprises a set of primers such as for example a set of first primers or a set of second primers. In some cases, the set of first or second primers can comprise a mixture of primers designed to hybridize to a plurality (e.g. about, more than, less than, or at least 2, 3, 4, 6, 8, 10, 20, 40, 80, 100, 125, 150, 200, 250, 300, 400, 500, 600, 800, 1000, 1500, 2000, 2500, 3000, 4000, 5000, 6000, 7000, 8000, 10,000, 20,000, or 25,000) target sequences. In some cases, the plurality of target sequences can comprise a group of related sequences, random sequences, a whole transcriptome or fraction (e.g. substantial fraction) thereof, or any group of sequences such as mRNA. Primers for use in the methods provided herein can be any of the primers listed in **Tables 1** and **2**, which are directed against the first and second adapter sequences listed in **Tables 3** and **4**, respectively.

[00115] **Table 1**: Primer sequences directed against first adapter listed in **Table 3**.

Primer (5'-3')
AAGCAGAAGACGGCATACGAGATGAGGTGGCTGCTGTCTTTCCCTCGTTTTCTCAAGCGACAC-
AAGCAGAAGACGGCATACGAGATGAGGTGGTGATCGGAGTGCAGAATCGTGGACTTCTAGTCT-
AAGCAGAAGACGGCATACGAGATGAGGTGGCCCAATGCGTTCTATATGCGTCTCAGCTGCGGC-
AAGCAGAAGACGGCATACGAGATGAGGTGGCTTGCGTGCACGAGAAGCATCGCCTCTCGAAGC
AAGCAGAAGACGGCATACGAGATGAGGTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
AAGCAGAAGACGGCATACGAGATGAGGTGGTTAGCACTCGGCCGCAATTCTGAGTAATCTGGC-
AAGCAGAAGACGGCATACGAGATGAGGTGGGGCCTGTCGCGGTCCGAGCGATAAGCACGATCT

AAGCAGAAGACGGCATACGAGATGAGGTGGTGACTGCTCATTGTGCATGTGGAGC GATTACCCAGT
AAGCAGAAGACGGCATACGAGATGAGGTGGGCTTGACTGGAGATGCGTAAAGCTT GACGACGATCT-
AAGCAGAAGACGGCATACGAGATGAGGTGGTGATGATACCCGATTCGCACCTGCG AAACGTGTTCTATG-
AAGCAGAAGACGGCATACGAGATGAGGTGGACTTCATACGCAATTCGAATCTACGC CACGTGTTCTTTGCGA-
AAGCAGAAGACGGCATACGAGATGAGGTGGTACGCAATTCGAATCTACGCCACGT GTTCTTTGCGA-
AAGCAGAAGACGGCATACGAGATGAGGTGGGCTTGACTACTGGAGATGCGTAAAG CTTGACGACGATCT-
AAGCAGAAGACGGCATACGAGATGAGGTGGCTTGCGTGCACGAGATTCAGCATCG CCTCTCGAGGAAGC-
AAGCAGAAGACGGCATACGAGATGAGGTGGCTGCTGTCTTTCCCTCGTTTTCTCAA GTTTGCGCAC-
AAGCAGAAGACGGCATACGAGATGAGGTGGTGATCGTCTTGCAGAATCGTGGACA GCTAGTCTGCT-
AAGCAGAAGACGGCATACGAGATGAGGTGGAGATACCGACGCGATGAAGCACGTT GCACCCTT-
AAGCAGAAGACGGCATACGAGATGAGGTGGTCGGATGAGCGAAGTTGCAATCCCG AACTTTCATGC-
AAGCAGAAGACGGCATACGAGATGAGGTGGAGATCGGAATTCCACACGTCTGAAT AACAGTCA-
AAGCAGAAGACGGCATACGAGATGAGGTGGGCCGCAGCTGAGACGCATATAGAAC GCATTGGGCGA-
AAGCAGAAGACGGCATACGAGATGAGGTGGCTGCTGTCTTTCCCTCGTTTTCTCAA GCGACAC-
AAGCAGAAGACGGCATACGAGATGAGGTGGTGATCGGAGTGCAGAATCGTGGACT TCTAGTCT-
AAGCAGAAGACGGCATACGAGATGAGGTGGCCCAATGCGTTCTATATGCGTCTCAG CTGCGGC-
AAGCAGAAGACGGCATACGAGATGAGGTGGCTTGCGTGCACGAGAAGCATCGCCT CTCGAAGC-
AAGCAGAAGACGGCATACGAGATGAGGTGGTGACTGGAGTTCAGACGTGTGCTCTT CCGATCT-
AAGCAGAAGACGGCATACGAGATGAGGTGGTTAGCACTCGGCCGCAATTCTGAGT

IDT Exhibit 1020.3641

AATCTGGC-
AAGCAGAAGACGGCATACGAGATGAGGTGGGGCCTGTCGCGGTCCGAGCGATAAG CACGATCT-
AAGCAGAAGACGGCATACGAGATGAGGTGGTGACTGCTCATTGTGCATGTGGAGC GATTACCCAGT-
AAGCAGAAGACGGCATACGAGATGAGGTGGGCTTGACTGGAGATGCGTAAAGCTT GACGACGATCT-
AAGCAGAAGACGGCATACGAGATGAGGTGGTGATGATACCCGATTCGCACCTGCG AAACGTGTTCTATG-
AAGCAGAAGACGGCATACGAGATGAGGTGGACTTCATACGCAATTCGAATCTACGC CACGTGTTCTTTGCGA-
AAGCAGAAGACGGCATACGAGATGAGGTGGTACGCAATTCGAATCTACGCCACGT GTTCTTTGCGA-
AAGCAGAAGACGGCATACGAGATGAGGTGGGCTTGACTACTGGAGATGCGTAAAG CTTGACGACGATCT-
AAGCAGAAGACGGCATACGAGATGAGGTGGCTTGCGTGCACGAGATTCAGCATCG CCTCTCGAGGAAGC-
AAGCAGAAGACGGCATACGAGATGAGGTGGCTGCTGTCTTTCCCTCGTTTTCTCAA GTTTGCGCAC-
AAGCAGAAGACGGCATACGAGATGAGGTGGTGATCGTCTTGCAGAATCGTGGACA GCTAGTCTGCT-
AAGCAGAAGACGGCATACGAGATGAGGTGGAGATACCGACGCGATGAAGCACGTT GCACCCTT-
AAGCAGAAGACGGCATACGAGATGAGGTGGTCGGATGAGCGAAGTTGCAATCCCG AACTTTCATGC-
AAGCAGAAGACGGCATACGAGATGAGGTGGAGATCGGAATTCCACACGTCTGAAT AACAGTCA-
AAGCAGAAGACGGCATACGAGATGAGGTGGGCCGCAGCTGAGACGCATATAGAAC GCATTGGGCGA-

[00116] **Table 2**: Primer sequences directed against second adapter listed in **Table 4**.

Primer (5'-3')
AATCTGACGATAACCGATGAGTCATACTCGCTTGGACTATACGACTGCCTTGTTCA
AATCTGACGATAACCGATGAGTCATACTCGCTTGGACTATACGACTGCCTTGTTCA
TTCGCATTACGTCTCGCATCTTACGATGGAGATCGTGCTGCTCTGGATACTGGCGA
AATGATTCCCGTTGCTCAATGGGAAGGCTTCTACACGACTGCGACCGCCG

IDT Exhibit 1020.3642

GCTACTCAGACGGCGACCTGCGCTTTGTGCTCTCGAAGCCGTCACGACCGAGTGGC CCA
CCTGATCCAGCGAGCTCATTGGAGATCTACACTCTGTATGTTGGCATTGACCCAGAC TCCTT
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGAT
AATCCAACGGCGGCTGGTGAGATCTACACTGAAGGAATGCTACACGACGTTAGACC CTT
TCGGACACGACGACTAGCGTCATGTGCTCTCATTCCCTACACGACCATCTGCACTT
AATGATACATCGACCTACGAGATCTACTGTGACGCTCCACTCGACGTCGTAGCTTA
TTTGATACGACCTCAGTGGAGATCTACACTCTTTCCCTAGATGACGCTGAAACTAG
ATTGTGACGATAACGGATGTGTCATACTCGCTTTGCCTAATCGACACGCTTCTTGA
AATCTGACGATAACCGATGAGTCATACTCGCTTGGACTATACGACTGCGAACTTGTT CA
TTTGATACGACCTCAGTGGAGATCTACACTCTTTCCCTAGATGACGCTTCTCGAGAA ACTAG
AATGATACGTTTGCGACCACCGAGATCTACACTCTTTCCCTACACGACAGAGTTCCG ATC
TCGGACACGACGACTAGCGTCATGTGCTCTCATTCCCTACACGACTGTCTGCAGCAT
AAGGTTTCCCGTTGCTCGATGGCAAGGCATGTACTCGACCGTGACGGTCCGG
TCGTTCACGACGACTAGCCTCATGTGCTCTCTTTGCCTACGTCTCGAACTGTAGGTA G
TCGTTCACGACGACTAGCCTCATGTGCTCTCTTTGCCTACGTCTCGTCGTCTTCCTCT
TACCTTACGCCGACCACCGACTACTAGACTGTATGCCTACACGACTCAGATGAAGT T
TGAACAAGGCAGTCGTATAGTCCAAGCGAGTATGACTCATCGGTTATCGTCAGATT
TGAACAAGGCAGTCGTATAGTCCAAGCGAGTATGACTCATCGGTTATCGTCAGATT
TCGCCAGTATCCAGAGCAGCACGATCTCCATCGTAAGATGCGAGACGTAATGCGAA
CGGCGGTCGCAGTCGTGTAGAAGCCTTCCCATTGAGCAACGGGAATCATT
TGGGCCACTCGGTCGTGACGGCTTCGAGAGCACAAAGCGCAGGTCGCCGTCTGAGT AGC
AAGGAGTCTGGGTCAATGCCAACATACAGAGTGTAGATCTCCAATGAGCTCGCTGG ATCAGG
ATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT
AAGGGTCTAACGTCGTGTAGCATTCCTTCAGTGTAGATCTCACCAGCCGCCGTTGGA TT
AAGTGCAGATGGTCGTGTAGGGAATGAGAGCACATGACGCTAGTCGTCGTGTCCGA

IDT Exhibit 1020.3643

TAAGCTACGACGTCGAGTGGAGCGTCACAGTAGATCTCGTAGGTCGATGTATCATT
CTAGTTTCAGCGTCATCTAGGGAAAGAGTGTAGATCTCCACTGAGGTCGTATCAAA
TCAAGAAGCGTGTCGATTAGGCAAAGCGAGTATGACACATCCGTTATCGTCACAAT
TGAACAAGTTCGCAGTCGTATAGTCCAAGCGAGTATGACTCATCGGTTATCGTCAG ATT
CTAGTTTCTCGAGAAGCGTCATCTAGGGAAAGAGTGTAGATCTCCACTGAGGTCGT ATCAAA
GATCGGAACTCTGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCAAACGTAT CATT
ATGCTGCAGACAGTCGTGTAGGGAATGAGAGCACATGACGCTAGTCGTCGTGTCCG A
CCGGACCGTCACGGTCGAGTACATGCCTTGCCATCGAGCAACGGGAAACCTT
CTACCTACAGTTCGAGACGTAGGCAAAGAGAGCACATGAGGCTAGTCGTCGTGAAC GA
AGAGGAAGACGACGAGACGTAGGCAAAGAGAGCACATGAGGCTAGTCGTCGTGAA CGA
AACTTCATCTGAGTCGTGTAGGCATACAGTCTAGTAGTCGGTGGTCGGCGTAAGGT A

[00117] The term "adapter" can refer to an oligonucleotide of known sequence, the ligation of which to a target polynucleotide or a target polynucleotide strand of interest enables the generation of amplification-ready products of the target polynucleotide or the target polynucleotide strand of interest. Various adapter designs can be used. Suitable adapter molecules include single or double stranded nucleic acid (DNA,RNA,or a combination thereof) molecules or derivatives thereof, stem-loop nucleic acid molecules, double stranded molecules comprising one or more single stranded overhangs of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 bases or longer, proteins, peptides, aptamers, organic molecules, small organic molecules, or any adapter molecules known in the art that can be covalently or non-covalently attached, such as for example by ligation, to the double stranded nucleic acid fragments. The adapters can be designed to comprise a double-stranded portion which can be ligated to double-stranded nucleic acid (or double-stranded nucleic acid with overhang) products.

[00118] Adapter oligonucleotides can have any suitable length, at least sufficient to accommodate the one or more sequence elements of which they are comprised. In some cases, adapters are about, less than about, or more than about 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100, 200, or more nucleotides in length. In some cases, the adapter is stem-loop or hairpin adapter, wherein the stem of the hairpin adapter is about, less than about, or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, or more nucleotides in length. Stems can be designed using a variety of different sequences that result in hybridization between the complementary regions on a hairpin

IDT Exhibit 1020.3644

adapter, resulting in a local region of double-stranded DNA. For example, stem sequences can be utilized that are from 15 to 18 nucleotides in length with equal representation of G:C and A:T base pairs. Such stem sequences are predicted to form stable dsDNA structures below their predicted melting temperatures of .about.45 degree C. Sequences participating in the stem of the hairpin can be perfectly complementary, such that each base of one region in the stem hybridizes via hydrogen bonding with each base in the other region in the stem according to Watson-Crick base-pairing rules. Alternatively, sequences in the stem can deviate from perfect complementarity. For example, there can be mismatches and or bulges within the stem structure created by opposing bases that do not follow Watson-Crick base pairing rules, and/or one or more nucleotides in one region of the stem that do not have the one or more corresponding base positions in the other region participating in the stem. Mismatched sequences can be cleaved using enzymes that recognize mismatches. The stem of a hairpin can comprise DNA, RNA, or both DNA and RNA. In some cases, the stem and/or loop of a hairpin, or one or both of the hybridizable sequences forming the stem of a hairpin, comprise nucleotides, bonds, or sequences that are substrates for cleavage, such as by an enzyme, including but not limited to endonucleases and glycosylases. The composition of a stem can be such that only one of the hybridizable sequences forming the stem is cleaved. For example, one of the sequences forming the stem can comprise RNA while the other sequence forming the stem consists of DNA, such that cleavage by an enzyme that cleaves RNA in an RNA-DNA duplex, such as RNase H, cleaves only the sequence comprising RNA. One or both strands of a stem and/or loop of a hairpin can comprise about, more than, less than, or at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 non-canonical nucleotides (e.g. uracil), and/or methylated nucleotides. In some cases, the loop sequence of a hairpin adapter is about, less than about, or more than about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, or more nucleotides in length.

[00119] An adapter can comprise at least two nucleotides covalently linked together. An adapter as used herein can contain phosphodiester bonds, although in some cases, as outlined below, nucleic acid analogs are included that can have alternate backbones, comprising, for example, phosphoramide (Beaucage et al., Tetrahedron 49(10):1925 (1993) and references therein; Letsinger, J. Org. Chem. 35:3800 (1970); Sprinzl et al., Eur. J. Biochem. 81:579 (1977); Letsinger et al., Nucl. Acids Res. 14:3487 (1986); Sawai et al, Chem. Lett. 805 (1984), Letsinger et al., J. Am. Chem. Soc. 110:4470 (1988); and Pauwels et al., Chemica Scripta 26:141 91986)), phosphorothioate (Mag et al., Nucleic Acids Res. 19:1437 (1991); and U.S. Pat. No. 5,644,048), phosphorodithioate (Briu et al., J. Am. Chem. Soc. 111:2321 (1989), O-methylphosphoroamidite linkages (see Eckstein, Oligonucleotides and Analogues: A Practical Approach, Oxford University Press), and peptide nucleic acid (also referred to herein as "PNA") backbones and linkages (see Egholm, J. Am. Chem. Soc. 114:1895 (1992); Meier et al., Chem. Int. Ed. Engl. 31:1008 (1992); Nielsen, Nature, 365:566 (1993); Carlsson et al., Nature 380:207 (1996), all of which are incorporated by reference). Other analog nucleic acids include those with bicyclic structures including locked nucleic acids (also referred to herein as "LNA"), Koshkin et al., J. Am. Chem. Soc. 120.13252 3 (1998); positive backbones (Denpcy et al., Proc. Natl. Acad. Sci. USA 92:6097 (1995); non-ionic

IDT Exhibit 1020.3645

backbones (U.S. Pat. Nos. 5,386,023, 5,637,684, 5,602,240, 5,216,141 and 4,469,863; Kiedrowshi et al., Angew. Chem. Intl. Ed. English 30:423 (1991); Letsinger et al., J. Am. Chem. Soc. 110:4470 (1988); Letsinger et al., Nucleoside & Nucleotide 13:1597 (1994); Chapters 2 and 3, ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research", Ed. Y. S. Sanghui and P. Dan Cook; Mesmaeker et al., Bioorganic & Medicinal Chem. Lett. 4:395 (1994); Jeffs et al., J. Biomolecular NMR 34:17 (1994); Tetrahedron Lett. 37:743 (1996)) and non-ribose backbones, including those described in U.S. Pat. Nos. 5,235,033 and 5,034,506, and Chapters 6 and 7, ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research", Ed. Y. S. Sanghui and P. Dan Cook. Nucleic acids containing one or more carbocyclic sugars are also included within the definition of nucleic acids (see Jenkins et al., Chem. Soc. Rev. (1995) pp 169 176). Several nucleic acid analogs are described in Rawls, C & E News Jun. 2, 1997 page 35. "Locked nucleic acids" are also included within the definition of nucleic acid analogs. LNAs are a class of nucleic acid analogues in which the ribose ring is "locked" by a methylene bridge connecting the 2'-O atom with the 4'-C atom. All of these references are hereby expressly incorporated by reference. These modifications of the ribose-phosphate backbone can be done to increase the stability and half-life of such molecules in physiological environments. For example, PNA:DNA and LNA-DNA hybrids can exhibit higher stability and thus can be used in some cases. Adapters can be single stranded or double stranded, as specified, or contain portions of both double stranded or single stranded sequence. Depending on the application, adapters can be DNA, RNA, or a hybrid, where the adapter contains any combination of deoxyribo- and ribo-nucleotides, and any combination of bases, including uracil, adenine, thymine, cytosine, guanine, inosine, xathanine hypoxathanine, isocytosine, isoguanine, etc.

[00120] As illustrated in **FIG. 2**, the first adapter as provided herein can be a double stranded nucleic acid or single stranded nucleic acid comprising a 3' overhang. As shown in **I** of **FIG. 2**, the first adapter comprises a partial duplex between two oligonucleotides, wherein a first oligonucleotide comprises a long strand comprising a known sequence, A, at the 5' end and a 3' overhang and a second oligonucleotide comprises a short strand comprising sequence complementary to sequence A, A', at the 3' end. The short strand in **I** of **FIG. 2** further comprises a block at the 3' and 5' end, which can serve to inhibit ligation. In some cases, the long strand comprises a block at the 5' end, thereby inhibiting ligation. As shown in **II** of **FIG. 2**, the first adapter comprises a single stranded oligonucleotide, wherein the 5' end of the oligonucleotide binds to a known sequence, A, located near the 3' end of the oligonucleotide, wherein the 5' end comprises sequence complementary to sequence A, A', and wherein the binding produces a 3' overhang. The 5' end and 3' end of the single stranded oligonucleotide adapter in **II** of **FIG. 2** can be connected through a linker. The linker can be a stem loop, non-nucleotide linker, or a combination thereof. The stem loop can comprise DNA, RNA, nucleotide analogs, or combinations thereof. The 5' end of the single stranded oligonucleotide adapter in **II** of **FIG. 2** can comprise a 5' block, which can inhibit ligation. Various constructs for useful second adaptors are anticipated. The second adaptors useful for carrying out the methods for producing directional polynucleotide libraries as provided herein

IDT Exhibit 1020.3646

can be dsDNA, partial duplex or stem-loop adaptors with one end suitable for ligation to the end of the dsDNA products produced by the methods provided herein, and the like. In some cases, a second adapter comprises a partial duplex between two oligonucleotides, wherein a first oligonucleotide comprises a long strand comprising a known sequence, B, and a second oligonucleotide comprising a short strand comprising sequence complementary to a portion of sequence B, B', wherein binding between the long strand and short strand generates a 3' overhang. The short strand of the second adapter can further comprise a block at the 3' and/or 5' end, which can serve to inhibit ligation. The 3' end of the long strand can comprise a block at the 3' end. In some cases, a second adapter comprises a partial duplex between two oligonucleotides, wherein a first oligonucleotide comprises a long strand comprising a known sequence, B, and a second oligonucleotide comprising a short strand comprising sequence complementary to a portion of sequence B, B', wherein binding between the long strand and short strand generates a 5' overhang. The short strand of the second adapter can further comprise a block at the 5' end, which can serve to inhibit ligation. The 3' and/or 5'end of the long strand can comprise a block, which can inhibit ligation. A block in any of the adapters provided herein can be any of the blocks provided herein. Adapters for use in the methods provided herein can be any of the first and/or second adapters listed in **Tables 3** and **4**.

[00121] Table 3: First adapter sequences for use in the methods provided herein.

Oligo A	Oligo B
CTG CTG TCT TTC CCT CGT TTT CTC AAG CGA CAC NNN NNN NNN	/5BioTEG/GTG TCG CTT GAG AAA ACG AGG GAA AGA CAG CAG /3AmMC6T/
TGA TCG GAG TGC AGA ATC GTG GAC TTC TAG TCT NNN NNN	/5BiodT/AGA CTA GAA GTC CAC GAT TCT GCA CTC CGA TCA /3AzideN/
CCC AAT GCG TTC TAT ATG CGT CTC AGC TGC GGC NNN NNN N	/5Biosg/GCC GCA GCT GAG ACG CAT ATA GAA CGC ATT GGG /3AmMC6T/
CTT GCG TGC ACG AGA AGC ATC GCC TCT CGA AGC NNN NNN NN	/5BioTEG/GCT TCG AGA GGC GAT GCT TCT CGT GCA CGC AAG /3ThiolMCD-6/
TGA CTG GAG TTC AGA CGT GTG CTC TTC CGA TCT NNN NNN NN	/5Biosg/AGA TCG GAA GAG CAC ACG TCT GAA CTC CAG TCA /3AmMO/
TTA GCA CTC GGC CGC AAT TCT GAG TAA TCT GGC NNN NNN NNN	/5DTPA/GCC AGA TTA CTC AGA ATT GCG GCC GAGTGC TAA /3AmMC6T/
GGC CTG TCG CGG TCC GAG CGA TAA GCA CGA TCT NNN NNN NNN N	/5DPTA/ACT GGG TAA TCG CTC CAC ATG CAC AAT GAG CAG TCA /3AmMO/

IDT Exhibit 1020.3647

TGA CTG CTC ATT GTG CAT GTG GAG CGA TTA CCC AGT NNN NNN NN	/5DPTA/ACT GGG TAA TCG CTC CAC ATG CAC AAT GAG CAG TCA /3AmMO/
GCT TGA CTG GAG ATG CGT AAA GCT TGA CGA CGA TCT NNN NNN	/52-Bio/AGA TCG TCG TCA AGC TTT AGC CAT CTC CAG TCA AGC/3AmMO/
TGA TGA TAC CCG ATT CGC ACC TGC GAA ACG TGT TCT ATG NNN NNNNN	/5BioTEG/CAT AGA ACA CGT TTC GCA GGT GCG AAT CGG GTA TCA TCA/33ThiolMC3-D/
ACT TCA TAC GCA ATT CGA ATC TAC GCC ACG TGT TCT TTG CGA NNN NNNNN	/5BioTEG/TCG CAA AGA ACA CGT GGC GTA GAT TCG AAT TGC GTA TGA AGT/33ThiolMC3-D/
TAC GCA ATT CGA ATC TAC GCC ACG TGT TCT TTG CGA NNN NNNNN	/5BioTEG/TCG CAA AGA ACA CGT GGC GTA GAT TCG AAT TGC GTA/33ThiolMC3-D/
GCT TGA CTA CTG GAG ATG CGT AAA GCT TGA CGA CGA TCT NNN NNN	/52-Bio/AGA TCG TCG TCA AGC TTT AGC CAT CTC CAG TAG TCA AGC/3AmMO/
CTT GCG TGC ACG AGA TTC AGC ATC GCC TCT CGA GGA AGC NNN NNN NN	/5BioTEG/GCT TCC TCG AGA GGC GAT GCT GAA TCT CGT GCA CGC AAG /3ThiolMCD-6/
CTG CTG TCT TTC CCT CGT TTT CTC AAG TTT GCG CAC NNN NNN NNN	/5BioTEG/GTG CGC AAA CTT GAG AAA ACG AGG GAA AGA CAG CAG /3AmMC6T/
TGA TCG TCT TGC AGA ATC GTG GAC AGC TAG TCT GCT NNN NNN	/5BiodT/AGC AGA CTA GCT GTC CAC GAT TCT GCA AGA CGA TCA /3AzideN/
AGA TAC CGA CGC GAT GAA GCA CGT TGC ACC CTT-NNN-NNN-NN	/5BioTEG/AAG GGT GCA ACG TGC TTC ATC GCG TCG GTA TCT/3AmMC6T/
TCG GAT GAG CGA AGT TGC AAT CCC GAA CTT TCA TGC-NNN-NNN	/5BioTEG/GCA TGA AAG TTC GGG ATT GCA ACT TCG CTC ATC CGA/3ThiolMCD-6/
AGA TCG GAA TTC CAC ACG TCT GAA TAA CAG TCA-NNN-NNN-N	/5BioTEG/TGA CTG TTA TTC AGA CGT GTG GAA TTC CGA TCT/3ThiolMCD-6/
GCC GCA GCT GAG ACG CAT ATA GAA CGC ATT GGG CGA NNN NNN N	/5Biosg/ TCG CCC AAT GCG TTC TAT ATG CGT CTC AGC TGC GGC /3AmMC6T/

IDT Exhibit 1020.3648

*/5Biosg/ TCG CCC AAT GCG TTC TAT ATG CGT CTC AGC TGC GGC ATT CAA GCC GCA GCT GAG ACG CAT ATA GAA CGC ATT GGG CGA NNN NNN N *Single stranded stem-loop first adapter; underlined sequence represents loop nucleotides	
*/5DPTA/ACT GGG TAA TCG CTC CAC ATG CAC AAT GAG CAG TCA ATT CAA TGA CTG CTC ATT GTG CAT GTG GAG CGA TTA CCC AGT NNN NNN NN *Single stranded stem-loop first adapter; underlined sequence represents loop nucleotides	
*/5BioTEG/GTG TCG CTT GAG AAA ACG AGG GAA AGA CAG CAG ATT CAA CTG CTG TCT TTC CCT CGT TTT CTC AAG CGA CAC NNN NNN NNN *Single stranded stem-loop first adapter; underlined sequence represents loop nucleotides	

[00122] Table 4: Second adapter sequences for use in the methods provided herein.

Oligo A	Oligo B
AATCTGACGATAACCGATGAGTCATACTCG CTT GGACTATACGACTGCCTTGTTCAGT	/5BioTEG /A*GTGCATCCTAG*/3ddC/
AATCTGACGATAACCGATGAGTCATACTCG CTTGGACTATACGACTGCCTTGTTCAGT	/5Biosg/A*CTGAACAAGGC*/3ddA/
TTCGCATTACGTCTCGCATCTTACGATGGA GATCGTGCTGCTCTGGATACTGGCGAAC	/52-Bio /G*TTCGCCAGTAT*/3ddC/
AATGATTCCCGTTGCTCAATGGGAAGGCTT CTACACGACTGCGAC CGCCGGA	/5Biosg/T*CCGGCGGTCGC*/3ddA/
GCTACTCAGACGGCGACCTGCGCTTTGTGC TCTCGAAGCCGTCACGACCGAGTGGCCCAG TC	5DPTA/G*ACTGGGCCACTC*/3ddG/

IDT Exhibit 1020.3649

CCTGATCCAGCGAGCTCATTGGAGATCTAC ACTCTGTATGTTGGCATTGACCCAGACTCC TTGA	/5BioTEG/T*CAAGGAGTCTG*/3ddG/
AATGATACGGCGACCACCGAGATCTACACT CTTTCCCTACACGACGCTCTTCCGATCT	/5Biosg/A*GATCGGAAGAG*/3ddC/
AATCCAACGGCGGCTGGTGAGATCTACACT GAAGGAATGCTACACGACGTTAGACCCTTG A	/5Biosg/T*CAAGGGTCTAA*/3ddC/
TCGGACACGACGACTAGCGTCATGTGCTCT CATTCCCTACACGACCATCTGCACTTGT	/5BioTEG/A*CAAGTGCAGAT*/3ddG/
AATGATACATCGACCTACGAGATCTACTGT GACGCTCCACTCGACGTCGTAGCTTAGT	/5Biosg/A*TAAGCTACGA*/3ddC/
TTTGATACGACCTCAGTGGAGATCTACACT CTTTCCCTAGATGACGCTGAAACTAGCG	/5Biosg/C*GCTAGTTTCAG*/3ddC/
ATTGTGACGATAACGGATGTGTCATACTCG CTTTGCCTAATCGACACGCTTCTTGATG	/5BioTEG/C*ATCAAGAAGCG*/3ddT/
AATCTGACGATAACCGATGAGTCATACTCG CTTGGACTATACGACTGCGAACTTGTTCAG T	/5Biosg/A*CTGAACAAGTTCGC*/3ddA/
TTTGATACGACCTCAGTGGAGATCTACACT CTTTCCCTAGATGACGCTTCTCGAGAAACT AGCG	/5Biosg/C*GCTAGTTTCTCGAGAAG*/3ddC/
AATGATACGTTTGCGACCACCGAGATCTAC ACTCTTTCCCTACACGACAGAGTTCCGATC TA	/5Biosg/T*AGATCGGAACTC*/3ddT/
TCGGACACGACGACTAGCGTCATGTGCTCT CATTCCCTACACGACTGTCTGCAGCATGA	/5BioTEG/T*CATGCTGCAGAC*/3ddA/

AAGGTTTCCCGTTGCTCGATGGCAAGGCAT GTACTCGACCGTGACGGTCCGGAG	/5Biosg/C*TCCGGACCGTCAC*/3ddG/
TCGTTCACGACGACTAGCCTCATGTGCTCT CTTTGCCTACGTCTCGAACTGTAGGTAGTA	/5BioTEG/T*ACTACCTACAGTT*/3ddC/
TCGTTCACGACGACTAGCCTCATGTGCTCT CTTTGCCTACGTCTCGTCGTCTTCCTCTCG	/5DPTA/C*GAGAGGAAGACGA*/3ddC/
TACCTTACGCCGACCACCGACTACTAGACT GTATGCCTACACGACTCAGATGAAGTTGT	/5DPTA/A*CAACTTCATCTG*/3ddA/

[00123] Various ligation processes and reagents are known in the art and can be useful for carrying out the methods provided herein. For example, blunt ligation can be employed. Similarly, a single dA nucleotide can be added to the 3′-end of the double-stranded DNA product, by a polymerase lacking 3′-exonuclease activity and can anneal to an adapter comprising a dT overhang (or the reverse). This design allows the hybridized components to be subsequently ligated (e.g., by T4 DNA ligase). Other ligation strategies and the corresponding reagents and known in the art and kits and reagents for carrying out efficient ligation reactions are commercially available (e.g, from New England Biolabs, Roche).

VII. Blocking Groups

[00124] Any of the adapters and/or primers used in the methods for generating directional polynucleotide libraries as provided herein can comprise a blocking group at the 5' and/or 3' end. Adapters and/or primers comprising a duplex or partial duplex can comprise a block at the 5' and/ or 3' end of one or both strands forming the duplex or partial duplex. A blocked end in any of the adapters or primers provided herein can be enzymatically unreactive to prevent adapter dimer formation and/or ligation. The blocking group can be a dideoxynucleotide (ddCMP, ddAMP, ddTMP, or ddGMP), various modified nucleotides (e.g. phosphorothioate-modified nucleotides), or non-nucleotide chemical moieties. In some cases, the blocking group comprises a nucleotide analog that comprises a blocking moiety. The blocking moiety can mean a part of the nucleotide analog that inhibits or prevents the nucleotide analog from forming a covalent linkage to a second nucleotide or nucleotide analog. For example, in the case of nucleotide analogs having a pentose moiety, a reversible blocking moiety can prevent formation of a phosphodiester bond between the 3′ oxygen of the nucleotide and the 5′ phosphate of the second nucleotide. Reversible blocking moieties can include phosphates, phosphodiesters, phosphotriesters, phosphorothioate esters,

IDT Exhibit 1020.3651

and carbon esters. In some cases, a blocking moiety can be attached to the 3′ position or 2′ position of a pentose moiety of a nucleotide analog. A reversible blocking moiety can be removed with a deblocking agent. The blocking group at a 5' and/or 3' end can be a spacer (C3 phosphoramidite, triethylene glycol (TEG), photo-cleavable, hexa-ethyleneglycol), inverted dideoxy-T, biotin, thiol, dithiol, hexanediol, digoxigenin, an azide, alkynes, or an amino modifier. A biotin blocking group can be photocleavable biotin, biotin-triethylene glycol (TEG), biotin-dT, desthiobiotin-TEG, biotin-azide, or dual biotin. A block at a 5' end can comprise a nucleotide at a 5' end that lacks a 5' phosphate. The 5' end can be removed by treatment with an enzyme. The enzyme can be a phosphatase. A block at a 3' end can comprise a nucleotide that lacks a free 3' hydroxyl. The ends (i.e. 5' and/or 3' ends) can further comprise phosphothioate bonds. The phosphothioate bonds can serve to protect any adapter or primer comprising the phosphothioate bond. The protection can be from nuclease degradation.

VIII. RNA-Dependent DNA Polymerases

[00125] RNA-dependent DNA polymerases for use in the methods and compositions provided herein can be capable of effecting extension of a primer according to the methods provided herein. Accordingly, an RNA-dependent DNA polymerase can be one that is capable of extending a nucleic acid primer along a nucleic acid template that is comprised at least predominantly of ribonucleotides. Suitable RNA-dependent DNA polymerases for use in the methods, compositions, and kits provided herein include reverse transcriptases (RTs). RTs are well known in the art. Examples of RTs include, but are not limited to, moloney murine leukemia virus (M- MLV) reverse transcriptase, human immunodeficiency virus (HIV) reverse transcriptase, rous sarcoma virus (RSV) reverse transcriptase, avian myeloblastosis virus (AMV) reverse transcriptase, rous associated virus (RAV) reverse transcriptase, and myeloblastosis associated virus (MAV) reverse transcriptase or other avian sarcoma-leukosis virus (ASLV) reverse transcriptases, and modified RTs derived therefrom. See e.g. US7056716. Many reverse transcriptases, such as those from avian myeoloblastosis virus (AMV-RT), and Moloney murine leukemia virus (MMLV-RT) comprise more than one activity (for example, polymerase activity and ribonuclease activity) and can function in the formation of the double stranded cDNA molecules. However, in some instances, it is preferable to employ a RT which lacks or has substantially reduced RNase H activity. RTs devoid of RNase H activity are known in the art, including those comprising a mutation of the wild type reverse transcriptase where the mutation eliminates the RNase H activity. Examples of RTs having reduced RNase H activity are described, e.g., in US20100203597. In these cases, the addition of an RNase H from other sources, such as that isolated from E. coli, can be employed for the degradation of the starting RNA sample and the formation of the double stranded cDNA. Combinations of RTs can also contemplated, including combinations of different non-mutant RTs, combinations of different mutant RTs, and combinations of one or more non-mutant RT with one or more mutant RT.

IDT Exhibit 1020.3652

IX. DNA-Dependent DNA Polymerases

[00126] DNA-dependent DNA polymerases for use in the methods and compositions provided herein can be capable of effecting extension of a nucleic acid comprising a free 3' hydroxyl. The nucleic acid comprising a free 3' hydroxyl can be on a primer and/or adapter as provided herein. The nucleic acid comprising a free 3' hydroxyl can be on a strand of a dsDNA (e.g. genomic DNA) generated by treatment of the dsDNA (e.g. genomic DNA) with a nicking enzyme. A DNA-dependent DNA polymerase can be one that is capable of extending a free 3' OH along a first strand cDNA in the presence of the RNA template or after selective removal of the RNA template. Exemplary DNA dependent DNA polymerases suitable for the methods provided herein include but are not limited to Klenow polymerase, with or without 3'-exonuclease, Bst DNA polymerase, Bca polymerase, .phi.29 DNA polymerase, Vent polymerase, Deep Vent polymerase, Taq polymerase, T4 polymerase, and E. coli DNA polymerase 1, derivatives thereof, or mixture of polymerases. In some cases, the polymerase does not comprise a 5'-exonuclease activity. In other cases, the polymerase comprises 5' exonuclease activity. In some cases, the extension of a free 3' OH can be performed using a polymerase comprising strong strand displacement activity such as, for example, Bst polymerase. In other cases, the extension of the free 3' OH can be performed using a polymerase comprising weak or no strand displacement activity. One skilled in the art can recognize the advantages and disadvantages of the use of strand displacement activity during any extension step in the methods provided herein, and which polymerases can be expected to provide strand displacement activity (see e.g., New England Biolabs Polymerases). For example, strand displacement activity can be useful in ensuring whole transcriptome coverage during the random priming and extension step or ensuring whole genomc coverage during the extension step following treatment of genomic DNA with a nicking enzyme.

[00127] In some cases, the double stranded products or fragments generated by the methods described herein can be end repaired to produce blunt ends for the adapter ligation applications described herein. Generation of the blunt ends on the double stranded products can be generated by the use of a single strand specific DNA exonuclease such as for example exonuclease 1, exonuclease 7 or a combination thereof to degrade overhanging single stranded ends of the double stranded products. Alternatively, the double stranded products can be blunt ended by the use of a single stranded specific DNA endonuclease for example but not limited to mung bean endonuclease or S1 endonuclease. Alternatively, the double stranded products can be blunt ended by the use of a polymerase that comprises single stranded exonuclease activity such as for example T4 DNA polymerase, any other polymerase comprising single stranded exonuclease activity or a combination thereof to degrade the overhanging single stranded ends of the double stranded products or fragments. In some cases, the polymerase comprising single stranded exonuclease activity can be incubated in a reaction mixture that does or does not comprise one or more dNTPs. In other cases, a combination of single stranded nucleic acid specific exonucleases and one or more polymerases can be used to blunt end the double stranded products of the extension reaction. In still other cases, the products of an extension reaction as provided herein can be made blunt ended by filling in

IDT Exhibit 1020.3653

the overhanging single stranded ends of the double stranded products. For example, the fragments can be incubated with a polymerase such as T4 DNA polymerase or Klenow polymerase or a combination thereof in the presence of one or more dNTPs to fill in the single stranded portions of the double stranded products. Alternatively, the double stranded products or fragments can be made blunt by a combination of a single stranded overhang degradation reaction using exonucleases and/or polymerases, and a fill-in reaction using one or more polymerases in the presence of one or more dNTPs.

[00128] In another embodiment, the adapter ligation applications described herein can leave a gap between one strand (e.g. non-ligation strand) of an adapters and a strand of a double stranded product or fragment. In these instances, a gap repair or fill-in reaction can be used to append the double stranded product or fragment with the sequence complementary to the other strand (e.g. ligation strand) of the adapter. Gap repair can be performed with any number of DNA dependent DNA polymerase described herein. In some cases, gap repair can be performed with a DNA dependent DNA polymerase with strand displacement activity. In some cases, gap repair can be performed using a DNA dependent DNA polymerase with weak or no strand displacement activity. In some cases, the ligation strand of the adapter can serve as the template for the gap repair or fill-in reaction. In some cases, gap repair can be performed using *Taq* DNA polymerase.

X. Cleavage Agents

[00129] The selective removal or cleavage of a polynucleotide comprising a non-canonical dNTP generated by the methods provided herein can be achieved through the use of enzymatic treatment of the polynucleotide. Enzymes that can be used for cleavage of the marked strand generated by the methods provided herein can include glycosylases such as Uracil-N-Glycosylase (UNG), which can selectively degrade the base portion of dUTP. Additional glycosylases which can be used to generate a first strand cDNA or polynucleotides comprising one or more non-canonical nucleotides as provided herein and their non-canonical or modified nucleotide substrates include 5-methylcytosine DNA glycosylase (5-MCDG), which can cleave the base portion of 5-methylcytosine (5-MeC) from the DNA backbone (Wolffe et al., Proc. Nat. Acad. Sci. USA 96:5894-5896, 1999); 3-methyladenosine-DNA glycosylase I, which can cleave the base portion of 3-methyl adenosine from the DNA backbone (see, e.g. Hollis et al (2000) Mutation Res. 460: 201-210); and/or 3-methyladenosine DNA glycosylase II, which can cleave the base portion of 3-methyladenosine, 7-methylguanine, 7-methyladenosine, and/3-methylguanine from the DNA backbone. See McCarthy et al (1984) EMBO J. 3:545-550. Multifunctional and mono-functional forms of 5-MCDG have been described. See Zhu et al., Proc. Natl. Acad. Sci. USA 98:5031-6, 2001; Zhu et al., Nuc. Acid Res. 28:4157-4165, 2000; and Neddermann et al., J. B. C. 271:12767-74, 1996 (describing bifunctional 5-MCDG; Vairapandi & Duker, Oncogene 13:933-938, 1996; Vairapandi et al., J. Cell. Biochem. 79:249-260, 2000 (describing mono-functional enzyme comprising 5-MCDG activity). In some cases, 5-MCDG preferentially cleaves fully methylated polynucleotide sites (e.g., CpG dinucleotides), and in other cases, 5-MCDG preferentially cleaves a hemi-methylated polynucleotide. For

example, mono-functional human 5-methylcytosine DNA glycosylase cleaves DNA specifically at fully methylated CpG sites, and can be relatively inactive on hemimethylated DNA (Vairapandi & Duker, supra; Vairapandi et al., supra). By contrast, chick embryo 5-methylcytosine-DNA glycosylase can have greater activity directed to hemi-methylated methylation sites. In some cases, the activity of 5-MCDG is potentiated (increased or enhanced) with accessory factors, such as recombinant CpG-rich RNA, ATP, RNA helicase enzyme, and proliferating cell nuclear antigen (PCNA). See U.S. Patent Publication No. 20020197639 A1. One or more agents can be used. In some cases, the one or more agents cleave a base portion of the same methylated nucleotide. In other cases, the one or more agents cleave a base portion of different methylated nucleotides. Treatment with two or more agents can be sequential or simultaneous.

[00130] In some cases, an abasic site in the DNA backbone of a first strand cDNA generated by the methods provided herein can be followed by fragmentation or cleavage of the backbone at the abasic site. Suitable agents (for example, an enzyme, a chemical and/or reaction conditions such as heat) capable of cleavage of the backbone at an abasic site include: heat treatment and/or chemical treatment (including basic conditions, acidic conditions, alkylating conditions, or amine mediated cleavage of abasic sites, (see e.g., McHugh and Knowland, Nucl. Acids Res. (1995) 23(10):1664-1670; Bioorgan. Med. Chem. (1991) 7:2351; Sugiyama, Chem. Res. Toxicol. (1994) 7: 673-83; Horn, Nucl. Acids. Res., (1988) 16:11559-71), and/ or the use of enzymes that catalyze cleavage of polynucleotides at abasic sites. For example, an enzyme that catalyzes cleavage of polynucleotides at abasic sites can be AP endonucleases (also called "apurinic, apyrimidinic endonucleases") (e.g., E. coli Endonuclease IV, available from Epicentre Tech., Inc, Madison Wis.), E. coli endonuclease III or endonuclease IV, E. coli exonuclease III in the presence of calcium ions. See, e.g. Lindahl, PNAS (1974) 71(9):3649-3653; Jendrisak, U.S. Pat. No. 6,190,865 B1; Shida, Nucleic Acids Res. (1996) 24(22):4572-76; Srivastava, J. Biol. Chem. (1998) 273(13):21203-209; Carey, Biochem. (1999) 38:16553-60; Chem Res Toxicol (1994) 7:673-683. As used herein "agent" encompasses reaction conditions such as heat. In some cases, the AP endonuclease, E. coli endonuclease IV, is used to cleave the phosphodiester backbone or phosphodiester bond at an abasic site. In some cases, cleavage is with an amine, such as N,N'-dimethylethylenediamine (DMED). See, e.g., McHugh and Knowland, supra.

[00131] In some cases, the polynucleotide (e.g. first strand cDNA) comprising one or more abasic sites can be treated with a nucleophile or a base. In some cases, the nucleophile is an amine such as a primary amine, a secondary amine, or a tertiary amine. For example, the abasic site can be treated with piperidine, moropholine, or a combination thereof. In some cases, hot piperidine (e.g., 1M at 90^0 C) may be used to cleave a polynucleotide comprising one or more abasic sites. In some cases, morpholine (e.g., 3M at 37^0 C or 65^0 C) can be used to cleave the polynucleotide comprising one or more abasic sites. Alternatively, a polyamine can be used to cleave the polynucleotide comprising one or more abasic sites. Suitable polyamines include for example spermine, spermidine, 1,4-diaminobutane, lysine, the tripeptide K--W--K, DMED, piperazine, 1,2-ethylenediamine, or any combination thereof. In some cases, the polynucleotide comprising one or more abasic sites can be treated with a reagent suitable for carrying out

a beta elimination reaction, a delta elimination reaction, or a combination thereof. In some cases, the methods provided herein provide for the use of an enzyme or combination of enzymes and a polyamine such as DMED under mild conditions in a single reaction mixture which does not affect the canonical or unmodified nucleotides and therefore may maintain the sequence integrity of the products of the method. Suitable mild conditions can include conditions at or near neutral pH. Other suitable conditions include pH of about 4.5 or higher, 5 or higher, 5.5 or higher, 6 or higher, 6.5 or higher, 7 or higher, 7.5 or higher, 8 or higher, 8.5 or higher, 9 or higher, 9.5 or higher, 10 or higher, or about 10.5 or higher. Still other suitable conditions include between about 4.5 and 10.5, between about 5 and 10.0, between about 5.5 and 9.5, between about 6 and 9, between about 6.5 and 8.5, between about 6.5 and 8.0, or between about 7 and 8.0. Suitable mild conditions also can include conditions at or near room temperature. Other suitable conditions include a temperature of about 10^0 C, 11^0 C, 12^0 C, 13^0 C, 14^0 C, 15^0 C, 16^0 C, 17^0 C, 18^0 C, 19^0 C, 20^0 C, 21^0 C, 22^0 C, 23^0 C, 24^0 C, 25^0 C, 26^0 C, 27^0 C, 28^0 C, 29^0 C, 30^0 C, 31^0 C, 32^0 C, 33^0 C, 34^0 C, 35^0 C, 36^0 C, 37^0 C, 38^0C, 39^0 C, 40^0 C, 41^0 C, 42^0 C, 43^0 C, 44^0 C, 45^0 C, 46^0 C, 47^0 C, 48^0C, 49^0 C, 50^0 C, 51^0C, 52^0 C, 53^0 C, 54^0 C, 55^0C, 56^0 C, 57^0 C, 58^0 C, 59^0 C, 60^0 C, 61^0 C, 62^0 C, 63^0 C, 64^0 C, 65^0 C, 66^0 C, 67^0 C, 68^0 C, 69^0 C, or 70^0 C or higher. Still other suitable conditions include between about 10^0 C and about 70^0 C, between about 15^0 C and about 65^0 C, between about 20^0 C and about 60^0 C, between about 20^0 C and about 55^0 C, between about 20^0 C and about 50^0 C, between about 20^0 C and about 45^0 C, between about 20^0 C and about 40^0 C, between about 20^0 C and about 35^0 C, or between about 20^0 C and about 30^0 C. In some cases, the use of mild cleavage conditions can increase final product yields, maintain sequence integrity, or render the methods provided herein more suitable for automation.

[00132] In embodiments involving fragmentation, the backbone of the polynucleotide comprising the abasic site can be cleaved at the abasic site, whereby two or more fragments of the polynucleotide can be generated. At least one of the fragments can comprise an abasic site, as described herein. Agents that cleave the phosphodiester backbone or phosphodiester bonds of a polynucleotide at an abasic site are provided herein. In some embodiments, the agent is an AP endonuclease such as E. coli AP endonuclease IV. In other embodiments, the agent is DMED. In other embodiments, the agent is heat, basic condition, acidic conditions, or an alkylating agent. In still other embodiments, the agent that cleaves the phosphodiester backbone at an abasic site is the same agent that cleaves the base portion of a nucleotide to form an abasic site. For example, glycosylases of the methods provided herein can comprise both a glycosylase and a lyase activity, whereby the glycosylase activity cleaves the base portion of a nucleotide (e.g., a modified nucleotide) to form an abasic site and the lyase activity cleaves the phosphodiester backbone at the abasic site so formed. In some cases, the glycosylase comprises both a glycosylase activity and an AP endonuclease activity.

[00133] It can be desirable to use agents or conditions that can affect the cleavage of the backbone at the abasic site to generate fragments comprising a blocked 3'-end, which cannot be extendable by a polymerase when the 3'-end is hybridized to a first adapter according to the methods described herein.

IDT Exhibit 1020.3656

[00134] Appropriate reaction media and conditions for carrying out the cleavage of a base portion of a non-canonical or modified nucleotide according to the methods provided herein are those that permit cleavage of a base portion of a non-canonical or modified nucleotide. Such media and conditions are known to persons of skill in the art, and are described in various publications, such as Lindahl, PNAS (1974) 71(9):3649-3653; and Jendrisak, U.S. Pat. No. 6,190,865 B1; U.S. Pat. No. 5,035,996; and U.S. Pat. No. 5,418,149. In one embodiment, UDG (Epicentre Technologies, Madison Wis.) is added to a nucleic acid synthesis reaction mixture, and incubated at 37^0 C for 20 minutes. In one embodiment, the reaction conditions are the same for the synthesis of a polynucleotide comprising a non-canonical or modified nucleotide and the cleavage of a base portion of the non-canonical or modified nucleotide. In another embodiment, different reaction conditions are used for these reactions. In some embodiments, a chelating regent (e.g. EDTA) is added before or concurrently with UNG in order to prevent a polymerase from extending the ends of the cleavage products.

In a one embodiment, the selection is done by incorporation of at least one modified nucleotide into one strand of a synthesized polynucleotide, and the selective removal is by treatment with an enzyme that displays a specific activity towards the at least one modified nucleotide. In some cases, the modified nucleotide being incorporated into one strand of the synthesized polynucleotide is deoxyuridine triphosphate (dUTP), and the selective cleavage is carried by out by UNG. UNG selectively degrades dUTP while it is neutral towards other dNTPs and their analogs. Treatment with UNG results in the cleavage of the N-glycosylic bond and the removal of the base portion of dU residues, forming abasic sites. In one embodiment, the UNG treatment is done in the presence of an apurinic/apyrimidinic endonuclease (APE) to create nicks at the abasic sites. Consequently, a polynucleotide strand with incorporated dUTP that is treated with UNG/APE can be cleaved. In another case, nick generation and cleavage is achieved by treatment with a polyamine, such as DMED, or by heat treatment.

XI. Methods of Amplification

[00135] The methods, compositions and kits described herein can be useful to generate amplification-ready products for downstream applications such as massively parallel sequencing (i.e. next generation sequencing methods) or hybridization platforms. Methods of amplification are well known in the art. Examples of PCR techniques that can be used include, but are not limited to, quantitative PCR, quantitative fluorescent PCR (QF-PCR), multiplex fluorescent PCR (MF-PCR), real time PCR(RT-PCR), single cell PCR, restriction fragment length polymorphism PCR(PCR-RFLP), PCR-RFLP/RT-PCR-RFLP, hot start PCR, nested PCR, in situ polony PCR, in situ rolling circle amplification (RCA), bridge PCR, picotiter PCR, digital PCR, droplet digital PCR, and emulsion PCR. Other suitable amplification methods include the ligase chain reaction (LCR), transcription amplification, molecular inversion probe (MIP) PCR, self-sustained sequence replication, selective amplification of target polynucleotide sequences, consensus sequence primed polymerase chain reaction (CP-PCR), arbitrarily primed polymerase chain reaction (AP-PCR), degenerate oligonucleotide-primed PCR (DOP-PCR) and nucleic

IDT Exhibit 1020.3657

acid based sequence amplification (NABSA), single primer isothermal amplification (SPIA, see e.g. U.S. Pat. No. 6,251,639), Ribo-SPIA, or a combination thereof. Other amplification methods that can be used herein include those described in U.S. Pat. Nos. 5,242,794; 5,494,810; 4,988,617; and 6,582,938. Amplification of target nucleic acids can occur on a bead. In other embodiments, amplification does not occur on a bead. Amplification can be by isothermal amplification, e.g., isothermal linear amplification. A hot start PCR can be performed wherein the reaction is heated to 95°C. for two minutes prior to addition of the polymerase or the polymerase can be kept inactive until the first heating step in cycle 1. Hot start PCR can be used to minimize nonspecific amplification. Other strategies for and aspects of amplification are described, e.g., in U.S. Patent Application Publication No. 2010/0173394 A1, published Jul. 8, 2010, which is incorporated herein by reference. In some cases, the amplification methods can be performed under limiting conditions such that only a few rounds of amplification (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 etc.), such as for example as is commonly done for cDNA generation. The number of rounds of amplification can be about 1-30, 1-20, 1-15, 1-10, 5-30, 10-30, 15-30, 20-30, 10-30, 15-30, 20-30, or 25-30.

[00136] Techniques for amplification of target and reference sequences are known in the art and include the methods described, e.g., in U.S. Pat. No. 7,048,481. Briefly, the techniques can include methods and compositions that separate samples into small droplets, in some instances with each containing on average less than about 5, 4, 3, 2, or one target nucleic acid molecule (polynucleotide) per droplet, amplifying the nucleic acid sequence in each droplet and detecting the presence of a target nucleic acid sequence. In some cases, the sequence that is amplified is present on a probe to the genomic DNA, rather than the genomic DNA itself. In some cases, at least 200, 175, 150, 125, 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, or 0 droplets have zero copies of a target nucleic acid.

[00137] PCR can involve in vitro amplification based on repeated cycles of denaturation, oligonucleotide primer annealing, and primer extension by thermophilic template dependent polynucleotide polymerase, which can result in the exponential increase in copies of the desired sequence of the polynucleotide analyte flanked by the primers. In some cases, two different PCR primers, which anneal to opposite strands of the DNA, can be positioned so that the polymerase catalyzed extension product of one primer can serve as a template strand for the other, leading to the accumulation of a discrete double stranded fragment whose length is defined by the distance between the 5' ends of the oligonucleotide primers.

[00138] LCR can involve use of a ligase enzyme to join pairs of preformed nucleic acid probes. The probes can hybridize with each complementary strand of the nucleic acid analyte, if present, and ligase can be employed to bind each pair of probes together resulting in two templates that can serve in the next cycle to reiterate the particular nucleic acid sequence.

[00139] SDA (Westin et al 2000, Nature Biotechnology, 18, 199-202; Walker et al 1992, Nucleic Acids Research, 20, 7, 1691-1696), can involve isothermal amplification based upon the ability of a restriction endonuclease such as HincII or BsoBI to nick the unmodified strand of a hemiphosphorothioate form of

IDT Exhibit 1020.3658

its recognition site, and the ability of an exonuclease deficient DNA polymerase such as Klenow exo minus polymerase, or Bst polymerase, to extend the 3'-end at the nick and displace the downstream DNA strand. Exponential amplification results from coupling sense and antisense reactions in which strands displaced from a sense reaction serve as targets for an antisense reaction and vice versa.

[00140] Some aspects of the methods described herein can utilize linear amplification of nucleic acids or polynucleotides. Linear amplification can refer to a method that involves the formation of one or more copies of the complement of only one strand of a nucleic acid or polynucleotide molecule, usually a nucleic acid or polynucleotide analyte. Thus, the primary difference between linear amplification and exponential amplification is that in the latter process, the product serves as substrate for the formation of more product, whereas in the former process the starting sequence is the substrate for the formation of product but the product of the reaction, i.e. the replication of the starting template, is not a substrate for generation of products. In linear amplification the amount of product formed increases as a linear function of time as opposed to exponential amplification where the amount of product formed is an exponential function of time.

[00141] In some cases, the amplification is exponential, e.g. in the enzymatic amplification of specific double stranded sequences of DNA by a polymerase chain reaction (PCR). In other embodiments the amplification method is linear. In other embodiments the amplification method is isothermal.

XII. Applications

[00142] One aspect of the methods and compositions disclosed herein is that they can be efficiently and cost-effectively utilized for downstream analyses, such as next generation sequencing or hybridization platforms, with minimal loss of biological material of interest. The methods described herein can be particularly useful for generating high throughput sequencing libraries from template DNA or RNA, for whole genome or whole transcriptome analysis, respectively.

[00143] For example, the methods described herein can be useful for sequencing by the method commercialized by Illumina, as described U.S. Pat. Nos. 5,750,341; 6,306,597; and 5,969,119. Directional (strand-specific) nucleic acid libraries can be prepared using the methods described herein, and the selected single-stranded nucleic acid is amplified, for example, by PCR. The resulting nucleic acid is then denatured and the single-stranded amplified polynucleotides can be randomly attached to the inside surface of flow-cell channels. Unlabeled nucleotides can be added to initiate solid-phase bridge amplification to produce dense clusters of double-stranded DNA. To initiate the first base sequencing cycle, four labeled reversible terminators, primers, and DNA polymerase can be added. After laser excitation, fluorescence from each cluster on the flow cell is imaged. The identity of the first base for each cluster is then recorded. Cycles of sequencing can be performed to determine the fragment sequence one base at a time.

[00144] In some cases, the methods described herein can be useful for preparing target polynucleotides for sequencing by the sequencing by ligation methods commercialized by Applied Biosystems (e.g.,

IDT Exhibit 1020.3659

SOLiD sequencing). Directional (strand-specific) nucleic acid libraries can be prepared using the methods described herein, and the selected single-stranded nucleic acid can then be incorporated into a water in oil emulsion along with polystyrene beads and amplified by for example PCR. In some cases, alternative amplification methods can be employed in the water-in-oil emulsion such as any of the methods provided herein. The amplified product in each water microdroplet formed by the emulsion interact, bind, or hybridize with the one or more beads present in that microdroplet leading to beads with a plurality of amplified products of substantially one sequence. When the emulsion is broken, the beads float to the top of the sample and are placed onto an array. The methods can include a step of rendering the nucleic acid bound to the beads stranded or partially single stranded. Sequencing primers are then added along with a mixture of four different fluorescently labeled oligonucleotide probes. The probes bind specifically to the two bases in the polynucleotide to be sequenced immediately adjacent and 3' of the sequencing primer to determine which of the four bases are at those positions. After washing and reading the fluorescence signal form the first incorporated probe, a ligase is added. The ligase cleaves the oligonucleotide probe between the fifth and sixth bases, removing the fluorescent dye from the polynucleotide to be sequenced. The whole process is repeated using a different sequence primer, until all of the intervening positions in the sequence are imaged. The process allows the simultaneous reading of millions of DNA fragments in a 'massively parallel' manner. This 'sequence-by-ligation' technique uses probes that encode for two bases rather than just one allowing error recognition by signal mismatching, leading to increased base determination accuracy.

[00145] In other embodiments, the methods are useful for preparing target polynucleotides for sequencing by synthesis using the methods commercialized by 454/Roche Life Sciences, including but not limited to the methods and apparatus described in Margulies et al., *Nature* (2005) 437:376-380 (2005); and U.S. Pat. Nos. 7,244,559; 7,335,762; 7,211,390; 7,244,567; 7,264,929; and 7,323,305. Directional (strand-specific) nucleic acid libraries can be prepared using the methods described herein, and the selected single-stranded nucleic acid can be amplified, for example, by PCR. The amplified products can then be immobilized onto beads, and compartmentalized in a water-in-oil emulsion suitable for amplification by PCR. In some cases, alternative amplification methods other than PCR can be employed in the water-in-oil emulsion such as any of the methods provided herein. When the emulsion is broken, amplified fragments can remain bound to the beads. The methods can include a step of rendering the nucleic acid bound to the beads single stranded or partially single stranded. The beads can be enriched and loaded into wells of a fiber optic slide so that there is approximately 1 bead in each well. Nucleotides can be flowed across and into the wells in a fixed order in the presence of polymerase, sulfhydrolase, and luciferase. Addition of nucleotides complementary to the target strand can result in a chemiluminescent signal that can be recorded such as by a camera. The combination of signal intensity and positional information generated across the plate can allow software to determine the DNA sequence.

[00146] In other embodiments, the methods are useful for preparing target polynucleotide(s) for sequencing by the methods commercialized by Helicos BioSciences Corporation (Cambridge, Mass.) as

IDT Exhibit 1020.3660

described in U.S. application Ser. No. 11/167,046, and U.S. Pat. Nos. 7,501,245; 7,491,498; 7,276,720; and in U.S. Patent Application Publication Nos. US20090061439; US20080087826; US20060286566; US20060024711; US20060024678; US20080213770; and US20080103058. Directional (strand-specific) nucleic acid libraries can be prepared using the methods described herein, and the selected single-stranded nucleic acid is amplified, for example, by PCR. The amplified products can then be immobilized onto a flow-cell surface. The methods can include a step of rendering the nucleic acid bound to the flow-cell surface stranded or partially single stranded. Polymerase and labeled nucleotides can then be flowed over the immobilized DNA. After fluorescently labeled nucleotides are incorporated into the DNA strands by a DNA polymerase, the surface can be illuminated with a laser, and an image can be captured and processed to record single molecule incorporation events to produce sequence data.

[00147] In some cases, the methods described herein can be useful for sequencing by the method commercialized by Pacific Biosciences as described in U.S. Patent Nos. 7462452; 7476504; 7405281; 7170050; 7462468; 7476503; 7315019; 7302146; 7313308; and U.S. Patent Application Publication Nos. US20090029385; US20090068655; US20090024331; and US20080206764. Directional (strand-specific) nucleic acid libraries can be prepared using the methods described herein, and the selected single-stranded nucleic acid is amplified, for example, by PCR. The nucleic acid can then be immobilized in zero mode waveguide arrays. The methods can include a step of rendering the nucleic acid bound to the waveguide arrays single stranded or partially single stranded. Polymerase and labeled nucleotides can be added in a reaction mixture, and nucleotide incorporations can be visualized via fluorescent labels attached to the terminal phosphate groups of the nucleotides. The fluorescent labels can be clipped off as part of the nucleotide incorporation. In some cases, circular templates are utilized to enable multiple reads on a single molecule.

[00148] Another example of a sequencing technique that can be used in the methods described herein is nanopore sequencing (see e.g. Soni G V and Meller A. (2007) *Clin Chem* 53: 1996-2001). A nanopore can be a small hole of the order of 1 nanometer in diameter. Immersion of a nanopore in a conducting fluid and application of a potential across it can result in a slight electrical current due to conduction of ions through the nanopore. The amount of current that flows is sensitive to the size of the nanopore. As a DNA molecule passes through a nanopore, each nucleotide on the DNA molecule obstructs the nanopore to a different degree. Thus, the change in the current passing through the nanopore as the DNA molecule passes through the nanopore can represent a reading of the DNA sequence.

[00149] Another example of a sequencing technique that can be used in the methods described herein is semiconductor sequencing provided by Life Techology's Ion Torrent (e.g., using the Ion Personal Genome Machine (PGM)). Ion Torrent technology can use a semiconductor chip with multiple layers, e.g., a layer with micro-machined wells, an ion-sensitive layer, and an ion sensor layer. Nucleic acids can be introduced into the wells, e.g., a clonal population of single nucleic can be attached to a single bead, and the bead can be introduced into a well. To initiate sequencing of the nucleic acids on the beads, one type of deoxyribonucleotide (e.g., dATP, dCTP, dGTP, or dTTP) can be introduced into the wells. When

IDT Exhibit 1020.3661

one or more nucleotides are incorporated by DNA polymerase, protons (hydrogen ions) can be released in the well, which can be detected by the ion sensor. The semiconductor chip can then be washed and the process can be repeated with a different deoxyribonucleotide. A plurality of nucleic acids can be sequenced in the wells of a semiconductor chip. The semiconductor chip can comprise chemical-sensitive field effect transistor (chemFET) arrays to sequence DNA (for example, as described in U.S. Patent Application Publication No. 20090026082). Incorporation of one or more triphosphates into a new nucleic acid strand at the 3′ end of the sequencing primer can be detected by a change in current by a chemFET. An array can have multiple chemFET sensors.

[00150] Another example of a sequencing technique that can be used in the methods described herein is DNA nanoball sequencing (as performed, e.g., by Complete Genomics; see e.g., Drmanac et al. (2010) Science 327: 78-81). DNA can be isolated, fragmented, and size selected. For example, DNA can be fragmented (e.g., by sonication) to a mean length of about 500 bp. Adapters (Adl) can be attached to the ends of the fragments. The adapters can be used to hybridize to anchors for sequencing reactions. DNA with adapters bound to each end can be PCR amplified. The adapter sequences can be modified so that complementary single strand ends bind to each other forming circular DNA. The DNA can be methylated to protect it from cleavage by a type IIS restriction enzyme used in a subsequent step. An adapter (e.g., the right adapter) can have a restriction recognition site, and the restriction recognition site can remain non-methylated. The non-methylated restriction recognition site in the adapter can be recognized by a restriction enzyme (e.g., AcuI), and the DNA can be cleaved by AcuI 13 bp to the right of the right adapter to form linear double stranded DNA. A second round of right and left adapters (Ad2) can be ligated onto either end of the linear DNA, and all DNA with both adapters bound can be PCR amplified (e.g., by PCR). Ad2 sequences can be modified to allow them to bind each other and form circular DNA. The DNA can be methylated, but a restriction enzyme recognition site can remain non-methylated on the left Adl adapter. A restriction enzyme (e.g., AcuI) can be applied, and the DNA can be cleaved 13 bp to the left of the Adl to form a linear DNA fragment. A third round of right and left adapter (Ad3) can be ligated to the right and left flank of the linear DNA, and the resulting fragment can be PCR amplified. The adapters can be modified so that they can bind to each other and form circular DNA. A type III restriction enzyme (e.g., EcoP15) can be added; EcoP15 can cleave the DNA 26 bp to the left of Ad3 and 26 bp to the right of Ad2. This cleavage can remove a large segment of DNA and linearize the DNA once again. A fourth round of right and left adapters (Ad4) can be ligated to the DNA, the DNA can be amplified (e.g., by PCR), and modified so that they bind each other and form the completed circular DNA template. Rolling circle replication (e.g., using Phi 29 DNA polymerase) can be used to amplify small fragments of DNA. The four adapter sequences can contain palindromic sequences that can hybridize and a single strand can fold onto itself to form a DNA nanoball (DNB(TM)) which can be approximately 200-300 nanometers in diameter on average. A DNA nanoball can be attached (e.g., by adsorption) to a microarray (sequencing flowcell). The flow cell can be a silicon wafer coated with silicon dioxide, titanium and hexamehtyldisilazane (HMDS) and a photoresist material. Sequencing can be performed by

IDT Exhibit 1020.3662

unchained sequencing by ligating fluorescent probes to the DNA. The color of the fluorescence of an interrogated position can be visualized by a high resolution camera. The identity of nucleotide sequences between adapter sequences can be determined.

[00151] In some cases, the sequencing technique can comprise paired-end sequencing in which both the forward and reverse template strand can be sequenced. In some cases, the sequencing technique can comprise mate pair library sequencing. In mate pair library sequencing, DNA can be fragments, and 2-5 kb fragments can be end-repaired (e.g., with biotin labeled dNTPs). The DNA fragments can be circularized, and non-circularized DNA can be removed by digestion. Circular DNA can be fragmented and purified (e.g., using the biotin labels). Purified fragments can be end-repaired and ligated to sequencing adapters.

[00152] In some cases, a sequence read is about, more than about, less than about, or at least about 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 525, 550, 575, 600, 625, 650, 675, 700, 725, 750, 775, 800, 825, 850, 875, 900, 925, 950, 975, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, or 3000 bases. In some cases, a sequence read is about 10 to about 50 bases, about 10 to about 100 bases, about 10 to about 200 bases, about 10 to about 300 bases, about 10

IDT Exhibit 1020.3663

to about 400 bases, about 10 to about 500 bases, about 10 to about 600 bases, about 10 to about 700 bases, about 10 to about 800 bases, about 10 to about 900 bases, about 10 to about 1000 bases, about 10 to about 1500 bases, about 10 to about 2000 bases, about 50 to about 100 bases, about 50 to about 150 bases, about 50 to about 200 bases, about 50 to about 500 bases, about 50 to about 1000 bases, about 100 to about 200 bases, about 100 to about 300 bases, about 100 to about 400 bases, about 100 to about 500 bases, about 100 to about 600 bases, about 100 to about 700 bases, about 100 to about 800 bases, about 100 to about 900 bases, or about 100 to about 1000 bases.

[00153] The number of sequence reads from a sample can be about, more than about, less than about, or at least about 100, 1000, 5,000, 10,000, 20,000, 30,000, 40,000, 50,000, 60,000, 70,000, 80,000, 90,000, 100,000, 200,000, 300,000, 400,000, 500,000, 600,000, 700,000, 800,000, 900,000, 1,000,000, 2,000,000, 3,000,000, 4,000,000, 5,000,000, 6,000,000, 7,000,000, 8,000,000, 9,000,000, or 10,000,000.

[00154] The depth of sequencing of a sample can be about, more than about, less than about, or at least about lx, 2x, 3x, 4x, 5x, 6x, 7x, 8x, 9x, 10x, 11x, 12x, 13x, 14x, 15x, 16x, 17x, 18x, 19x, 20x, 21x, 22x, 23x, 24x, 25x, 26x, 27x, 28x, 29x, 30x, 31x, 32x, 33x, 34x, 35x, 36x, 37x, 38x, 39x, 40x, 41x, 42x, 43x, 44x, 45x, 46x, 47x, 48x, 49x, 50x, 51x, 52x, 53x, 54x, 55x, 56x, 57x, 58x, 59x, 60x, 61x, 62x, 63x, 64x, 65x, 66x, 67x, 68x, 69x, 70x, 71x, 72x, 73x, 74x, 75x, 76x, 77x, 78x, 79x, 80x, 81x, 82x, 83x, 84x, 85x, 86x, 87x, 88x, 89x, 90x, 91x, 92x, 93x, 94x, 95x, 96x, 97x, 98x, 99x, l00x, 110x, 120x, 130x, 140x, 150x, 160x, 170x, 180x, 190x, 200x, 300x, 400x, 500x, 600x, 700x, 800x, 900x, l000x, 1500x, 2000x, 2500x, 3000x, 3500x, 4000x, 4500x, 5000x, 5500x, 6000x, 6500x, 7000x, 7500x, 8000x, 8500x, 9000x, 9500x, or 10,000x. The depth of sequencing of a sample can about 1x to about 5x, about 1x to about 10x, about 1x to about 20x, about 5x to about 10x, about 5x to about 20x, about 5x to about 30x, about 10x to about 20x, about 10x to about 25x, about 10x to about 30x, about 10x to about 40x, about 30x to about 100x, about 100x to about 200x, about 100x to about 500x, about 500x to about 1000x, about 1000x, to about 2000x, about 1000x to about 5000x, or about 5000x to about 10,000x. Depth of sequencing can be the number of times a sequence (e.g., a genome) is sequenced. In some cases, the Lander/Waterman equation is used for computing coverage. The general equation can be: C = LN/G, where C = coverage; G = haploid genome length; L = read length; and N = number of reads.

[00155] In some cases, different barcodes can be added (e.g., by using primers and/or adapters) to polynucleotides generated from template nucleic acids by methods described herein, wherein the template nucleic acids are derived from different samples, and the different samples can be pooled and analyzed in a multiplexed assay. The barcode can allow the determination of the sample from which a template nucleic acid originated. Pooling of the libraries generated from the various samples can be performed at different stages following appending of barcode sequences, dependent on the stage of appending the barcodes

IDT Exhibit 1020.3664

XIII. Compositions and Reaction mixtures

[00156] The present methods further provide one or more compositions or reaction mixtures. In some cases, the reaction mixture comprises: (a) template RNA; (b) a primer comprising a random sequence; (c) a reverse transcriptase; (d) a mixture of unmodified dNTPs and non-canonical dNTP (e.g. dUTP); (e) a first adapter comprising a long strand comprising a 3' overhang and a known sequence A and a short strand; (f) a DNA polymerase; (g) a mixture of unmodified dNTPs; (h) a second adapter comprising a long strand comprising a 3' overhang and a known sequence B and a short strand comprising a block at the 3' end. In some cases, the reaction mixture further comprises (e) amplification primers directed to unique priming sites created at each end of the polynucleotides following ligation of the second adapter and, optionally, extension of the end of the polynucleotide comprising second adapter sequence as described herein. In some cases, the reaction mixture further comprises (f) sequencing primers directed against sequences present in one or more of the adapter sequences appended to the ends of the polynucleotides generated by the methods provided herein. In some embodiments the primers (b) comprise sequences selected for preferential hybridizing to a desired group of templates, such as primers that preferentially hybridized to all transcripts other than the structural RNA (such as rRNA). In some embodiments the first adapter (e) comprises a stem-loop oligonucleotide with a 3' overhang comprising random sequences.

XIV. Kits

[00157] Any of the compositions described herein can be comprised in a kit. In a non-limiting example, the kit, in a suitable container, comprises: an adapter or several adapters, one or more of oligonucleotide primers and reagents for ligation, primer extension and amplification. The kit can also comprise means for purification, such as a bead suspension, and nucleic acid modifying enzymes.

[00158] The containers of the kits will generally include at least one vial, test tube, flask, bottle, syringe or other containers, into which a component can be placed, and , suitably aliquotted. Where there is more than one component in the kit, the kit also will generally contain a second, third or other additional container into which the additional components can be separately placed. However, various combinations of components can be comprised in a container.

[00159] When the components of the kit are provided in one or more liquid solutions, the liquid solution can be an aqueous solution. However, the components of the kit can be provided as dried powder(s). When reagents and/or components are provided as a dry powder, the powder can be reconstituted by the addition of a suitable solvent.

[00160] The present methods provide kits containing one or more compositions described herein and other suitable reagents suitable for carrying out the methods described herein. The methods described herein provide, e.g., diagnostic kits for clinical or criminal laboratories, or nucleic acid amplification, or RNA-seq library preparation kits, or analysis kits for general laboratory use. The present methods thus include kits which include some or all of the reagents to carry out the methods described herein, e.g.,

IDT Exhibit 1020.3665

sample preparation reagents, oligonucleotides, binding molecules, stock solutions, nucleotides, polymerases, enzymes, positive and negative control oligonucleotides and target sequences, test tubes or plates, fragmentation or cleavage reagents, detection reagents, purification matrices, and an instruction manual. In some cases the kit contains first strand complementary DNA primers comprising random sequences at the 3'-end. In some cases the first strand cDNA primers contained in the kits comprise sequences hybridizable to selected group of targets, such as all transcripts other than rRNA. In some cases, the kit contains a modified or non-canonical nucleotide. Suitable modified or non-canonical nucleotides include any nucleotides provided herein including but not limited to dUTP. In some cases, the kit comprises a cleavage agent. In some cases, the cleavage agent is a glycosylase and a chemical agent, or an enzyme. The glycosylase can be UNG. The chemical agent can be a polyamine. The polyamine can be DMED. The enzyme can be an endonuclease. The endonuclease can be endonuclease VIII or APE. In some cases, the kit contains a first adapter/primer comprising a first universal sequence and a 3' overhang, wherein the 3' overhang comprises sequence directed against sequence present at the 3' end of a polynucleotide comprising a 3' end block. In some cases the kit contains one of more oligonucleotide first adapters comprising a 3'-overhang wherein the 3'-overhang comprises random sequence. In some cases the first primer comprises a stem-loop oligonucleotide. In some cases the first adapter further comprises barcode sequence and universal sequence. In some cases, the kit contains a second adapter comprising a second universal sequence. In some cases, the kit contains a first primer directed against a portion of a sequence complementary to the universal sequence present in the first adapter and a second primer comprising sequence directed against the universal sequence present in the second adapter or its complement.

[00161] In some cases, the kit can contain one or more reaction mixture components, or one or more mixtures of reaction mixture components. In some cases, the reaction mixture components or mixtures thereof can be provided as concentrated stocks, such as 1.1x, 1.5 x, 2 x, 2.5 x, 3 x, 4 x, 5 x, 6 x, 7 x, 10 x, 15 x, 20 x, 25 x, 33 x, 50 x, 75 x, 100 x or higher concentrated stock. The reaction mixture components can include any of the compositions provided herein including but not limited to buffers, salts, divalent cations, azeotropes, chaotropes, dNTPs, labeled nucleotides, non-canonical or modified nucleotides, dyes, fluorophores, biotin, enzymes (such as endonucleases, exonucleases, glycosylases), or any combination thereof.

[00162] In some cases, the kit can contain one or more oligonucleotide primers, such as the oligonucleotide primers provided herein. For example, the kit can contain one or more oligonucleotide primers comprising sequence directed the adapter sequences appended to the ends of the polynucleotides generated by the methods provided herein. In some cases the kit can contain tailed primers comprising a 3'-portion hybridizable to the target nucleic acid (e.g. sequence present in a first and/or second adapter sequence) and a 5'-portion which is not hybridizable to the target nucleic acid. In some cases, the kit can contain chimeric primers comprising an RNA portion and a DNA portion. In some cases, the 5' portion of the tailed primers comprises one or more barcode or other identifier sequences. In some cases, the

IDT Exhibit 1020.3666

identifier sequences comprise flow cell sequences, TruSeq primer sequence, and/or second read barcode sequences.

[00163] In some cases, the kit can contain one or more polymerases or mixtures thereof. In some cases, the one or more polymerases or mixtures thereof can comprise strand displacement activity. Suitable polymerases include any of the polymerases provided herein. The kit can further contain one or more polymerase substrates such as for example dNTPs, non-canonical or modified nucleotides, or nucleotide analogs.

[00164] In some cases, the kit can contain one or more means for purification of the nucleic acid products, removing of the fragmented products from the desired products, or combination of the above. Suitable means for the purification of the nucleic acid products include but are not limited to single stranded specific exonucleases, affinity matrices, nucleic acid purification columns, spin columns, ultrafiltration or dialysis reagents, or electrophoresis reagents including but not limited acrylamide or agarose, or any combination thereof.

[00165] In some cases, the kit can contain one or more reagents for producing blunt ends. For example, the kit can contain one or more of single stranded DNA specific exonucleases including but not limited to exonuclease 1 or exonuclease 7; a single stranded DNA specific endonucleases such as mung bean exonuclease or S1 exonuclease, one or more polymerases such as for example T4 DNA polymerase or Klenow polymerase, or any mixture thereof. Alternatively, the kit can contain one or more single stranded DNA specific exonucleases, endonucleases and one or more polymerases, wherein the reagents are not provided as a mixture. Additionally, the reagents for producing blunt ends can comprise dNTPs.

[00166] In some cases, the kit can contain one or more reagents for preparing the double stranded products for ligation to adapter molecules. For example, the kit can contain dATP, dCTP, dGTP, dTTP, or any mixture thereof. In some cases, the kit can contain a polynucleotide kinase, such as for example T4 polynucleotide kinase. Additionally, the kit can contain a polymerase suitable for producing a 3' extension from the blunt ended double stranded DNA fragments. Suitable polymerases can be included, for example, exo-Klenow polymerase.

[00167] In some cases, the kit can contain one or more adapter molecules such as any of the adapter molecules provided herein. Suitable adapter molecules include single or double stranded nucleic acid (DNA or RNA) molecules or derivatives thereof, stem-loop nucleic acid molecules, double stranded molecules comprising one or more single stranded overhangs of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 bases or longer, proteins, peptides, aptamers, organic molecules, small organic molecules, or any adapter molecules known in the art that can be covalently or non-covalently attached, such as for example by ligation, to the double stranded DNA fragments. In some cases, the kit contains adapters, wherein the adapters can be duplex adapters wherein one strand comprises a known or universal sequence, while the other strand comprises a 5' and/or 3' block. The long-strand can also comprise a 5' or 3' block. In a further embodiment, the duplex adapter is a partial duplex adapter. In some cases, the partial duplex adapter comprises a long strand comprising a known or universal sequence, and a short strand comprising a 5'

IDT Exhibit 1020.3667

and 3' block. The long-strand can also comprise a 5' or 3' block. In some cases, the 3' block is blocked with a terminal dideonucleotide.

[00168] In some cases, the kit can contain one or more reagents for performing gap or fill-in repair on the ligation complex formed between the adapter(s) and the double stranded products of the methods described herein. The kit can contain a polymerase suitable for performing gap repair. Suitable polymerases can be included, for example, *Taq* DNA polymerase.

[00169] The kit can further contain instructions for the use of the kit. For example, the kit can contain instructions for generating directional polynucleotide libraries or directional cDNA libraries representing the whole or a part of the transcriptome or genome useful for large scale analysis of including but not limited to e.g., , pyrosequencing, sequencing by synthesis, sequencing by hybridization, single molecule sequencing, nanopore sequencing, and sequencing by ligation, high density PCR, digital PCR, massively parallel Q-PCR, and characterizing amplified nucleic acid products generated by the methods described herein, or any combination thereof. The kit can further contain instructions for mixing the one or more reaction mixture components to generate one or more reaction mixtures suitable for the methods described herein. The kit can further contain instructions for hybridizing the one or more oligonucleotide primers to a nucleic acid template. The kit can further contain instructions for extending the one or more oligonucleotide primers with for example a polymerase and/or modified dNTPs. The kit can further contain instructions for treating the DNA products with a cleavage agent. In some cases, the cleavage agent is a glycosylase and a chemical agent, or an enzyme. The glycosylase can be UNG. The chemical agent can be a polyamine. The polyamine can be DMED. The enzyme can be an endonuclease. The endonuclease can be endonuclease VIII or APE. The kit can further contain instructions for purification of any of the products provided by any of the steps of the methods provided herein. The kit can further contain instructions for producing blunt ended fragments, for example by removing single stranded overhangs or filling in single stranded overhangs, with for example single stranded DNA specific exonucleases, polymerases, or any combination thereof. The kit can further contain instructions for phosphorylating the 5' ends of the double stranded DNA fragments produced by the methods described herein. The kit can further contain instructions for ligating one or more adapter molecules to the double stranded DNA fragments.

[00170] A kit will can include instructions for employing, the kit components as well the use of any other reagent not included in the kit. Instructions can include variations that can be implemented.

[00171] Unless otherwise specified, terms and symbols of genetics, molecular biology, biochemistry and nucleic acid used herein follow those of standard treatises and texts in the field, e.g. Kornberg and Baker, *DNA Replication*, Second Edition (W.H. Freeman, New York, 1992); Lehninger, *Biochemistry*, Second Edition (Worth Publishers, New York, 1975); Strachan and Read, Human *Molecular Genetics*, Second Edition (Wiley-Liss, New York, 1999); Eckstein, editor, *Oligonucleotides and Analogs: A Practical Approach* (Oxford University Press, New York, 1991); Gait, editor, *Oligonucleotide Synthesis: A Practical Approach* (IRL Press, Oxford, 1984); and the like.

IDT Exhibit 1020.3668

EXAMPLES

Example 1: Stranded library preparation from 100ng total RNA input

[00172] The process described in **FIG. 3** was employed for the generation of stranded cDNA sequencing libraries from Universal Human Reference (UHR) total RNA samples (100ng) following a process workflow as in **FIG. 3**.

[00173] a.) Synthesis of first strand cDNA comprising dU: 2 µl of First Strand Primer Mix (NuGEN, 0334-32) and 2 µl of H_2O were added to 2 µl of Universal Human Reference RNA (50ng/µl; Agilent). The mixture was incubated 65°C for 5 min. and cool on ice. The following mixture was added to the above: 2.5 µl of First Strand Buffer Mix (NuGEN, 0334-32), 0.5 µl of First Strand Enzyme Mix (NuGEN, 0334-32), 0.375 µl of 1 mM dUTP and 0.625 µl of H_2O. First strand cDNA synthesis was carried out at 40°C for 30min. followed by incubation at 70°C for 10 min.

[00174] b.) Fragmentation of first strand cDNA: 0.5 µl USER Enzyme (New England BioLabs) was added to the first strand cDNA synthesis reaction mixture above and the reaction mixture was incubated at 37°C for 30 min. followed by incubation at 95°C for 10min.

[00175] c.) RNA Hydrolysis: The RNA input was hydrolyzed by addition of 2 µl 1N NaOH to the cDNA fragmentation reaction mixture above, and incubation of the reaction mixture at 95°C for 15 min., followed by neutralization of the reaction mixture by the addition of 2 µl 1N HCl to the cooled reaction mixture.

[00176] d.)Purification: The fragmented first strand cDNA was purified using ssDNA/RNA Clean & Concentrator (Zymo Research) following the manufacturer instruction and the purified fragmented first strand cDNA was eluted in 10 µl of H_2O.

[00177] e.) Conversion of the all fragments of first strand cDNA to dsDNA with appended first adaptor at one end: 10µl of the purified fragmented and 3'-blocked first strand cDNA was mixed with 1.5µl of 10xNEBuffer2 (New England BioLabs), 1.5µl of 2.5 mM dNTPs, 0.5µl of 10µM First adaptor (33 bp dsDNA with 8-base 3' overhang of random sequences) hybridizable to the blocked 3'-end of the fragmented first strand cDNA and 1µl of H_2O. The mixture was incubated at 65°C for 5 min, and cool on ice. Extension of the hybridized first adaptor along the first strand cDNA fragments was carried out by the addition of 0.5µl Bsu DNA Polymerase, (Large Fragment New England BioLabs) and incubating the reaction mixture at 25°C for 15 min, 37°C for 15 min, followed by 70°C for 10 min.

[00178] f.) Polishing DNA Ends: The above reaction mixture was combined with 0.5µl T4 DNA Polymerase (Enzymatics) and the reaction mixture was incubated at 25°C for 30 min, followed by 70°C for 10 min.

[00179] g.) Ligation of Second Adaptor to the blunt end of the ds cDNA produced as above: The ligation was carried out by the addition of the following to the above reaction mixture: 6µl of 5xQuick Ligation Buffer (New England BioLabs), 2.5µl of 20 µM Second Adaptor, 1.5µl of Quick Ligase (New

IDT Exhibit 1020.3669

England BioLabs), and 5µl of H_2O. The reaction mixture was incubated at 25°C for 30 min, followed by 70°C for 10 min.

[00180] h.) Purification: The ligation products, dsDNA with first adaptor appended at one end, and second adaptor at the other end, were purified using 0.8 volume of Agencourt Ampure XP (Beckman Coulter), and eluted in 25 µl.

[00181] i.) PCR Amplification: The library of stranded cDNA products with appended first and second adaptors prepared as described above, was PCR amplified with primers comprising sequences specific to the first and the second adaptor, and barcodes enabling multiplex sequencing, for 17 cycles using the following PCR program: 70°C 5 min, 17x(94°C 30sec, 60°C 30sec, 72°C 1 min) 72°C 5min.

[00182] j.) Purification: The PCR products, amplified stranded cDNA library, were purified using 1 volume of Agencourt Ampure XP (Beckman Coulter) following the manufacturer instruction.

[00183] A size distribution of one directional sequencing library generated from 100ng UHR total RNA was analyzed using BioAnalyzer (Agilent). The size distribution of the said library is shown in **FIG. 6**.

Example 2: Generation of stranded cDNA library from 1ng total RNA input

[00184] a.) Synthesis of first strand cDNA comprising dU: 2 µl of First Strand Primer Mix (NuGEN, 0334-32) and 2 µl of H_2O were added to 2 µl of Universal Human Reference RNA (0.5ng/µl; Agilent). The mixture was incubated 65°C for 5 min. and cool on ice. The following mixture was added to the above: 2.5 µl of First Strand Buffer Mix (NuGEN, 0334-32), 0.5 µl of First Strand Enzyme Mix (NuGEN, 0334-32), 0.375 µl of 1 mM dUTP and 0.625 µl of H_2O. First strand cDNA synthesis was carried out at 40°C for 30min. followed by incubation at 70°C for 10 min.

[00185] b.) Fragmentation of first strand cDNA: 0.5 µl USER Enzyme (New England BioLabs) was added to the first strand cDNA synthesis reaction mixture above and the reaction mixture was incubated at 37°C for 30 min. followed by incubation at 95°C for 10min.

[00186] c.) RNA Hydrolysis: The RNA input was hydrolyzed by addition of 2 µl 1N NaOH to the cDNA fragmentation reaction mixture above, and incubation of the reaction mixture at 95°C for 15 min., followed by neutralization of the reaction mixture by the addition of 2 µl 1N HCl to the cooled reaction mixture.

[00187] d.) Purification: The fragmented first strand cDNA was purified using ssDNA/RNA Clean & Concentrator (Zymo Research) following the manufacturer instruction and the purified fragmented first strand cDNA was eluted in 10 µl of H_2O.

[00188] e.) Conversion of the all fragments of first strand cDNA to dsDNA with appended first adaptor at one end: 10µl of the purified fragmented and 3'-blocked first strand cDNA was mixed with 1.5µl of 10xNEBuffer2 (New England BioLabs), 1.5µl of 2.5 mM dNTPs, 0.5µl of 10µM First adaptor (33 bp dsDNA with 8-base 3' overhang of random sequences) hybridizable to the blocked 3'-end of the fragmented first strand cDNA and 1µl of H_2O. The mixture was incubated at 65°C for 5 min, and cool on

IDT Exhibit 1020.3670

ice. Extension of the hybridized first adaptor along the first strand cDNA fragments was carried out by the addition of 0.5µl Bsu DNA Polymerase, (Large Fragment New England BioLabs) and incubating the reaction mixture at 25°C for 15 min, 37°C for 15 min, followed by 70°C for 10 min.

[00189] f.) Polishing DNA Ends: The above reaction mixture was combined with 0.5µl T4 DNA Polymerase (Enzymatics) and the reaction mixture was incubated at 25°C for 30 min, followed by 70°C for 10 min.

[00190] g.) Purification: The DNA was purified using 1.5X volume of Agencourt Ampure XP (Beckman Coµlter), and eluted in 18µl of H_2O

[00191] h.) Ligation of Second Adaptor to the blunt end of the ds cDNA produced as above: The ligation was carried out by the addition of the following to the above purified DNA product: 5µl of 5xQuick Ligation Buffer (New England BioLabs), 0.625µl of 20 µM Second Adaptor, and 1.5µl of Quick Ligase (New England BioLabs). The reaction mixture was incubated at 25°C for 30 min, followed by 70°C for 10 min.

[00192] i.) Purification: The ligation products, dsDNA with first adaptor appended at one end and second adaptor at the other end, were purified using 0.8X volume of Agencourt Ampure XP (Beckman Coulter), and eluted in 25 µl of H2O.

[00193] j.) PCR Amplification was carried out in two steps with a purification step between the two steps.

[00194] First step PCR was carried out for 18 cycles using the following PCR program: 70°C 5 min, 18x(94°C 30sec, 60°C 30sec, 72°C 1 min) 72°C 5min.

[00195] PCR products from this step were purified using 0.8X volume of Agencourt Ampure XP (Beckman Coulter).

[00196] The purified PCR products were further amplified for 7 cycles using the following PCR program: 7x(94°C 30sec, 60°C 30sec, 72°C 1 min) 72°C 5min

[00197] This two step PCR was undertaken with the goal of diminishing the potential generation of primer-dimer artifacts.

[00198] k.) Purification: The PCR products, amplified stranded cDNA library, were purified using 1X volume of Agencourt Ampure XP (Beckman Coulter) following the manufacturer instruction.

Example 3: RNA Strand Retention Efficiency and transcriptome sequencing quality.

[00199] Strand retention efficiency using the methods provided herein was validated experimentally by assessing the strand bias of sequence reads that map to the coding exons of human mRNAs, 3'-UTR and 5'-UTR regions as well as rRNA. Directional cDNA libraries generated according to the methods and compositions provided herein were generated from 100ng and 1ng of total UHR RNA, as described in examples 1 and 2. Single end 40 nucleotide reads were generated using the Illumina Genome Analyzer II. The results of the sequencing data as well as strand retention efficiency summarized in **FIG. 9. FIG. 9.** showed greater than 95% strand retention and minimal reads generated from rRNA for libraries

IDT Exhibit 1020.3671

generated from 100 ng (Sample 1, s4_L2DR14; Sample 2 s4_L2DR15) and 1 ng of total UHR RNA (Sample 3, BC14).

[00200] The quality of transcriptome sequencing generated from directional cDNA libraries described in Examples 1 and 2, employing the methods and compositions provided herein, were further demonstrated from the sequencing data. Non biased whole transcriptome sequencing is demonstrated by analysis of 5'- to 3' representation, as shown for libraries generated from 100 ng (Sample 1, s4_L2DR14; Sample 2 s4_L2DR15; **FIG. 7**)and 1 ng of total UHR RNA (Sample 3, BC14; **FIG. 10**). Furthermore, the choice of first strand cDNA primers utilized for the generation of the directional cDNA sequencing libraries described in Examples 1 and 2, leads to generation of libraries with minimal representation of rRNA.

[00201] The methods and compositions provided herein afford highly reproducible gene expression profiling employing directional cDNA sequencing libraries from total RNA samples as shown by the correlation of sequencing data, reads per kilobase of transcript per million (RPKM), for the libraries s4_L2DR14 and s4_L2DR15 generated as described in Example 1, as shown in **FIG. 8.**

Example 4: Stranded library preparation from total RNA isolated from a single cell:

[00202] The process depicted in **FIG. 1** is employed for the generation of stranded cDNA sequencing libraries from total RNA isolated from a single cell following a process workflow as in **FIG. 3**, following isolation of the RNA from a single cell.

[00203] a.) A single cell is lysed in a cell lysis buffer.

[00204] b.) Synthesis of first strand cDNA comprising dU: 2 µl of First Strand Primer Mix (NuGEN, 0334-32) and 2 µl of H_2O is added to the cell lysate. The mixture is incubated 65°C for 5 min. and cooled on ice. The following mixture is added to the above: 2.5 µl of First Strand Buffer Mix (NuGEN, 0334-32), 0.5 µl of First Strand Enzyme Mix (NuGEN, 0334-32), 0.375 µl of 1 mM dUTP and 0.625 µl of H_2O. First strand cDNA synthesis is carried out at 40°C for 30min. followed by incubating at 70°C for 10 min.

[00205] b.) Fragmentation of first strand cDNA: 0.5 µl USER Enzyme (New England BioLabs) is added to the first strand cDNA synthesis reaction mixture above and the reaction mixture is incubated at 37°C for 30 min. followed by incubation at 95°C for 10min.

[00206] c.) RNA Hydrolysis: The RNA input is hydrolyzed by addition of 2 µl 1N NaOH to the cDNA fragmentation reaction mixture above, and incubation of the reaction mixture at 95°C for 15 min., followed by neutralization of the reaction mixture by the addition of 2 µl 1N HCl to the cooled reaction mixture.

[00207] d.)Purification: The fragmented first strand cDNA is purified using ssDNA/RNA Clean & Concentrator (Zymo Research) following the manufacturer instruction and the purified fragmented first strand cDNA is eluted in 10 µl of H_2O.

[00208] e.) Conversion of the all fragments of first strand cDNA to dsDNA with appended first adaptor at one end: 10µl of the purified fragmented and 3'-blocked first strand cDNA is mixed with 1.5µl of

IDT Exhibit 1020.3672

10xNEBuffer2 (New England BioLabs), 1.5µl of 2.5 mM dNTPs, 0.5µl of 10µM First adaptor (33 bp dsDNA with 8-base 3' overhang of random sequences) hybridizable to the blocked 3'-end of the fragmented first strand cDNA and 1µl of H_2O. The mixture is incubated at 65°C for 5 min, and cooled on ice. Extension of the hybridized first adaptor along the first strand cDNA fragments is carried out by the addition of 0.5µl Bsu DNA Polymerase, (Large Fragment New England BioLabs) and incubating the reaction mixture at 25°C for 15 min, 37°C for 15 min, followed by 70°C for 10 min.

[00209] f.) Polishing DNA Ends: The above reaction mixture is combined with 0.5µl T4 DNA Polymerase (Enzymatics) and the reaction mixture is incubated at 25°C for 30 min, followed by 70°C for 10 min.

[00210] g.) Ligation of Second Adaptor to the blunt end of the ds cDNA produced as above: The ligation is carried out by the addition of the following to the above reaction mixture: 6µl of 5xQuick Ligation Buffer (New England BioLabs), 2.5µl of 20 µM Second Adaptor, 1.5µl of Quick Ligase (New England BioLabs), and 5µl of H_2O. The reaction mixture is incubated at 25°C for 30 min, followed by 70°C for 10 min.

[00211] h.) Purification: The ligation products, dsDNA with first adaptor appended at one end, and second adaptor at the other end, is purified using 0.8 volume of Agencourt Ampure XP (Beckman Coulter), and eluted in 25 µl.

[00212] i.) PCR Amplification: The library of stranded cDNA products with appended first and second adaptors prepared as described above, is PCR amplified with primers comprising sequences specific to the first and the second adaptor, and barcodes enabling multiplex sequencing, for 17 cycles using the following PCR program: 70°C 5 min, 17x(94°C 30sec, 60°C 30sec, 72°C 1 min) 72°C 5min.

[00213] j.) Purification: The PCR products, amplified stranded cDNA library, is purified using 1 volume of Agencourt Ampure XP (Beckman Coulter) following the manufacturer instruction.

[00214] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

IDT Exhibit 1020.3673

CLAIMS

WHAT IS CLAIMED IS:

1. A method for generating a directional cDNA library, the method comprising:
   a) annealing one or more primers to a template RNA;
   b) extending the one or more primers in the presence of a reaction mixture comprising dATP, dCTP, dGTP, dTTP, and dUTP, wherein the reaction mixture comprises a ratio of dUTP to dTTP, wherein the ratio permits incorporation of dUTP at a desired density, thereby generating a one or more first strand complementary DNAs (cDNAs) comprising dUTP incorporated at a desired density;
   c) selectively cleaving the one or more first strand cDNAs comprising dUTPs incorporated at a desired density with uracil-N-glycosylase (UNG) and an agent capable of cleaving a phosphodiester backbone at an abasic site created by the UNG, wherein the cleaving generates a plurality of first strand cDNA fragments of a desired size comprising a blocked 3' end;
   d) annealing a first adapter comprising a partial duplex and a 3' overhang to a 3' end of one or more of the plurality of first strand cDNA fragments comprising a blocked 3' end, wherein the first adapter comprises a sequence A, and wherein the annealing comprises hybridizing a random sequence at the 3' overhang to a complementary sequence present at the 3' end of the one or more of the plurality of first strand cDNA fragments comprising a blocked 3' end;
   e) extending the 3' overhang hybridized to the complementary sequence with a DNA polymerase, wherein one or more double stranded cDNA fragments comprising the sequence A at one end is generated; and
   f) ligating a second adapter comprising a sequence B to the one or more double stranded cDNA fragments comprising the sequence A at one end, wherein the ligating generates one or more double stranded cDNA fragments comprising the sequence A at one end and the sequence B at an opposite end, thereby generating the directional polynucleotide library.
2. A method for whole transcriptome directional sequencing, the method comprising:
   a) annealing one or more primers to a template RNA;
   b) extending the primer in the presence of a reaction mixture comprising dATP, dCTP, dGTP, dTTP, and dUTP, wherein the reaction mixture comprises a ratio of dUTP to dTTP, wherein the ratio permits incorporation of dUTP at a desired density, thereby generating one or more first strand complementary DNAs (cDNAs) comprising dUTP incorporated at a desired density;
   c) selectively cleaving the one or more first strand cDNAs comprising dUTPs incorporated at a desired density with uracil-N-glycosylase (UNG) and an agent

IDT Exhibit 1020.3674

capable of cleaving a phosphodiester backbone at an abasic site created by the UNG, wherein the cleaving generates a plurality of first strand cDNA fragments of a desired size comprising a blocked 3' end;

d) annealing a first adapter comprising a partial duplex and a 3' overhang to a 3' end of one or more of the plurality of first strand cDNA fragments comprising a blocked 3' end, wherein the first adapter comprises a sequence A, and wherein the annealing comprises hybridizing a random sequence at the 3' overhang to a complementary sequence present at the 3' end of the one or more of the plurality of first strand cDNA fragments comprising a blocked 3' end;

e) extending the 3' overhang hybridized to the complementary sequence with a DNA polymerase, wherein one or more double stranded cDNA fragments comprising the sequence A at one end is generated;

f) ligating a second adapter comprising a sequence B to the one or more double stranded cDNA fragments comprising the sequence A at one end, wherein the ligating generates one or more double stranded cDNA fragments comprising the sequence A at one end and the sequence B at an opposite end thereby generating a directional cDNA library; and

g) amplifying and/or sequencing the directional cDNA library.

3. A method for generating a directional cDNA library, the method comprising:

a) treating a template dsDNA with a nicking enzyme, wherein the treating generates one or more breaks in a phosphodiester backbone of one strand of the template dsDNA, wherein the break produces one or more 3' hydroxyls in the one strand;

b) extending the one or more 3' hydroxyls, wherein the extending is performed in the presence of a reaction mixture comprising dATP, dCTP, dGTP, dTTP, and dUTP, wherein the reaction mixture comprises a ratio of dUTP to dTTP, wherein the ratio permits incorporation of dUTP at a desired density, thereby generating one or more first strand complementary DNAs (cDNAs) comprising dUTP incorporated at a desired density;

c) selectively cleaving the one or more first strand cDNAs comprising dUTPs incorporated at a desired density with uracil-N-glycosylase (UNG) and an agent capable of cleaving a phosphodiester backbone at an abasic site created by the UNG, wherein the cleaving generates a plurality of first strand cDNA fragments of a desired size comprising a blocked 3' end;

d) annealing a first adapter comprising a partial duplex and a 3' overhang to a 3' end of one or more of the plurality of first strand cDNA fragments comprising a blocked 3' end, wherein the first adapter comprises a sequence A, and wherein the annealing comprises hybridizing a random sequence at the 3' overhang to a complementary

IDT Exhibit 1020.3675

sequence present at the 3' end of the one or more of the plurality of first strand cDNA fragments comprising a blocked 3' end;

e) extending the 3' overhang hybridized to the complementary sequence with a DNA polymerase, wherein one or more double stranded cDNA fragments comprising the sequence A at one end is generated; and

f) ligating a second adapter comprising a sequence B to the one or more double stranded cDNA fragments comprising the sequence A at one end, wherein the ligating generates one or more double stranded cDNA fragments comprising the sequence A at one end and the sequence B at an opposite end thereby generating a directional cDNA library.

4. A method for whole genome sequencing, the method comprising:

a) treating genomic DNA with a nicking enzyme, wherein the treating generates one or more breaks in a phosphodiester backbone of a one strand of the genomic DNA, wherein the breaks produce one or more 3' hydroxyls in the one strand;

b) extending the one or more 3' hydroxyls, wherein the extending is performed in the presence of a reaction mixture comprising dATP, dCTP, dGTP, dTTP, and dUTP, wherein the reaction mixture comprises a ratio of dUTP to dTTP, wherein the ratio permits incorporation of dUTP at a desired density, thereby generating one or more first strand complementary DNAs (cDNAs) comprising dUTP incorporated at a defined frequency;

c) selectively cleaving the one or more first strand cDNA comprising dUTPs incorporated at a desired density with uracil-N-glycosylase (UNG) and an agent capable of cleaving a phosphodiester backbone at an abasic site created by the UNG, wherein the cleaving generates a plurality of first strand cDNA fragments of a desired size comprising a blocked 3' end;

d) annealing a first adapter comprising a partial duplex and a 3' overhang to a 3' end of one or more of the plurality of first strand cDNA fragments comprising a blocked 3' end, wherein the first adapter comprises a sequence A, and wherein the annealing comprises hybridizing a random sequence at the 3' overhang to a complementary sequence present at the 3' end of the one or more of the plurality of first strand cDNA fragments comprising a blocked 3' end;

e) extending the 3' overhang hybridized to the complementary sequence with a DNA polymerase, wherein one or more double stranded cDNA fragments comprising the sequence A at one end is generated;

f) ligating a second adapter comprising a sequence B to the one or more double stranded cDNA fragments comprising the sequence A at one end, wherein the ligating generates one or more double stranded cDNA fragments comprising the

IDT Exhibit 1020.3676

sequence A at one end and the sequence B at an opposite end thereby generating a directional cDNA library; and

g) amplifying and/or sequencing the directional cDNA library.

5. The method of claim 1 or 2, wherein the one or more primers comprise a random primer.

6. The method of claim 1 or 2, wherein the one or more primers comprise a sequence specific to a target template RNA or group of RNAs.

7. The method of claim 6, wherein the group of RNAs comprises substantially all transcripts.

8. The method of claim 6, wherein the group of RNAs does not comprise structural RNA, wherein the structural RNA comprises ribosomal RNA (rRNA).

9. The method of claim 1, 2, 3, or 4, wherein the agent capable of cleaving a phosphodiester backbone comprises an enzyme, chemical agent, and/or heat.

10. The method of claim 9, wherein the chemical agent is a polyamine.

11. The method of claim 10, wherein the polyamine is N,N-dimethylethylenediamine (DMED).

12. The method of claim 9, wherein the enzyme is an endonuclease.

13. The method of claim 12, wherein the endonuclease is endonuclease VIII.

14. The method of claim 1, 2, 3, or 4, wherein the partial duplex comprises a long strand and a short strand, wherein the long strand comprises the sequence A that forms a duplex with the short strand and a 3' overhang.

15. The method of claim 14, wherein the short strand further comprises a block at a 3' and/or a 5' end.

16. The method of claim 1, 2, 3, or 4, wherein the first adapter comprises a plurality of first adapters, wherein the random sequence on each of the plurality of first adapters is different than the random sequence on another of the plurality of first adapters, and wherein each of the plurality of first adapters comprises the sequence A.

17. The method of claim 16, wherein step d) results in substantially all of the plurality of first strand cDNA fragments of a desired size comprising a blocked 3' end generated in step c) further comprising one of the plurality of first adapters annealed the 3' end.

18. The method of claim 1, 2, 3, or 4, wherein the first adapter further comprises a stem loop, wherein the stem loop links a 5' end of a long strand of the partial duplex with a 3' end of a short strand of the partial duplex, and wherein the long strand comprises the sequence A and the 3' overhang.

19. The method of claim 14, wherein the first adapter further comprises a block at a 5' end of the long strand.

20. The method of claim 16, wherein the first adapter further comprises a block at a 5' end of the short strand.

IDT Exhibit 1020.3677

21. The method of claim 1, 2, 3, or 4, wherein the 3' overhang comprises at least 6, 7, 8, or 9 nucleotides.

22. The method of claim 1, 2, 3, or 4, wherein the second adapter comprises a partial duplex, wherein the partial duplex comprises a long strand hybridized to a short strand, wherein the long strand comprises the sequence B and an overhang.

23. The method of claim 22, wherein the long strand comprises the sequence B and a 3' overhang, and wherein the short strand comprises a block at a 3' end.

24. The method of claim 23, wherein the ligating generates the one or more double stranded cDNA fragments comprising the sequence A at one end and the sequence B at an opposite end, wherein the sequence A is at a 5' end on one end and the sequence B is at a 3' end on the opposite end.

25. The method of claim 22, wherein the long strand comprises the sequence B and a 5' overhang, and wherein the short strand comprises a block at a 5' end.

26. The method of claim 25, wherein the ligating generates the one or more double stranded cDNA fragments comprising the sequence A at one end and the sequence B at an opposite end, wherein the sequence A is at a 5' end on one end and the sequence B is at a 5' end on the opposite end.

27. The method of claim 26, wherein a 3' end of the opposite end is extended using the sequence B as a template, thereby generating one or more double stranded cDNA fragments comprising the sequence A at a 5' end on one end and a sequence complementary to the sequence B, B', at a 3' end on the opposite end.

28. The method of claim 1 or 3, further comprising amplifying the directional cDNA library, thereby generating amplified products.

29. The method of claim 28, further comprising an additional step of sequencing the amplified products.

30. The method of claim 28, wherein the amplification comprises SPIA.

31. The method of claim 28, wherein the amplification comprises a use of primers, wherein one or more of the primers comprises one or more barcode sequences.

32. The method of claim 29, wherein the sequencing comprises next generation sequencing.

33. The method of claim 2 or 4, wherein the amplification comprises SPIA.

34. The method of claim 2 or 4, wherein the amplification comprises a use of primers, wherein one or more of the primers comprises a barcode sequence.

35. The method of claim 2 or 4, wherein the sequencing comprises next generation sequencing.

36. The method of claim 1 or 2, further comprising degrading the template RNA following step b.).

37. The method of claim 1 or 2, wherein the cleaving comprises exposing the template RNA sample to an RNase.

IDT Exhibit 1020.3678

38. The method of claim 3 or 4, wherein the nicking enzyme comprises a strand specific nicking enzyme.

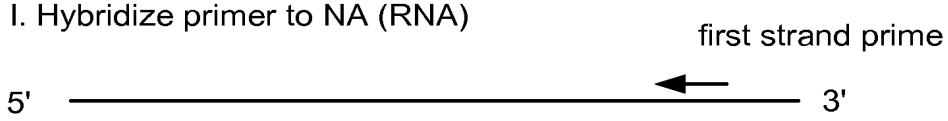
39. The method of claim 3 or 4, wherein the extending the one or more 3' hydroxyls in step b) is performed with a DNA polymerase comprising strand displacement activity.

40. The method of claim 1, 2, 3, or 4, wherein the ligating comprises blunt end ligation, wherein the one or more double stranded cDNA fragments comprising the sequence A at one end generated in step e) are end repaired prior to step f).

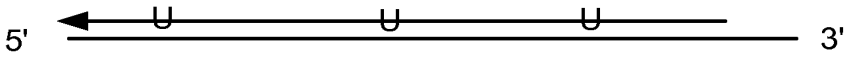
41. The method of claim 1, 2, 3, or 4, wherein the first and/or second adapter further comprises one or more barcodes.

IDT Exhibit 1020.3679

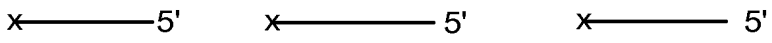
I. Hybridize primer to NA (RNA)



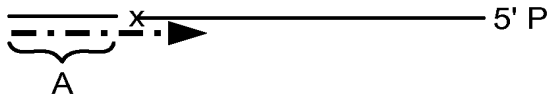
II. Primer extension incorporating dU



III. Degrade template NA and treat primer extension products with UNG and cleave to generate fragments of desired size range with blocked 3'-ends



IV. Anneal first adaptor comprising a 3'-overhang comprising random sequence at the 3'-end and sequence A (one exemplary fragment shown)

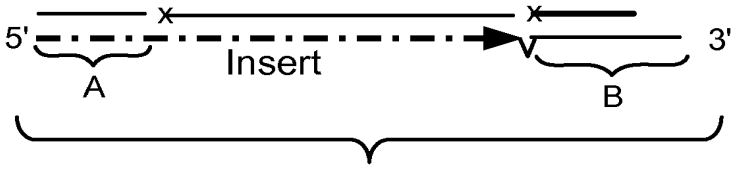


V. Extend hybridized 3'-end of the of adaptor along the fragment comprising a blocked 3'-end generated in step III with a DNA polymerase



Double stranded fragments with an appended sequence A

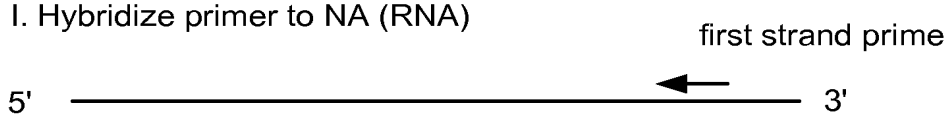
VI. Ligate a second adaptor comprising sequence B



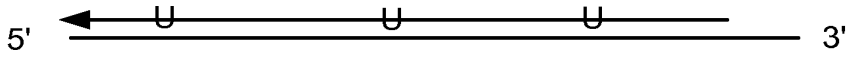
Strand specific product with defined sequences, A and B, at the 5' and 3'-ends of the insert(s) respectively

FIG. 1A

I. Hybridize primer to NA (RNA)



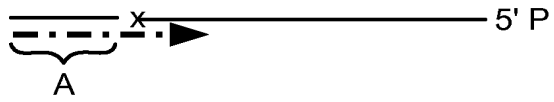
II. Primer extension incorporating dU



III. Degrade template NA and treat primer extension products with UNG and cleave to generate fragments of desired size range with blocked 3'-ends



IV. Anneal first adaptor comprising a 3'-overhang comprising random sequence at the 3'-end and sequence A (one exemplary fragment shown)

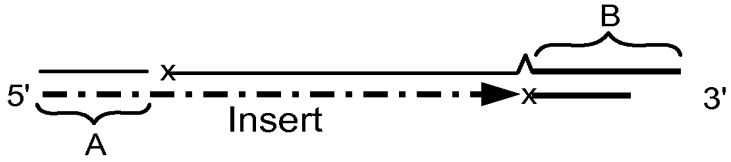


V. Extend hybridized 3'-end of the of adaptor along the fragment comprising a blocked 3'-end generated in step III with a DNA polymerase

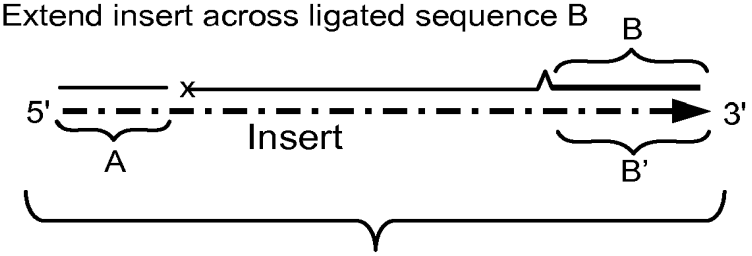


Double stranded fragments with an appended sequence A

VI. Ligate a second adaptor comprising sequence B



VII. Extend insert across ligated sequence B



Strand specific product with defined sequences, A and B, at the 5' and 3'-ends of the insert(s) respectively

FIG. 1B

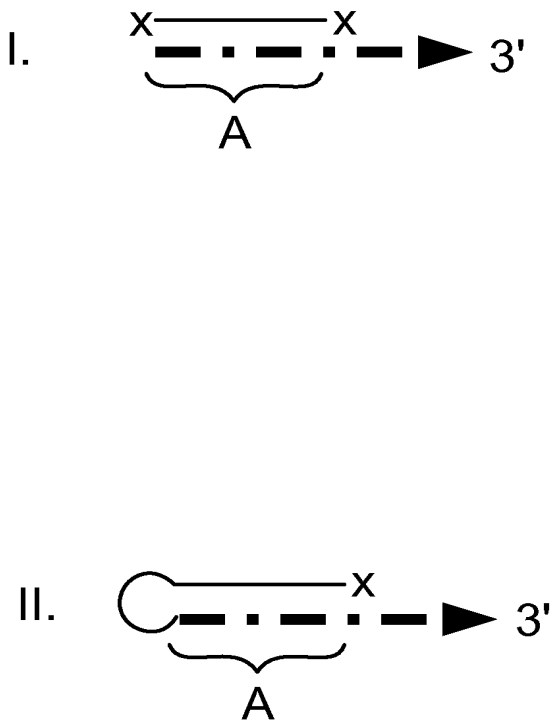


FIG. 2

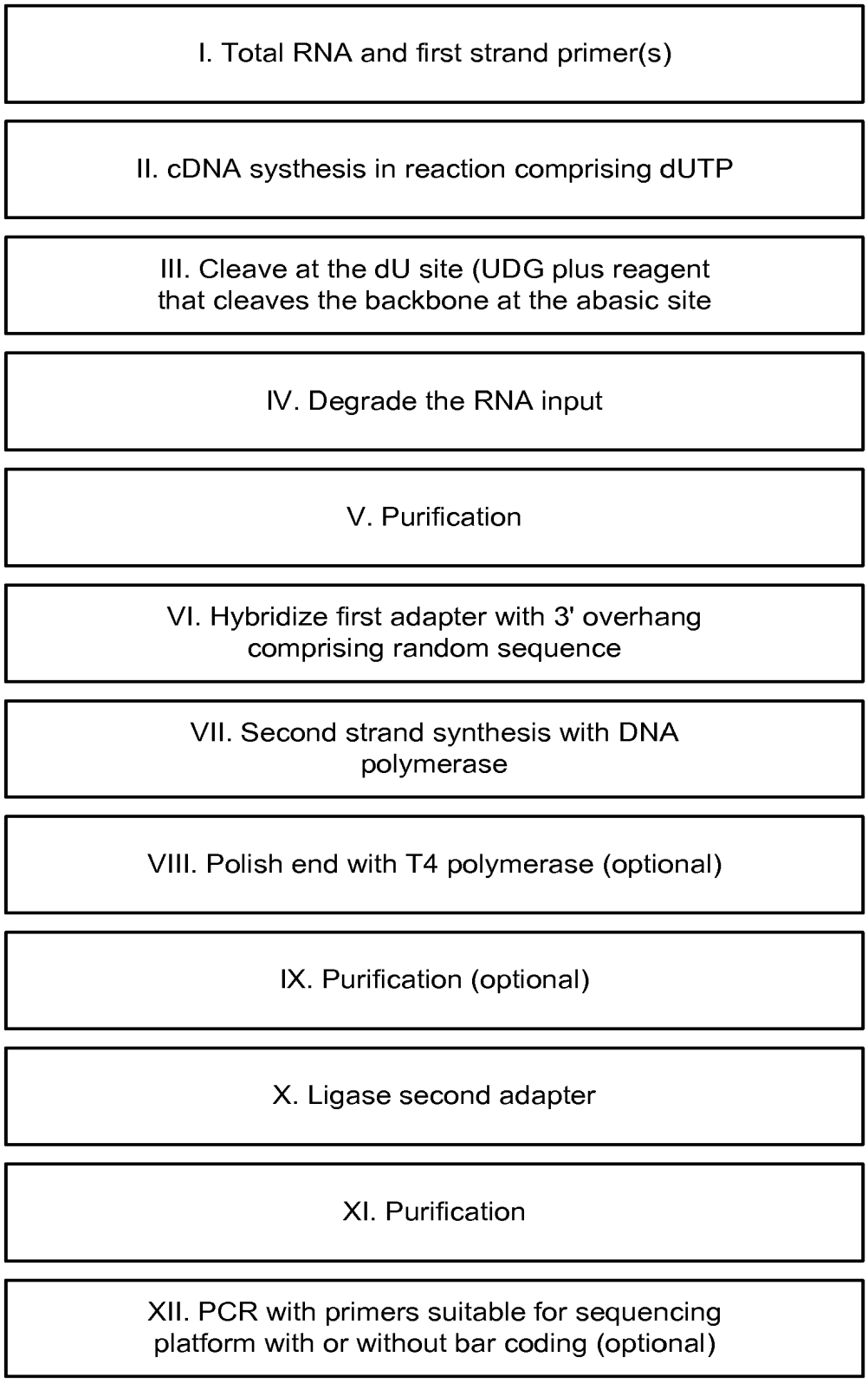

I. Total RNA and first strand primer(s)
II. cDNA systhesis in reaction comprising dUTP
III. Cleave at the dU site (UDG plus reagent that cleaves the backbone at the abasic site
IV. Degrade the RNA input
V. Purification
VI. Hybridize first adapter with 3' overhang comprising random sequence
VII. Second strand synthesis with DNA polymerase
VIII. Polish end with T4 polymerase (optional)
IX. Purification (optional)
X. Ligase second adapter
XI. Purification
XII. PCR with primers suitable for sequencing platform with or without bar coding (optional)

FIG. 3

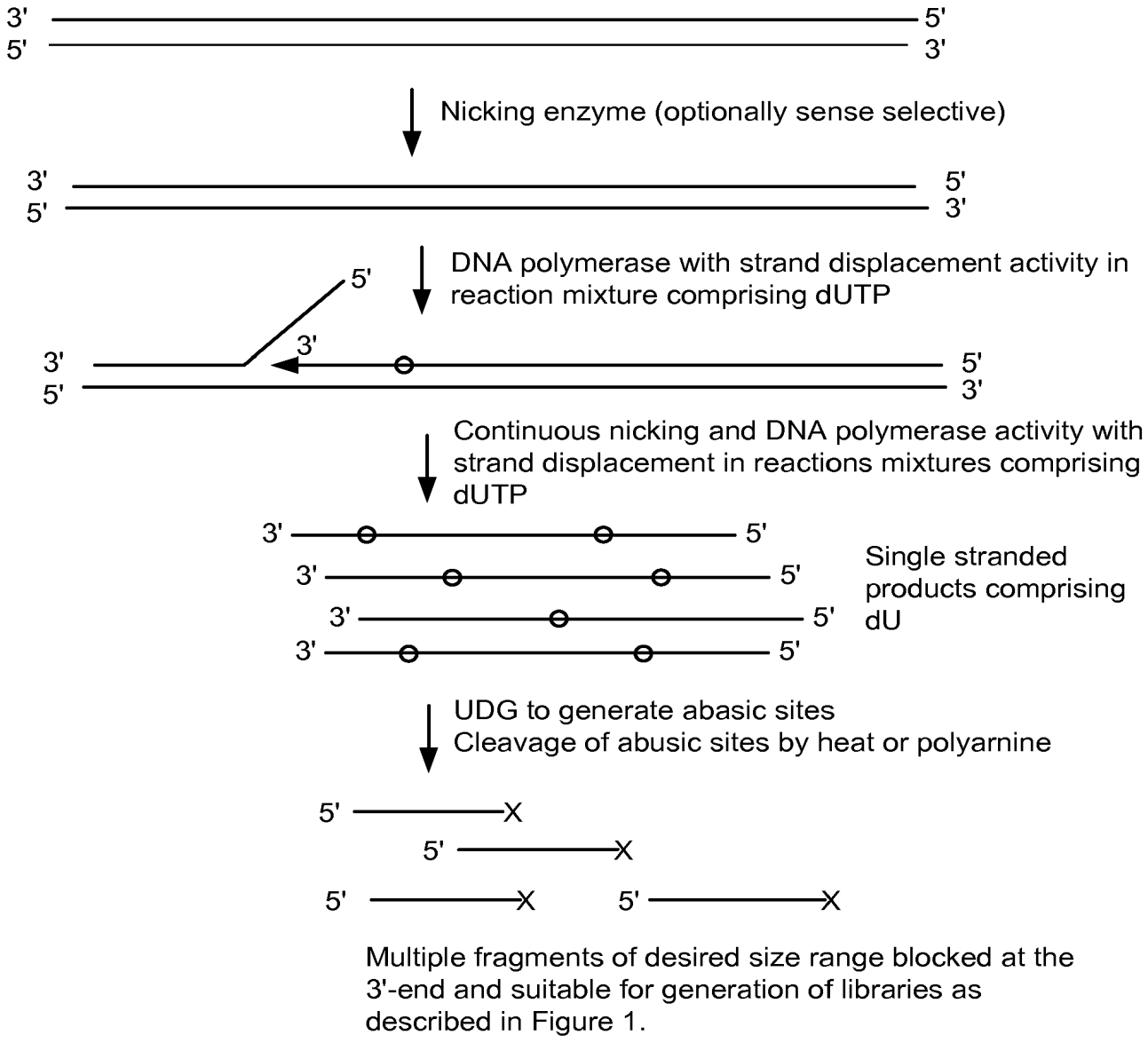

3'
5'
5'
3'
Nicking enzyme (optionally sense selective)
3'
5'
5'
3'
DNA polymerase with strand displacement activity in reaction mixture comprising dUTP
5'
3'
3'
5'
5'
3'
Continuous nicking and DNA polymerase activity with strand displacement in reactions mixtures comprising dUTP
3'
5'
3'
5'
3'
5'
3'
5'
Single stranded products comprising dU
UDG to generate abasic sites
Cleavage of abusic sites by heat or polyarnine
5'
X
5'
X
5'
X
5'
X
Multiple fragments of desired size range blocked at the 3'-end and suitable for generation of libraries as described in Figure 1.

FIG. 4

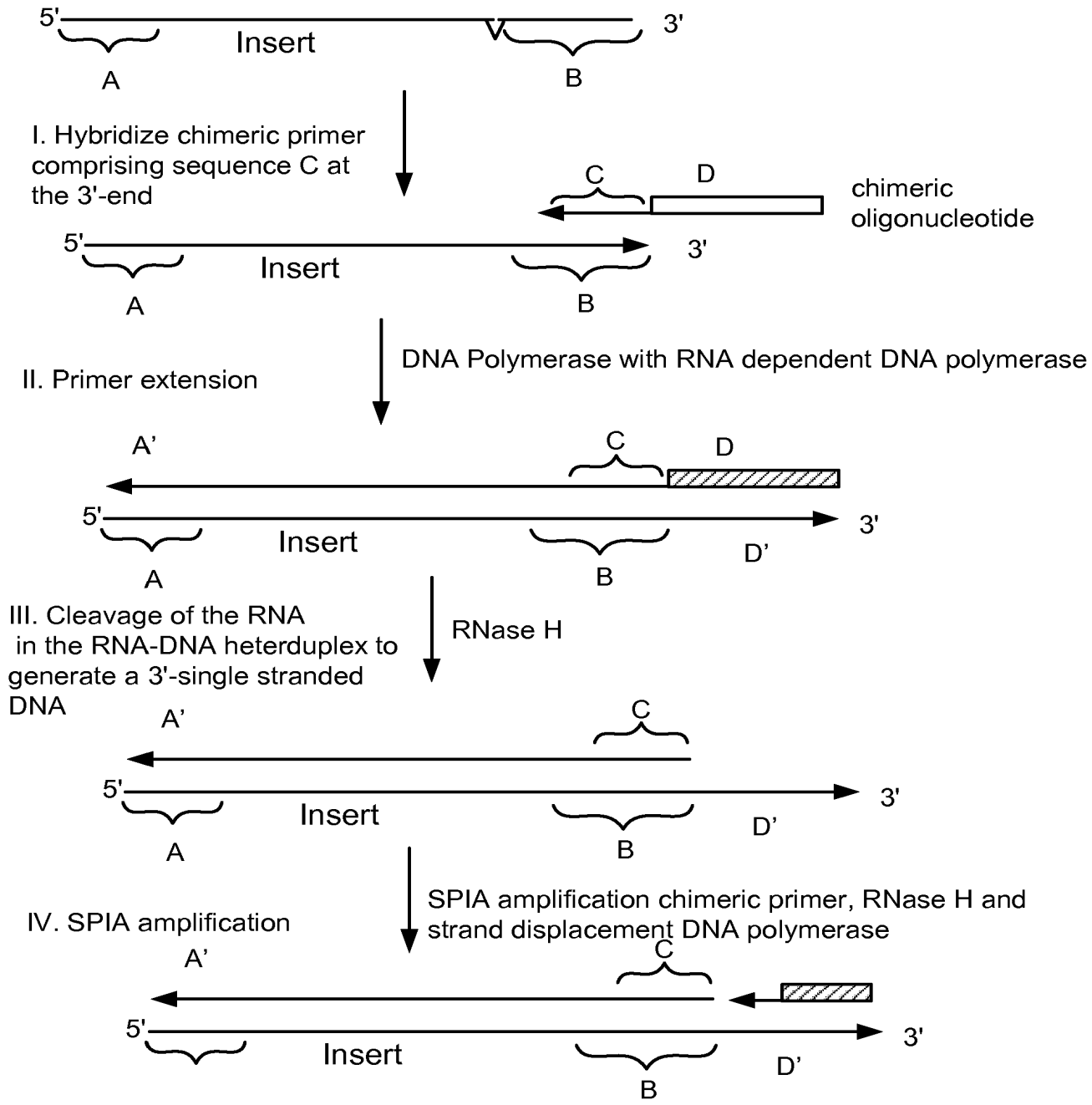

5'
Insert
3'
A
B
I. Hybridize chimeric primer comprising sequence C at the 3'-end
C
D
chimeric oligonucleotide
5'
Insert
3'
A
B
II. Primer extension
DNA Polymerase with RNA dependent DNA polymerase
A'
C
D
5'
Insert
3'
D'
A
B
III. Cleavage of the RNA in the RNA-DNA heterduplex to generate a 3'-single stranded DNA
RNase H
A'
C
5'
Insert
3'
D'
A
B
IV. SPIA amplification
SPIA amplification chimeric primer, RNase H and strand displacement DNA polymerase
A'
C
5'
Insert
3'
D'
B

FIG. 5

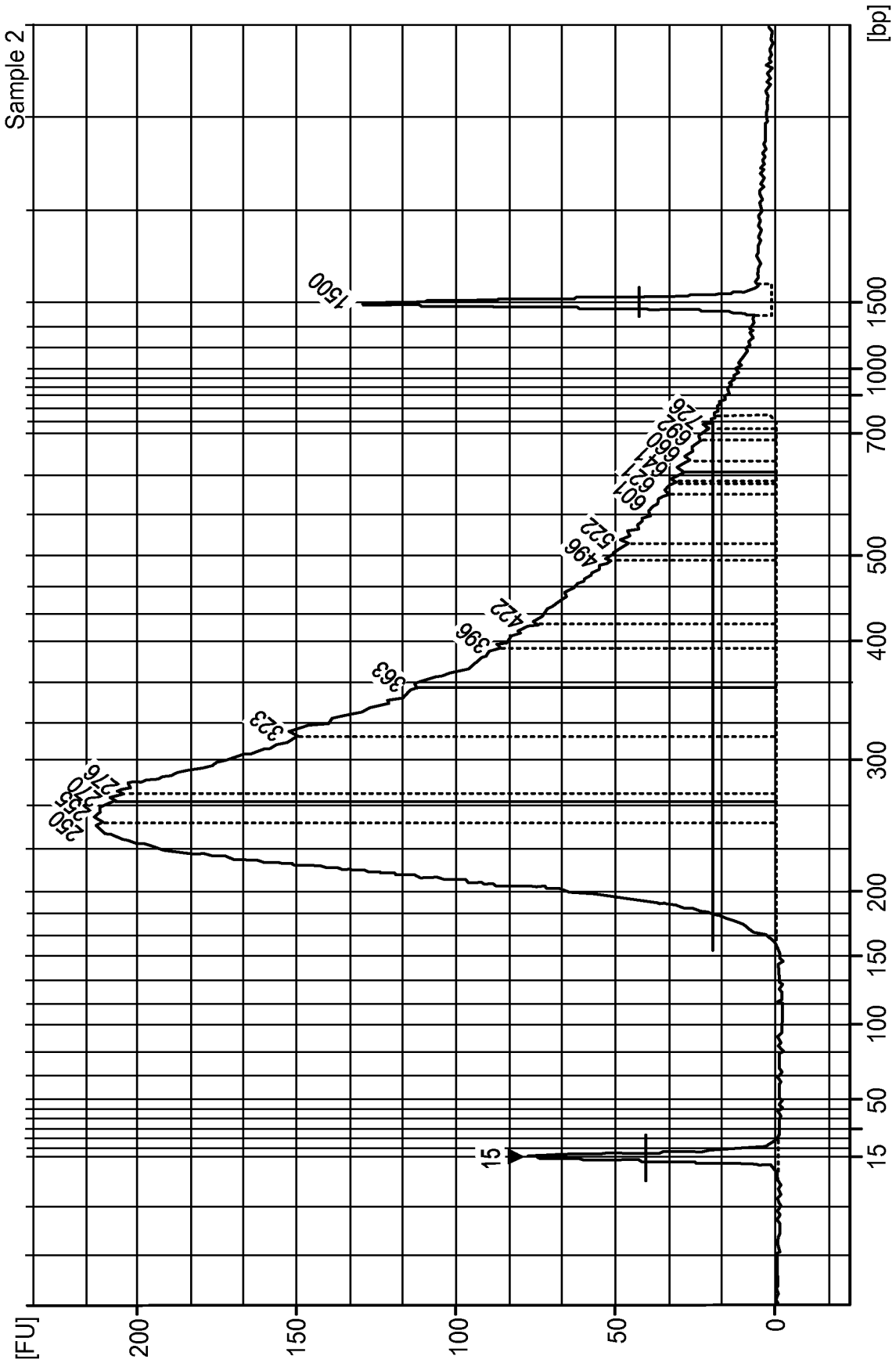

Sample 2
[FU]
200
150
100
50
0
[bp]
15
50
100
150
200
300
400
500
700
1000
1500
15
1500
726
692
660
641
621
601
522
496
422
396
363
323
276
270
255
250

FIG. 6

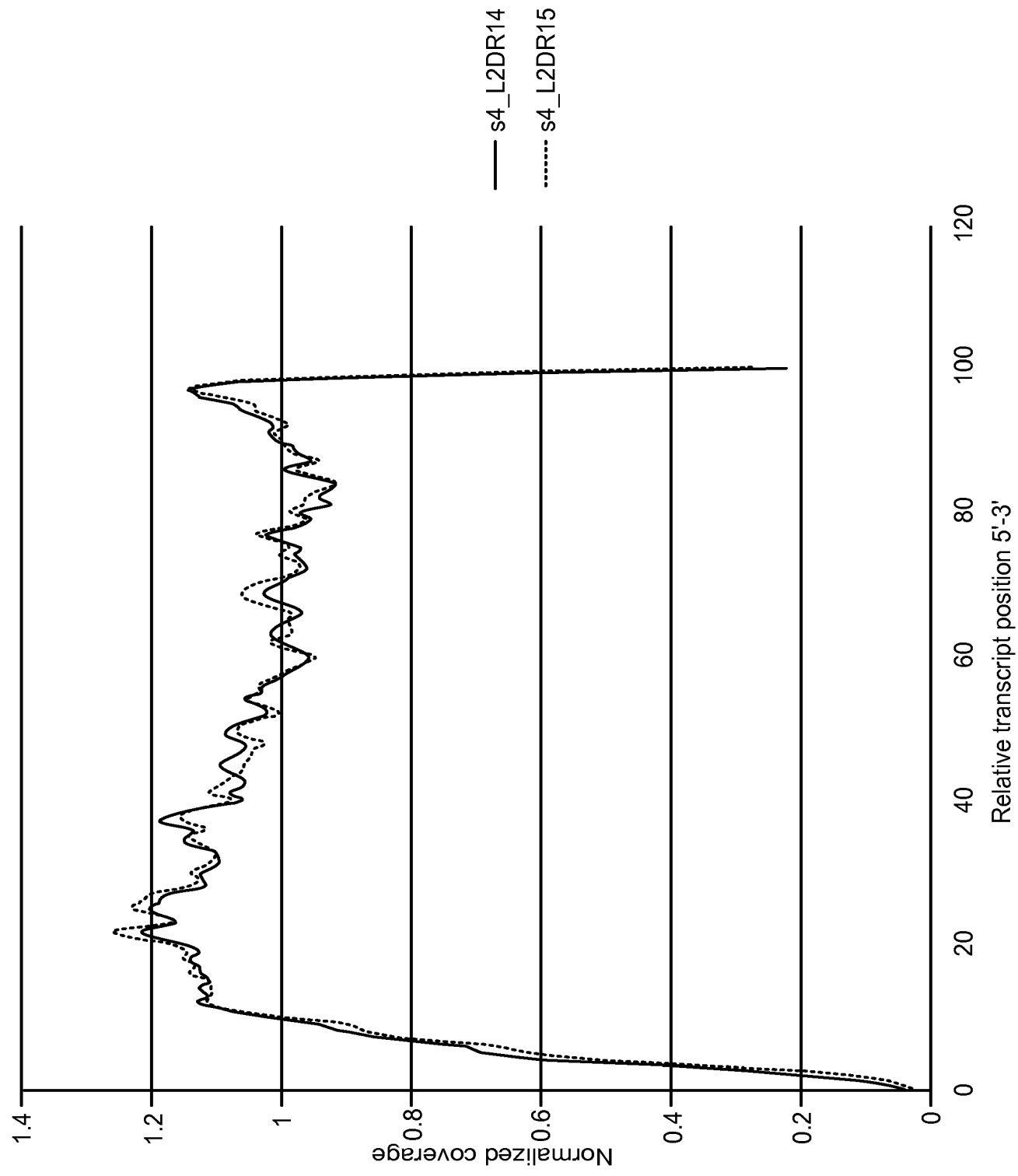

s4_L2DR14
s4_L2DR15
Normalized coverage
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
20
40
60
80
100
120
Relative transcript position 5'-3'

FIG. 7

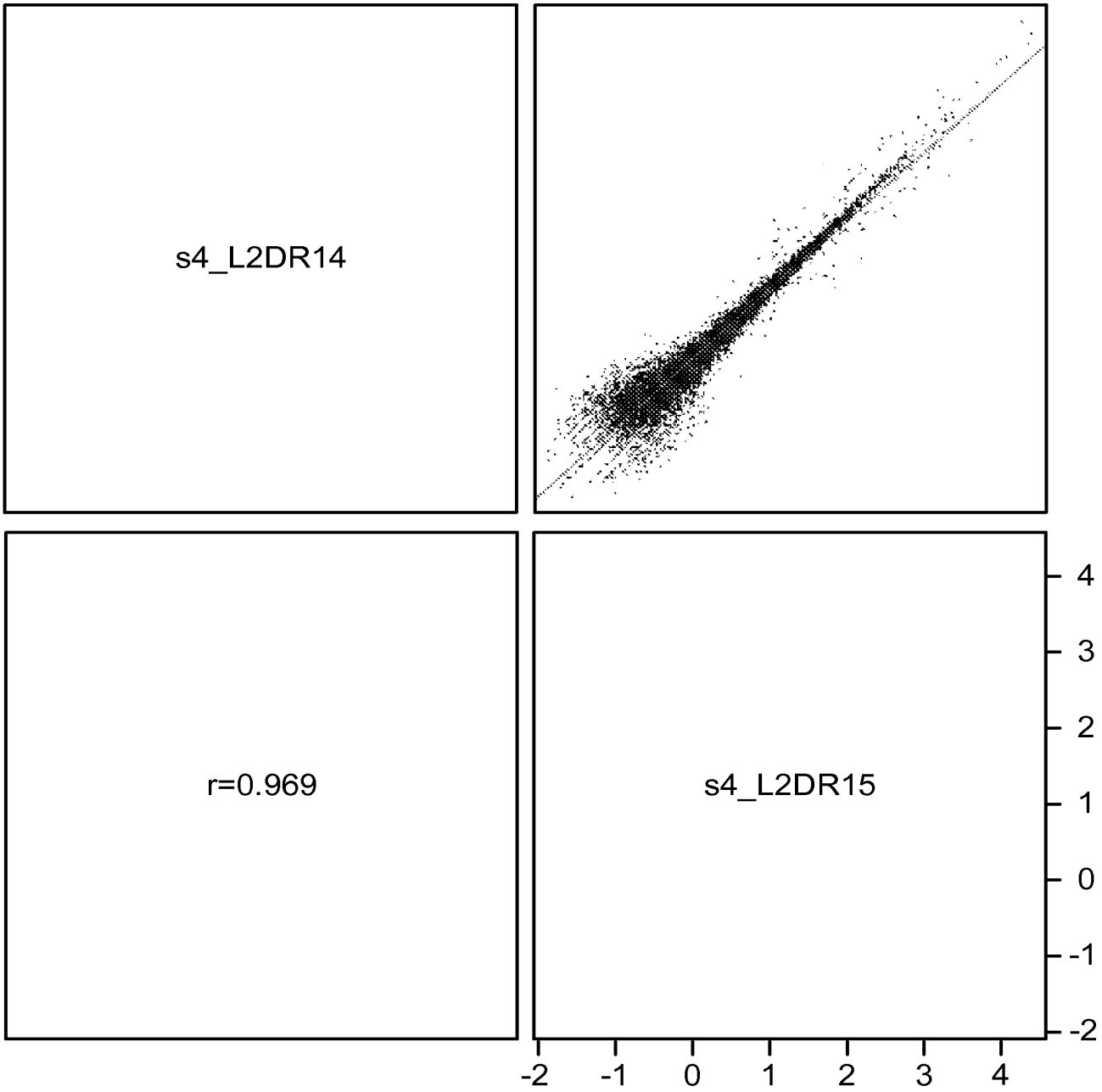


FIG. 8

Sample	Input	% of all		% of mapped				RefSeq breakdown (counts)			RefSeq strand retention			
		Not aligned	aligned	rRNA	non-rRNA	mito rRNA	cyto rRNAA	exon	intron	intergenic	coding	5'UTR	3'UTR	rRNA
1	100 ng	12%	88%	6.6%	93.4%	1.6%	5.0%	27.7%	38.6%	33.7%	97.2%	96.5%	95.1%	99.5%
2	100 ng	7%	93%	7.1%	92.9%	1.5%	5.6%	31.0%	39.0%	30.1%	97.8%	97.1%	95.8%	99.6%
3	1 ng	13%	87%	9.2%	90.8%	0.95%	8.23%	25.5%	38.4%	36.1%	97.5%	96.6%	95.7%	99.6%

FIG. 9

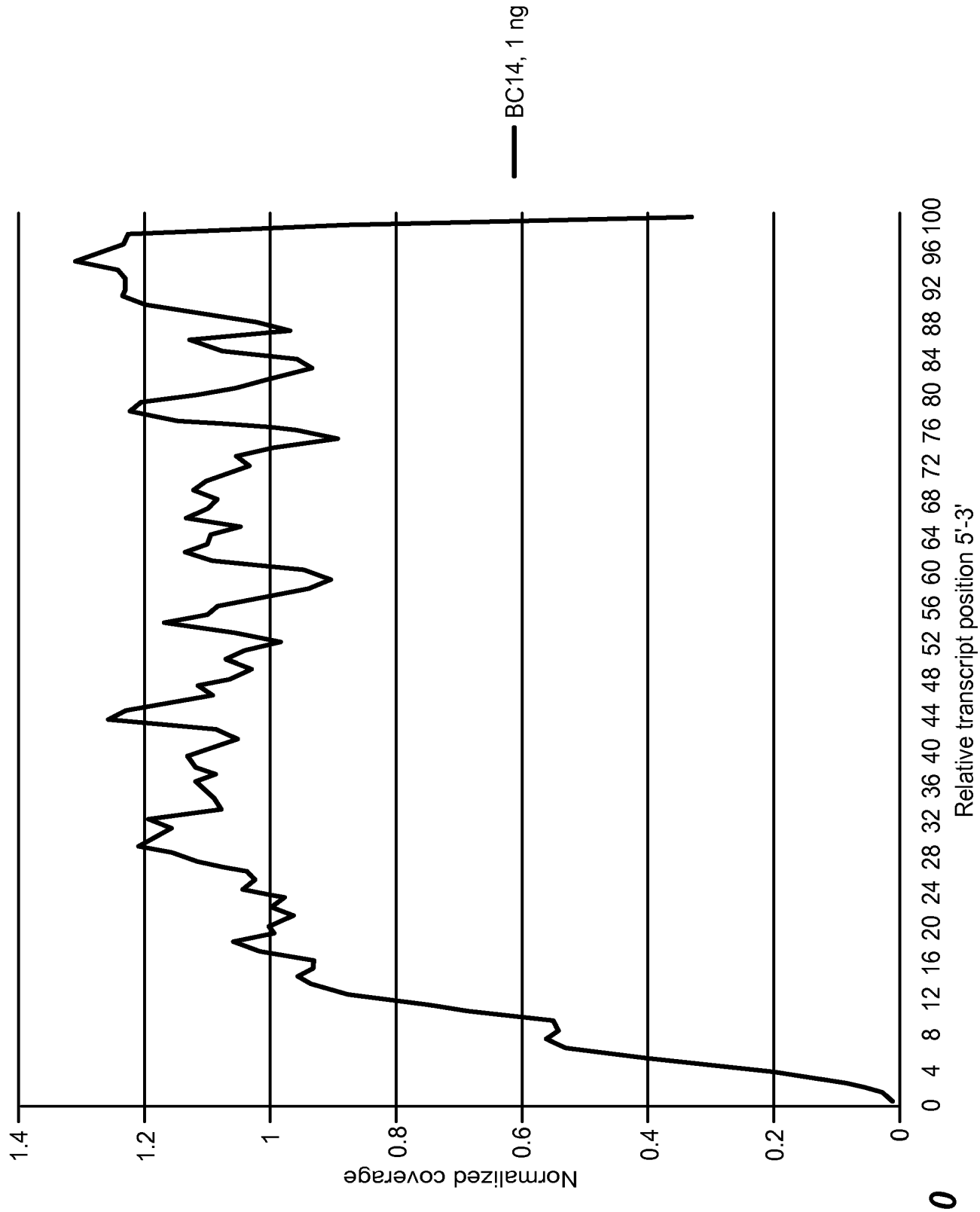

BC14, 1 ng
Normalized coverage
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0 4 8 12 16 20 24 28 32 36 40 44 48 52 56 60 64 68 72 76 80 84 88 92 96 100
Relative transcript position 5'-3'

FIG. 10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US14/24581

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C40B 40/06; C12P 19/34 (2014.01)
USPC - 506/16, 15, 13, 23
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): C40B 40/06; C12P 19/34 (2014.01)
USPC: 435/91.2, 91.1, 89, 85, 84, 72, 41; 506/16, 15, 13, 7, 23; 514/44R, 43, 42, 23.1

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
MicroPatent (US-G, US-A, EP-A, EP-B, WO, JP-bib, DE-C,B, DE-A, DE-T, DE-U, GB-A, FR-A); Google Scholar; Pubmed; ScienceDirect; overhang, '3',' '5',' 'directional library,' 'adapter molecule,' polynucleotide

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2011/0224105 A1 (KURN, N et al.) September 15, 2011; figure 7; paragraphs [0006], [0023], [0053], [0079], [0087], [0104]-[0107], [0129], [0130], [0134]	1-41
Y	BORODINA, T et al. A Strand-Specific Library Preparation Protocol For RNA Sequencing. Methods Enzymol. 2011, Vol. 500; pages 79-98; abstract; page 82, third paragraph; page 83, first paragraph. DOI: 10.1016/B978-0-12-385118-5.00005-0.	1-41
Y	US 2012/0283145 A1 (WANG, Y) November 8, 2012; paragraphs [0013], [0064]	13/12/9/1-13/12/9/4

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

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

Date of the actual completion of the international search	Date of mailing of the international search report
08 July 2014 (08.07.2014)	29 JUL 2014
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Shane Thomas PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

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

Espacenet

Bibliographic data: WO2015131107 (A1) — 2015-09-03

REDUCED REPRESENTATION BISULFITE SEQUENCING WITH DIVERSITY ADAPTORS

Inventor(s): SCHROEDER BENJAMIN G [US] ± (SCHROEDER, BENJAMIN, G)

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

Classification: - international: *C12P19/34*
- cooperative: C12Q1/6806; C12Q1/6827; C12Q1/6874 more

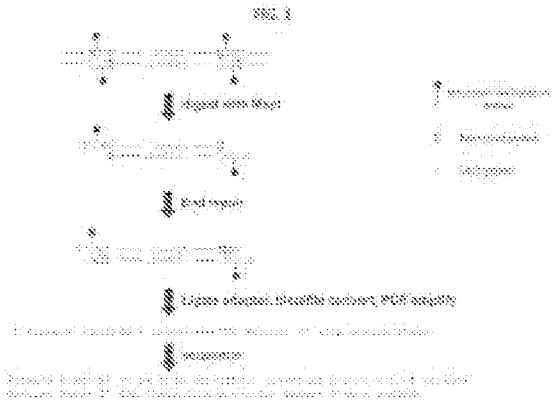
Application number: WO2015US18112 20150227 Global Dossier

Priority number(s): US201461968982P 20140321 ; US201461946617P 20140228

Also published as: US2015284769 (A1) US9745614 (B2)

Abstract of WO2015131107 (A1)

Described herein are methods, compositions and kits for the generation of bisulfite-converted libraries useful for conducting reduced representation bisulfite sequencing (RRBS). The methods described herein can be employed to generate RRBS libraries in a manner that is easier and more cost-efficient than conventional RRBS methods, and can be efficiently sequenced with next generation sequencing (NGS) techniques without the need for genomic, higher diversity sequencing controls such as PhiX spike-ins.



IDT Exhibit 1020.3692

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 September 2015 (03.09.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/131107 A1

(51) International Patent Classification:
C12P 19/34 (2006.01)

(21) International Application Number:
PCT/US2015/018112

(22) International Filing Date:
27 February 2015 (27.02.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/946,617 28 February 2014 (28.02.2014) US
61/968,982 21 March 2014 (21.03.2014) US

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

(72) Inventor: **SCHROEDER, Benjamin, G.**; 1964 Shoreview Avenue, San Mateo, CA 94401 (US).

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

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

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

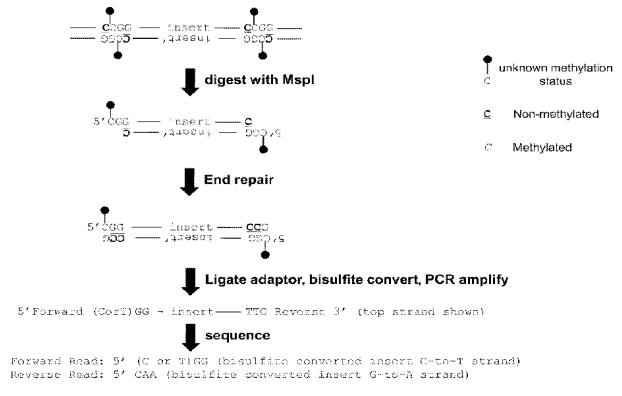
Published:

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

WO 2015/131107 A1

(54) Title: REDUCED REPRESENTATION BISULFITE SEQUENCING WITH DIVERSITY ADAPTORS

FIG. 1



(57) Abstract: Described herein are methods, compositions and kits for the generation of bisulfite-converted libraries useful for conducting reduced representation bisulfite sequencing (RRBS). The methods described herein can be employed to generate RRBS libraries in a manner that is easier and more cost-efficient than conventional RRBS methods, and can be efficiently sequenced with next generation sequencing (NGS) techniques without the need for genomic, higher diversity sequencing controls such as PhiX spike-ins.

IDT Exhibit 1020.3693

REDUCED REPRESENTATION BISULFITE SEQUENCING WITH DIVERSITY ADAPTORS

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application No. 61/946,617, filed on February 28, 2014, and U.S. Provisional Application No. 61/968,982, filed on March 21, 2014, each of which is herein incorporated by reference in its entirety.

BACKGROUND

[0002] DNA methylation is an epigenetic modification that can occur at the C5 position of cytosine residues. In mammals, 5-methylcytosine can appear in the CpG dinucleotide context (Ramsahoye *et al.*, *Proc Natl Acad Sci USA* 97:5237–5242, 2000). Recent data suggests that approximately 25 % of all cytosine methylation identified in stem cells can occur in non-CpG context (*see* Ziller *et al.*, *PLoS Genet.* 7(12):e1002389, 2011). Although CpG dinucleotides can be underrepresented in the genome, stretches of sequences known as CpG islands can exist that are rich in CpG dinucleotides. These CpG islands can be associated with promoter regions and span several hundred nucleotides or more.

[0003] Epigenomics, i.e., the study of the complete set of epigenetic modifications on the genetic material of a cell, has revealed that epigenetic modifications (e.g., DNA methylation) can play a role in mammalian development and disease. For example, DNA methylation is implicated in embryonic development, genomic imprinting and X-chromosome inactivation through regulation of transcriptional activity, chromatin structure and chromatin stability (Robertson, *Nat Review Genet* 6:597-610, 2005). Increased DNA methylation (hypermethylation) at promoter regions of genes can be associated with transcriptional silencing, whereas decreased methylation (hypomethylation) at promoter regions can be associated with increased gene activity. Aberrant methylation patterns can be associated with various human pathologies, including tumor formation and progression (Feinberg and Fogelstein, *Nature* 301:89-92, 1983; Esteller, *Nat Review Genet* 8: 286-298, 2007; and Jones and Paylin, *Cell* 128:683-692, 2007). Therefore, analysis of DNA methylation status across the human genome can be of interest.

[0004] Methods for measuring DNA methylation at specific genomic loci include, for example, immunoprecipitation of methylated DNA, methyl-binding protein enrichment of methylated fragments, digestion with methylation-insensitive restriction enzymes, and bisulfite conversion followed by Sanger sequencing (reviewed in Laird, *Nat Review Genet* 11: 191-203,

IDT Exhibit 1020.3694

2010). Bisulfite treatment can convert unmethylated cytosine residues into uracils (the readout of which can be thymine after amplification with a polymerase). Methylcytosines can be protected from conversion by bisulfite treatment to uracils. Following bisulfite treatment, methylation status of a given cytosine residue can be inferred by comparing the sequence to an unmodified reference sequence.

[0005] Techniques have been developed for profiling methylation status of the whole genome, i.e. the methylome, at a single-base resolution using high throughput sequencing technologies. Bisulfite conversion of genomic DNA combined with next generation sequencing (NGS), or BS-seq, is one strategy. Because of the high cost still associated with genome-wide methylation sequencing, variations of BS-seq technology that enable genome partitioning to enrich for regions of interest can be used. One such variation is reduced representation BS-seq (RRBS), which can involve digestion of a DNA sample with a methylation-insensitive restriction endonuclease that has CpG dinucleotide as a part of its recognition site, followed by bisulfite sequencing of the selected fragments (Meissner *et al.*, *Nucleic Acids Res*. 33(18):5868-5877, 2005).RRBS provides less coverage than whole genome bisulfite sequencing, but allows the researcher to obtain quantitative DNA methylation information across many features with approximately 50-fold fewer sequencing reads, resulting in substantial sequencing cost reduction. Since the RRBS technique was first described (Meissner, et al. (2005) *Nucleic Acids Res* 33(18):5868), there have been many enhancements and improvements (see for example, Boyle, et al. (2012) *Genome Biol* 13:R92). The current approach utilizes the methylation insensitive restriction enzyme MspI, which recognizes CCGG. As a result of partial fragmentation during bisulfite conversion, PCR, and efficiency of cluster generation, only a subset of these fragments, typically under 300 bp in length, can be sequenced. However these smaller fragments are derived from genomic DNA that has a high frequency of MspI sites and therefore a high frequency of potential CpG methylation sites. This is how RRBS can achieve reduced representation. One feature of RRBS can be that all forward reads start with either TGG or CGG. This feature can create a lack of color balance that can be problematic for some NGS sequencers. For example, the complete lack of color balance at bases 2 and 3 can make these samples very challenging for both Illumina MiSeq and HiSeq instruments. The methods, compositions and kits provided herein can allow RRBS libraries to be generated easily and sequenced more efficiently by increasing the diversity of nucleotide bases sequenced during the initial cycles of sequencing. The increased diversity can increase the efficiency and accuracy of cluster identification by sequencers during certain types of next generation sequencing applications (e.g., Illumina based sequencing by synthesis).

IDT Exhibit 1020.3695

SUMMARY

[0006] In one aspect, provided herein is a method for increasing diversity of sequences at ends of a plurality of polynucleotides in a sample, the method comprising introducing a pool of oligonucleotide sequences to 3' ends of the plurality of polynucleotides, a. wherein each oligonucleotide sequence in the pool of oligonucleotide sequences comprises a sequence complementary to a sequencing primer, b. wherein the pool of oligonucleotide sequences comprises a first sequence, wherein a terminal 3' nucleotide base of the sequencing primer is capable of annealing to a terminal 5' nucleotide base of the first sequence, wherein following introduction the 5' end of the first sequence is adjacent to a 3' end of a polynucleotide from the plurality of polynucleotides, c. wherein the pool of oligonucleotide sequences comprises a second sequence, wherein the second sequence terminates at its 5' end with an additional base relative to the 5' end of the first sequence, wherein following introduction the 5' end of the second sequence is adjacent to a 3' end of a polynucleotide from the plurality of polynucleotides; d. wherein the pool of oligonucleotide sequences comprises a third sequence, wherein the third sequence terminates at its 5' end with two additional bases relative to the 5' end of the first sequence, wherein following introduction the 5' end of the third sequence is adjacent to a 3' end of a polynucleotide from the plurality of polynucleotides; e. wherein the pool of oligonucleotide sequences comprises a fourth sequence, wherein the fourth sequence terminates at its 5' end with three additional bases relative to the 5' end of the first sequence, wherein following introduction the 5' end of the fourth sequence is adjacent to a 3' end of a polynucleotide from the plurality of polynucleotides; and f. wherein the introducing generates a plurality of oligonucleotide-polynucleotide complexes, thereby increasing diversity of sequences at ends of the plurality of polynucleotides. In some cases, the sample is a whole genome sample. In some cases, the sample comprises a polynucleotide library. In some cases, the method further comprises fragmenting the polynucleotides prior to introducing the pool of oligonucleotide sequences to the 3' ends of the plurality of polynucleotides. In some cases, the fragmenting is performed by contacting the polynucleotides with an enzyme. In some cases, the enzyme is a restriction enzyme. In some cases, the restriction enzyme is a methylation insensitive restriction enzyme. In some cases, the methylation insensitive restriction enzyme is MspI. In some cases, the additional base of the second sequence is an adenine, thymine, or cytosine base. In some cases, each of the two additional bases of the third sequence is an adenine, thymine, or cytosine base. In some cases, a 3' most base of the three additional bases of the fourth sequence comprises a thymine or cytosine base, and each of the two other bases of the three additional bases of the fourth sequence comprises an adenine, thymine, or cytosine base. In some cases, each oligonucleotide sequence

in the pool of oligonucleotide sequences is within a strand of a duplexed adaptor. In some cases, the duplexed adaptor is a partial duplexed adaptor comprising a short strand and a long strand. In some cases, the duplexed adaptor is a partial duplexed adaptor comprising two strands of equal length. In some cases, the short strand is a non-ligation strand of the duplexed adaptor, and the long strand is a ligation strand of the duplexed adaptor. In some cases, the non-ligation strand comprises a 5' overhang. In some cases, the 5' overhang comprises two bases. In some cases, the two bases are 3'-GC-5'. In some cases, the introducing comprises a ligation reaction. In some cases, the ligation reaction comprises cohesive end ligation. In some cases, the ligation reaction comprises blunt end ligation. In some cases, the plurality of polynucleotides are strands of double-stranded nucleic acids, wherein each 5' end of the double-stranded nucleic acids comprises a ligation end and each 3' end comprises a non-ligation end, wherein each of the non-ligation ends of the double-stranded nucleic acid comprises part of the polynucleotide sequence. In some cases, the 3'-GC-5' of the non-ligation strand of the duplexed adaptor anneals to a 5'-CG-3' sequence of the ligation end of the double-stranded sequence comprising the polynucleotide sequence. In some cases, the 3'-GC-5' of the non-ligation strand of the duplexed adaptor is 5' to the terminal 5' nucleotide base of the first sequence. In some cases, the 3'-GC-5' of the non-ligation strand of the duplexed adaptor is 5' to the additional base of the second sequence relative to the 5' end of the first sequence. In some cases, the 3'-GC-5' of the non-ligation strand of the duplexed adaptor is 5' to the two additional bases of the third sequence relative to the 5' end of the first sequence. In some cases, the 3'-GC-5' of the non-ligation strand of the duplexed adaptor is 5' to the three additional bases of the fourth sequence relative to the 5' end of the first sequence. In some cases, the ligation strand is capable of ligating to a 5' end of the plurality of polynucleotides, and wherein a nick is formed between the non-ligation strand and the end of the plurality of polynucleotides. In some cases, the method further comprises an extending step, wherein the non-ligation end of the plurality of polynucleotides is extended using a polymerase in the presence of a mixture of dNTPs using the ligation strand of the adaptor as template, thereby introducing the pool of oligonucleotide sequences to the 3' ends of the plurality of polynucleotides. In some cases, the ligation strand of each adaptor duplex in the pool comprise cytosine analogs resistant to bisulfite treatment and the mixture of dNTPs comprises dATP, dTTP, dGTP, and dCTP, whereby the extending generates the plurality of oligonucleotide sequence-polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment. In some cases, the ligation strand of each adaptor duplex in the pool of

IDT Exhibit 1020.3697

adaptor duplexes comprise cytosine bases sensitive to bisulfite treatment and the mixture of dNTPs comprises dATP, dTTP, dGTP, and a dCTP analog resistant to bisulfite treatment, whereby the extending generates the plurality of oligonucleotide sequence-polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment. In some cases, each oligonucleotide sequence in the pool of oligonucleotide sequences is a primer, wherein the pool of oligonucleotide sequences comprises the primer and a complement thereof. In some cases, the introducing comprises an amplification reaction. In some cases, the amplification reaction comprises polymerase chain reaction. In some cases, each of the primers in the pool comprises cytosine bases sensitive to bisulfite treatment, wherein the amplification is conducted in the presence of a mixture of dNTPs comprising dATP, dTTP, dGTP, and a dCTP analog resistant to bisulfite treatment, thereby generating the plurality of oligonucleotide sequence-polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment. In some cases, each of the primers in the pool of primers comprises cytosine analogs resistant to bisulfite treatment, wherein the amplification is conducted in the presence of a mixture of dNTPs comprising dATP, dTTP, dGTP, and dCTP, thereby generating the plurality of oligonucleotide sequence-polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment. In some cases, the cytosine analog resistant to bisulfite treatment is 5-methylcytosine, 5-hydroxymethylcytosine, or 5-propynylcytosine. In some cases, the cytosine analog resistant to bisulfite treatment is 5-methylcytosine. In some cases, the method further comprises amplifying the plurality of oligonucleotide sequence-polynucleotide complexes, wherein the amplifying comprises solid-phase nucleic acid amplification to generate a plurality of clusters, wherein each cluster represents a plurality of copies of each of the plurality of oligonucleotide sequence-polynucleotide complexes. In some cases, the method further comprises treating the plurality of oligonucleotide sequence-polynucleotides complexes with bisulfite prior to the amplification, wherein the treating generates a plurality of oligonucleotide sequence-polynucleotide complexes comprising non-complementary ends. In some cases, the treating with bisulfite prior to amplification generates the sequence complementary to a sequencing primer in each

IDT Exhibit 1020.3698

oligonucleotide sequence in the pool of oligonucleotide sequences comprises. In some cases, the method further comprises sequencing each of the plurality of clusters, wherein the sequencing comprises hybridizing the sequencing primer to the sequence complementary to a sequencing primer, and conducting sequencing by synthesis. In some cases, the introducing of the pool of oligonucleotide sequences to the polynucleotides improves identification of each of the plurality of clusters generated from the plurality of oligonucleotide sequence-polynucleotide complexes relative to the identification of each cluster from a plurality of clusters generated from the plurality of polynucleotides not introduced to the pool of oligonucleotide sequences. In some cases, the sequencing comprises paired-end sequencing.

[0007] In one aspect, provided herein is a method for generating a bisulfite converted library, the method comprising: a. introducing a pool of oligonucleotide sequences to each end of a plurality of polynucleotides from a sample, i. wherein each oligonucleotide sequence in the pool of oligonucleotide sequences comprises a sequence complementary to a sequencing primer and a sequence element adjacent to the sequence complementary to a sequencing primer, ii. wherein the pool of oligonucleotide sequences comprises a first sequence, wherein the first sequence terminates at its 3' end with the sequence element, wherein following ligation the 3' end of the sequence element is adjacent to a 5' end of a polynucleotide from the plurality of polynucleotides, iii. wherein the pool of oligonucleotide sequences comprises a second sequence, wherein the second oligonucleotide sequence terminates at its 3' end with an additional base relative to the 3' end of the first sequence, wherein following ligation the 3' end of the second sequence is adjacent to a 5' end of a polynucleotide from the plurality of polynucleotides; iv. wherein the pool of oligonucleotide sequences comprises a third sequence, wherein the third sequence terminates at its 3' end with two additional bases relative to the 3' end of the first sequence, wherein following ligation the 3' end of the third sequence is adjacent to a 5' end of a polynucleotide from the plurality of polynucleotides; v. wherein the pool of oligonucleotide sequences comprises a fourth sequence, wherein the fourth sequence terminates at its 3' end with three additional bases relative to the 3' end of the first sequence, wherein following ligation the 3' end of the fourth sequence is adjacent to a 5' end of a polynucleotide from the plurality of polynucleotides; vi. wherein the introducing generates a plurality of oligonucleotide sequence-polynucleotide complexes; and b. treating the plurality of oligonucleotide sequence-polynucleotide complexes with bisulfite, thereby generating a bisulfite converted polynucleotide library. In some cases, the sample is a whole genome sample. In some cases, the sample comprises a polynucleotide library. In some cases, the method further comprises fragmenting the polynucleotides prior to introducing the pool of oligonucleotide sequences to the 3' ends of the

IDT Exhibit 1020.3699

plurality of polynucleotides. In some cases, the fragmenting is performed by contacting the polynucleotides with an enzyme. In some cases, the enzyme is a restriction enzyme. In some cases, the restriction enzyme is a methylation insensitive restriction enzyme. In some cases, the methylation insensitive restriction enzyme is MspI. In some cases, the additional base of the second sequence is an adenine, thymine, or guanine base. In some cases, each of the two additional bases of the third sequence is an adenine, thymine, or guanine base. In some cases, a 5' most base of the three additional bases of the fourth sequence comprises an adenine or guanine base, and each of the two other bases of the three additional bases of the fourth sequence comprises an adenine, thymine, or guanine base. In some cases, each oligonucleotide sequence in the pool of oligonucleotide sequences is within a strand of a duplexed adaptor. In some cases, the duplexed adaptor is a partial duplexed adaptor. In some cases, the strand is a ligation strand of the duplexed adaptor. In some cases, the non-ligation strand comprises a 5' overhang. In some cases, the 5' overhang comprises two bases. In some cases, the two bases are 3'-GC-5'. In some cases, the introducing comprises a ligation reaction. In some cases, the ligation reaction comprises cohesive end ligation. In some cases, the ligation reaction comprises blunt end ligation. In some cases, the plurality of polynucleotides are double-stranded nucleic acids, wherein each 5' end of the double-stranded nucleic acids comprises a ligation end and each 3' end comprises a non-ligation end. In some cases, each duplexed adaptor in the pool comprises a ligation strand and a non-ligation strand, wherein the ligation strand is capable of ligating to the ligation end of the plurality of polynucleotides, and wherein a nick is formed between the non-ligation strand and the non-ligation end of the plurality of polynucleotides. In some cases, the method further comprises an extending step, wherein the non-ligation end of the plurality of polynucleotides is extended using a polymerase in the presence of a mixture of dNTPs using the ligation strand of the adaptor as template, thereby introducing the pool of oligonucleotide sequences to the 3' ends of the plurality of polynucleotides. In some cases, the ligation strand of each adaptor duplex in the pool comprise cytosine analogs resistant to bisulfite treatment and the mixture of dNTPs comprises dATP, dTTP, dGTP, and dCTP, whereby the extending generates the plurality of oligonucleotide sequence-polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment. In some cases, the ligation strand of each adaptor duplex in the pool of adaptor duplexes comprise cytosine bases sensitive to bisulfite treatment and the mixture of dNTPs comprises dATP, dTTP, dGTP, and a dCTP analog resistant to bisulfite treatment, whereby the extending generates the plurality of oligonucleotide sequence-

IDT Exhibit 1020.3700

polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment. In some cases, each oligonucleotide sequence in the pool of oligonucleotide sequences is within a primer, wherein the pool of oligonucleotide sequences comprises the primer and a complement thereof. In some cases, the introducing comprises an amplification reaction. In some cases, the amplification reaction comprises polymerase chain reaction. In some cases, each of the primers in the pool comprises cytosine bases sensitive to bisulfite treatment, wherein the amplification is conducted in the presence of a mixture of dNTPs comprising dATP, dTTP, dGTP, and a dCTP analog resistant to bisulfite treatment, thereby generating the plurality of oligonucleotide sequence -polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence -polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment. In some cases, each of the primers in the pool of primers comprises cytosine analogs resistant to bisulfite treatment, wherein the amplification is conducted in the presence of a mixture of dNTPs comprising dATP, dTTP, dGTP, and dCTP, thereby generating the plurality of oligonucleotide sequence-polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment. In some cases, the cytosine analog resistant to bisulfite treatment is 5-methylcytosine, 5-hydroxymethylcytosine, or 5-propynylcytosine. In some cases, the cytosine analog resistant to bisulfite treatment is 5-methylcytosine. In some cases, the method further comprises amplifying the plurality of amplifying the plurality of oligonucleotide sequence-polynucleotide complexes. In some cases, the amplifying comprises solid-phase nucleic acid amplification to generate a plurality of clusters, wherein each cluster represents a plurality of copies of each of the plurality of oligonucleotide sequence-polynucleotide complexes. In some cases, the method further comprises treating the plurality of oligonucleotide sequence-polynucleotides complexes with bisulfite prior to the amplification, wherein the treating generates a plurality of oligonucleotide sequence-polynucleotide complexes comprising non-complementary ends. In some cases, the sequence complementary to a sequencing primer in one end of the plurality of oligonucleotide sequence-polynucleotide complexes comprising non-complementary ends is complementary to a reverse read sequencing primer, and wherein the sequence complementary to a sequencing primer in an opposite end of the non-complementary

ends in the plurality of oligonucleotide sequence-polynucleotide complexes comprising non-complementary ends is complementary to a forward read sequencing primer. In some cases, the method further comprises sequencing each of the plurality of clusters, wherein the sequencing comprises hybridizing the sequencing primer to the sequence complementary to a sequencing primer, and conducting sequencing by synthesis, wherein the introducing of the pool of oligonucleotide sequences to the polynucleotides improves identification of each of the plurality of clusters generated from the plurality of oligonucleotide sequence-polynucleotide complexes relative to the identification of each cluster from a plurality of clusters generated from the plurality of polynucleotides not introduced to the pool of oligonucleotide sequences. In some cases, the sequence element comprises a barcode, a universal sequence, a linker sequence, or a random sequence.

[0008] In one aspect, provided herein is a kit comprising a pool of oligonucleotide sequences, a. wherein each oligonucleotide sequence in the pool of oligonucleotide sequences comprises a sequence complementary to a sequencing primer, b. wherein the pool of oligonucleotide sequences comprises a first oligonucleotide sequence, wherein a terminal 3' nucleotide base of the sequencing primer is capable of annealing to a terminal 5' nucleotide base of the first oligonucleotide sequence, c. wherein the pool of oligonucleotide sequences comprises a second oligonucleotide sequence, wherein the second oligonucleotide sequence terminates at a 5' end with an additional base relative to the 5' end of the first oligonucleotide sequence; d. wherein the pool of oligonucleotide sequences comprises a third oligonucleotide sequence, wherein the third oligonucleotide sequence terminates at a 5' end with two additional bases relative to the 5' end of the first oligonucleotide sequence; and e. wherein the pool of oligonucleotide sequences comprises a fourth oligonucleotide sequence, wherein the fourth oligonucleotide sequence terminates at a 5' end with three additional bases relative to the 5' end of the first oligonucleotide sequence. In some cases, each oligonucleotide sequence in the pool of oligonucleotide sequences is within a primer, wherein the pool of oligonucleotide sequences comprises the primer and a complement thereof. In some cases, the primer comprises cytosine analogs resistant to bisulfite conversion. In some cases, the primer comprises cytosine bases sensitive to bisulfite conversion. In some cases, each oligonucleotide sequence in the pool of oligonucleotide sequences is within a strand of a duplexed adaptor. In some cases, each duplexed adaptor in the pool comprises a ligation strand and a non-ligation strand. In some cases, the ligation strand and the non-ligation strand of each of duplexed adaptor in the pool comprise cytosine bases sensitive to bisulfite treatment. In some cases, the kit further comprises a mixture of dNTPs comprising dATP, dTTP, dGTP, and a dCTP analog resistant to bisulfite treatment. In

IDT Exhibit 1020.3702

some cases, the ligation strand and the non-ligation strand of each of duplexed adaptor in the pool comprise cytosine analogs resistant to bisulfite treatment. In some cases, the kit further comprises a mixture of dNTPs comprising dATP, dTTP, dGTP, and a dCTP analog sensitive to bisulfite treatment. In some cases, the kit further comprises a restriction enzyme. In some cases, the restriction enzyme is MspI. In some cases, the kit further comprises a DNA polymerase. In some cases, the DNA polymerase comprises exonuclease activity. In some cases, the DNA polymerase comprises strand displacement activity. In some cases, the non-ligation strand comprises a block at a 5' and 3' end, whereby the non-ligation strand is enzymatically unreactive. In some cases, the block at the 5' end comprises a biotin moiety. In some cases, the block at the 3' end comprises a dideoxy residue. In some cases, the cytosine analog resistant to bisulfite treatment is 5-methylcytosine, 5-hydroxymethylcytosine, or 5-propynylcytosine. In some cases, the cytosine analog resistant to bisulfite treatment is 5-methylcytosine.

[0009] In one aspect, provided herein is a method for analyzing sequence reads, the method comprising: a. generating a plurality of sequence reads, wherein the generating comprises conducting sequencing by synthesis on a plurality of clusters, wherein each cluster of the plurality of clusters comprises copies of one oligonucleotide sequence-polynucleotide from a plurality of oligonucleotide sequence-polynucleotides, wherein the sequencing by synthesis comprises hybridizing a sequencing primer to sequence complementary to the sequencing primer present in each cluster of the plurality of clusters, wherein the plurality of clusters are produced from solid-phase nucleic acid amplification of the plurality of oligonucleotide sequence-polynucleotides, wherein each of the plurality of oligonucleotide sequence-polynucleotides comprises a single nucleotide base extender sequence, a dinucleotide base extender sequence, a trinucleotide base extender sequence or no nucleotide base extender sequence between the sequence complementary to the sequencing primer of an oligonucleotide sequence and an end of a polynucleotide; and b. analyzing each of the plurality of sequence reads, wherein the analyzing comprises trimming or removing the extender sequence *in silico* from each of the plurality of sequence reads. In some cases, the plurality of oligonucleotide sequence-polynucloetides are generated by introducing a pool of oligonucleotide sequences to 3' ends of the plurality of polynucleotides from a sample, a. wherein each oligonucleotide sequence in the pool of oligonucleotide sequences comprises a sequence complementary to a sequencing primer, b. wherein the pool of oligonucleotide sequences comprises a first sequence, wherein a terminal 3' nucleotide base of the sequencing primer is capable of annealing to a terminal 5' nucleotide base of the first sequence, wherein following introduction the 5' end of the first sequence is adjacent to a 3' end of a polynucleotide from the plurality of polynucleotides, c. wherein the pool of

IDT Exhibit 1020.3703

oligonucleotide sequences comprises a second sequence, wherein the second sequence terminates at its 5' end with an additional base relative to the 5' end of the first sequence comprising the single nucleotide base extender, wherein following introduction the 5' end of the second sequence is adjacent to a 3' end of a polynucleotide from the plurality of polynucleotides; d. wherein the pool of oligonucleotide sequences comprises a third sequence, wherein the third sequence terminates at its 5' end with two additional bases relative to the 5' end of the first sequence comprising the dinucleotide base extender, wherein following introduction the 5' end of the third sequence is adjacent to a 3' end of a polynucleotide from the plurality of polynucleotides; e. wherein the pool of oligonucleotide sequences comprises a fourth sequence, wherein the fourth sequence terminates at its 5' end with three additional bases relative to the 5' end of the first sequence comprising the trinucleotide base extender, wherein following introduction the 5' end of the fourth sequence is adjacent to a 3' end of a polynucleotide from the plurality of polynucleotides; and f. wherein the introducing generates a plurality of oligonucleotide-polynucleotide complexes, thereby increasing diversity of sequences at ends of the plurality of polynucleotides. In some cases, the sample is a whole genome sample. In some cases, the sample comprises a polynucleotide library. In some cases, the method further comprises fragmenting the polynucleotides prior to introducing the pool of oligonucleotide sequences to the 3' ends of the plurality of polynucleotides. In some cases, the fragmenting is performed by contacting the polynucleotides with an enzyme. In some cases, the enzyme is a restriction enzyme. In some cases, the restriction enzyme is a methylation insensitive restriction enzyme. In some cases, the methylation insensitive restriction enzyme is MspI. In some cases, the additional base of the second sequence is an adenine, thymine, or cytosine base. In some cases, each of the two additional bases of the third sequence is an adenine, thymine, or cytosine base. In some cases, a 3' most base of the three additional bases of the fourth sequence comprises a thymine or cytosine base, and each of the two other bases of the three additional bases of the fourth sequence comprises an adenine, thymine, or cytosine base. In some cases, each oligonucleotide sequence in the pool of oligonucleotide sequences is within a strand of a duplexed adaptor. In some cases, the duplexed adaptor is a partial duplexed adaptor comprising a long strand and a short strand. In some cases, the short strand is a non-ligation strand of the duplexed adaptor and the long strand is a ligation strand. In some cases, the non-ligation strand comprises a 5' overhang. In some cases, the 5' overhang comprises two bases. In some cases, the two bases are 3'-GC-5'. In some cases, the introducing comprises a ligation reaction. In some cases, the ligation reaction comprises cohesive end ligation. In some cases, the ligation reaction comprises blunt end ligation. In some cases, the plurality of polynucleotides are strands of

double-stranded nucleic acids, wherein each 5' end of the double-stranded nucleic acids comprises a ligation end and each 3' end comprises a non-ligation end, wherein each of the non-ligation endsof the double-stranded nucleic acid comprises part of the polynucleotide sequence. In some cases, the 3'-GC-5' of the non-ligation strand of the duplexed adaptor anneals to a 5'-CG-3' sequence of the ligation end of the double-stranded sequence comprising the polynucleotide sequence. In some cases, the 3'-GC-5' of the non-ligation strand of the duplexed adaptor is 5' to the terminal 5' nucleotide base of the first sequence. In some cases, the 3'-GC-5' of the non-ligation strand of the duplexed adaptor is 5' to the additional base of the second sequence relative to the 5' end of the first sequence. In some cases, the 3'-GC-5' of the non-ligation strand of the duplexed adaptor is 5' to the two additional bases of the third sequence relative to the 5' end of the first sequence. In some cases, the 3'-GC-5' of the non-ligation strand of the duplexed adaptor is 5' to the three additional bases of the fourth sequence relative to the 5' end of the first sequence. In some cases, each duplexed adaptor in the pool comprises a ligation strand and a non-ligation strand, wherein the ligation strand is capable of ligating to a 5' end of the plurality of polynucleotides, and wherein a nick is formed between the non-ligation strand and the 3' end of the plurality of polynucleotides. In some cases, the method further comprises an extending step, wherein the 3' end of the plurality of polynucleotides is extended using a polymerase in the presence of a mixture of dNTPs using the ligation strand of the adaptor as template, thereby introducing the pool of oligonucleotide sequences to the 3' ends of the plurality of polynucleotides. In some cases, the ligation strand of each adaptor duplex in the pool comprise cytosine analogs resistant to bisulfite treatment and the mixture of dNTPs comprises dATP, dTTP, dGTP, and dCTP, whereby the extending generates the plurality of oligonucleotide-polynucleotide complexes, wherein each of the plurality of oligonucleotide-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment. In some cases, the ligation strand of each adaptor duplex in the pool of adaptor duplexes comprise cytosine bases sensitive to bisulfite treatment and the mixture of dNTPs comprises dATP, dTTP, dGTP, and a dCTP analog resistant to bisulfite treatment, whereby the extending generates the plurality of oligonucleotide-polynucleotide complexes, wherein each of the plurality of oligonucleotide-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment. In some cases, each oligonucleotide in the pool of oligonucleotides is a primer, wherein the pool of oligonucleotides comprises the primer and a complement thereof. In some cases, the introducing

IDT Exhibit 1020.3705

comprises an amplification reaction. In some cases, the amplification reaction comprises polymerase chain reaction. In some cases, each of the primers in the pool comprises cytosine bases sensitive to bisulfite treatment, wherein the amplification is conducted in the presence of a mixture of dNTPs comprising dATP, dTTP, dGTP, and a dCTP analog resistant to bisulfite treatment, thereby generating the plurality of oligonucleotide sequence-polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment. In some cases, each of the primers in the pool of primers comprises cytosine analogs resistant to bisulfite treatment, wherein the amplification is conducted in the presence of a mixture of dNTPs comprising dATP, dTTP, dGTP, and dCTP, thereby generating the plurality of oligonucleotide sequence-polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment. In some cases, the cytosine analog resistant to bisulfite treatment is 5-methylcytosine, 5-hydroxymethylcytosine, or 5-propynylcytosine. In some cases, the method further comprises treating the plurality of oligonucleotide sequence-polynucleotides complexes with bisulfite prior to the amplification, wherein the treating generates a plurality of oligonucleotide sequence-polynucleotide complexes comprising non-complementary ends. In some cases, the treating with bisulfite prior to amplification generates the sequence complementary to a sequencing primer in each oligonucleotide sequence in the pool of oligonucleotide sequences comprises. In some cases, the generating the plurality of sequence reads comprises paired-end sequencing.

[0010] In one aspect, provided herein is a method for increasing diversity of sequences at ends of a plurality of polynucleotide fragments in a library, the method comprising: a. fragmenting polynucleotides to generate polynucleotide fragments; b. ligating a pool of adaptor sequences to 3' ends of the plurality of polynucleotide fragments, wherein each adaptor sequence in the pool of adaptor sequences comprises a sequence complementary to a sequencing primer, wherein a 3' end of the sequencing primer anneals at least four bases from the 5' end of the adaptor sequence, and wherein each of four different types of nucleotide bases are present in each of the at least three terminal nucleotides at the 5' ends of the adaptors, thereby increase diversity of sequences at ends of the plurality of polynucleotide fragments.

IDT Exhibit 1020.3706

INCORPORATION BY REFERENCE

[0011] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] The novel features described herein are set forth with particularity in the appended claims. A better understanding of the features and advantages described herein will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles described herein are utilized, and the accompanying drawings of which:

[0013] FIG. 1 illustrates the basic principle of RRBS. Nucleic acid Library inserts (or genomic DNA) is digested with MspI, bisulfite converted (a chemical reaction that converts non-methylated C to U, but leaves 5-mC unreacted), and sequenced. C in the sequence indicates the presence of 5-mC at that position. Note how the first base of every sequence contains a methylation measurement.

[0014] FIG. 2A-B illustrates a method for generating the diversity adapters used in the methods provided herein. **FIG. 2A** illustrates a long strand of nucleotides (R01446) and 4 different short strands (i.e., Ben383, Ben384, Ben385, and Ben341) such that each short strand comprises a stretch of 12 nucleotides that comprise bases complementary to the long strand (R01446). **FIG. 2B** illustrates the method used to generate a diversity adaptor using the long strand (R01446) and a representative short strand (Ben383) from **FIG. 2A.**

[0015] FIG. 3 illustrates an alternative method for generating an RRBS library.

[0016] FIG. 4A-B illustrates the sequencing results of forward reads (FIG. 4A) and reverse reads (FIG. 4B) for sequencing an RRBS library generated using diversity-adapters comprising 3 "DDD" bases.

[0017] FIG. 5A-B illustrates the sequencing results of forward reads (FIG. 5A) and reverse reads (FIG. 5B) for sequencing an RRBS library generated using diversity-adapters comprising 2 "DD" bases.

[0018] FIG. 6A-B illustrates the sequencing results of forward reads (FIG. 6A) and reverse reads (FIG. 6B) for sequencing an RRBS library generated using diversity-adapters comprising a single "D" base.

[0019] FIG. 7A-B illustrates the sequencing results of forward reads (FIG. 7A) and reverse reads (FIG. 7B) for sequencing an RRBS library generated using diversity-adapters comprising no "D" bases.

[0020] FIG. 8 illustrates expected sequencing reads (forward and reverse reads) for sequencing an RRBS library with a pool of diversity-adapters comprising a variable number of "D" bases.

IDT Exhibit 1020.3707

The expected forward or reverse reads can be from N to NNNNNN bases in the first 6 sequencing cycles, wherein an N base can be an H, D, Y or B base.

[0021] FIG. 9A-B illustrates the sequencing results of forward reads (FIG. 9A) and reverse reads (FIG. 9B) for sequencing an RRBS library with a pool of diversity-adapters comprising a variable number of "D" bases.

[0022] FIG. 10 illustrates the sequencing results for sequencing an RRBS library with a 5-plex pool of diversity-adapters comprising a variable number of "D" bases, wherein the sequencing was performed without using a PhiX control as described in Example 2.

[0023] FIG. 11 illustrates the % base for each cycle for the sequencing reactions performed in **FIG. 10**.

[0024] FIG. 12 illustrates the % >Q20 score for each cycle for the sequencing reactions performed in **FIG. 10**.

[0025] FIG. 13 illustrates a Q-score heat map for the sequencing reactions performed in **FIG. 10**.

[0026] FIG. 14 illustrates the sequencing results for sequencing a blood sample using a 16 plex sequencing system (e.g., Ovation® Blood Seq Library System) as described in Example 2.

[0027] FIG. 15 illustrates a Q-score heat map for the sequencing reactions performed in **FIG. 14.**

[0028] FIG. 16 illustrates a run summary for the sequencing reactions performed in **FIG. 10.**

[0029] FIG. 17 illustrates a run summary for the sequencing reactions performed in **FIG. 14.**

[0030] FIG. 18A-C illustrates proposed rules for trimming diversity "D" bases *in silico* from sequence reads from nucleic acid inserts generated with mixed diversity-Adaptor sequences comprising 0 to 3 "D" bases (D0-D3). **FIG. 18A** illustrates a first rule. **FIG. 18B** illustrates a second rule. **FIG. 18C** illustrates a third rule.

[0031] FIG. 19 illustrates the total theoretical diversity possible for introducing to the ends of a polynucleotide upon introduction of a pool of diversity adaptors comprising diversity-adapters comprising 0 to 3 "D" bases (D0-D3).

[0032] FIG. 20 illustrates total CpGs analyzed using the modified RRBS methods provided herein in comparison to Whole Genome Bisulfite Sequencing (WGBS).

[0033] FIG. 21 illustrates a comparison of the average CpG per read and average CpG per uniquely aligned read between the modified RRBS method as provided herein vs. WGBS.

[0034] FIG. 22 illustrates the feature coverage using the modified RRBS as provided herein.

IDT Exhibit 1020.3708

[0035] FIG. 23 illustrates bioanalyzer traces of MspI-digested human genomic DNA (left), and the resulting RRBS sequencing library (top right) including an example of the nucleotide distribution (Forward Reads) obtained using said method (bottom right).

[0036] FIG. 24 illustrates a comparison between RRBS and other DNA methylation profiling technologies.

[0037] FIG. 25 illustrates a comparison of the *in silico* assignment of CpGs into classes for RRBS vs. whole genome bisulfite sequencing.

[0038] FIG. 26 illustrates an example of a dark sequencing approach used that can be used separately or in conjunction with RRBS.

[0039] FIG. 27A illustrates a polynucleotide insert from an RRBS generated using diversity adapters as provided herein. **FIG. 27B** illustrates the increased color balance in the initial sequence reads of an RRBS library generated using diversity adapters as provided herein and comprising polynucleotide inserts as shown in **FIG. 27A**.

[0040] FIG. 28 illustrates library size distribution on Bioanalyzer High Sensitivity DNA Chip.

[0041] FIG. 29 illustrates determination of the number of PCR cycles to perform during the library amplification.

[0042] FIG. 30 illustrates results from selecting Plot > deltaRn vs. Cycle

[0043] FIG. 31 illustrates cluster identification requires sequence diversity. The Illumina software identifies clusters over the first several cycles of sequencing. During sequencing of normal, high diversity clusters (top), overlapping clusters can be distinguished because they are different colors. If overlapping clusters contain the same sequence (bottom), theycan be mistaken as a single cluster.

[0044] FIG. 32 illustrates random N6 sequence adjacent to the barcode that can be used to mark PCR duplicates for *in silico* removal.

[0045] FIG. 33 illustrates high coverage overlap between three NuGEN RRBS technical replicates, each made from 25 ng of IMR90 gDNA. CpG's covered at 10X or greater are shown, after PCR duplicate removal.

[0046] FIG. 34 illustrates concordance in methylation levels. Concordance for CpG's covered at 20X or greater depth between NuGEN RRBS technical replicates (left) or between NuGEN RRBS (30 M reads) and published whole genome bisulfite sequencing (WGBS) from Lister et al., *Nature*, 2009, 462:315–322 (1180 M reads), which is herein incorporated by reference in its entirety.

[0047] FIG. 35 illustrates types of sequence diversity added to the beginning of forward reads in RRBS libraries generated using diversity adaptors with 0, 1, 2, or 3 diversity "D" bases.

IDT Exhibit 1020.3709

[0048] FIG. 36 illustrates the process of generating an RRBS library using MspI digestion and diversity adaptors lacking diversity bases.

[0049] FIG. 37 illustrates a schematic of a nucleic acid insert from an RRBS library generated using the library preparation process provided herein

[0050] FIG. 38 illustrates an overview of the RRBS library preparation process described in Example 6.

DETAILED DESCRIPTION

I. Overview

[0051] Provided herein are methods, compositions, and kits for introducing sequence diversity into genomic regions comprising low sequence diversity. The genomic regions comprising low sequence diversity can be highly conserved genomic regions between organisms (e.g., 16s rRNA sequences between microbial species). The genomic regions comprising low sequence diversity can be repetitive sequences within a genome (e.g., CpG islands within mammalian genomes). The methods described herein can introduce sequence diversity into genomic regions comprising low sequence diversity by introducing oligonucleotide sequence to one or both ends of a polynucleotide. The introduction can be through ligation (e.g., blunt-end or cohesive-end) using adaptors such that the adaptors comprise the oligonucleotide sequence introduced to one or both ends of the polynucleotide. The introduction can through amplification of a target sequence in a polynucleotide (e.g., genomic region) using primers such that the primers comprise the oligonucleotide sequence introduced to one or both ends of the polynucleotide. The primers can comprise sequence complementary to a target sequence in a nucleic acid as well as the oligonucleotide sequence. The oligonucleotide sequence can be present in the primer as a tail such that it is non-complementary to the target sequence. In general, the oligonucleotide sequence can comprise sequence complementary to a sequencing primer, one or more barcodes, universal sequence, random sequence, and/or sequence useful for introducing sequence diversity at one or both ends of a polynucleotide (e.g., extender sequence as provided herein). The oligonucleotide sequence can be configured such that sequence complementary to a sequencing primer is immediately adjacent to the end of the polynucleotide upon introduction (i.e., through amplification or ligation) of the oligonucleotide sequence complementary to a sequencing primer. In some cases, the oligonucleotide sequence further comprises an extender sequence such that the sequence complementary to a sequencing primer precedes the extender sequence such that, upon introduction, the extender sequence is immediately adjacent to the end of the polynucleotide.

IDT Exhibit 1020.3710

[0052] In the methods, compositions and kits provided herein, the extender sequence in an oligonucleotide sequence as described herein can be a variable number of nucleotides in length. The extender sequence can be less than, more than, or about 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 nucleotides in length. In some cases, the extender sequence is a single nucleotide in length. In some cases, the extender sequence is two nucleotides in length. In some cases, the extender sequence is three nucleotides in length. In some cases, an additional sequence element is present in the oligonucleotide sequence such that the additional sequence is located between the sequence complementary to a sequencing primer and the extender sequence such that upon introduction the extender sequence is immediately adjacent to the end of the polynucleotide and is followed by the additional sequence element which is followed by the sequence complementary to a sequencing primer. In some cases, the oligonucleotide sequence can comprise one or more additional sequence elements such that the one or more additional sequence elements are not located between the sequence complementary to a sequencing primer and an extender sequence such that the one or more additional sequence elements precedes the sequence complementary to a sequencing primer. In this case, upon introduction of the oligonucleotide sequence to one or both ends of a polynucleotide, the extender sequence is immediately adjacent to the end of the polynucleotide and is followed by the sequence complementary to a sequencing primer which is followed by the one or more additional sequence elements. In some cases, the oligonucleotide sequence can comprise a plurality of additional sequence elements such that one of the plurality of additional sequence elements is located between the sequence complementary to a sequencing primer and the extender sequence while one or more additional sequence elements of the plurality of additional sequence elements are not located between the sequence complementary to a sequencing primer and an extender sequence. The one or more additional sequence elements of the plurality of additional sequence elements precede the sequence complementary to a sequencing primer. In this case, upon introduction of the oligonucleotide sequence to one or both ends of a polynucleotide, the extender sequence is immediately adjacent to the end of the polynucleotide and is followed in succession by one of the plurality of additional sequence elements, the sequence complementary to a sequencing primer and the one or more additional sequence elements of the plurality of additional sequence elements.

[0053] An additional sequence element can be one or more barcodes, universal adaptor sequence, linker sequence, and/or random sequence. The sequencing primer can be a custom sequencing primer or can be a sequencing primer compatible with a commercially available NGS sequencing system.

IDT Exhibit 1020.3711

[0054] The polynucleotides can be in a sample. The sample can be any sample as provided herein. The polynucleotides can be fragmented prior to the introduction of an oligonucleotide sequence as described herein. Fragmentation can be mechanical (e.g., sonication) or chemical (e.g., enzymatic) as provided herein.

[0055] Following introduction of the oligonucleotide sequence to the ends of a polynucleotide, the resulting oligonucleotide sequence-polynucleotide complex can be amplified. In some cases, oligonucleotide sequence is introduced to both ends of a polynucleotide, and subsequent amplification is performed using primers directed to complementary sequence in the oligonucleotide sequences introduced to the ends of the polynucleotide. The primers used for amplification can introduce further sequence elements to the oligonucleotide sequence-polynucleotide complex formed after introduction of the oligonucleotide sequence. The sequence elements introduced during amplification can be barcode(s) or sequence complementary to a capture probe on a solid surface. The solid surface can be a bead or planar surface (e.g., flow cell). The amplification can be a solid-state amplification (e.g., cluster amplification) that generates a cluster comprising a plurality of copies of the oligonucleotide sequence-polynucleotide complex. The methods provided herein can further comprise sequencing the amplified oligonucleotide-polynucleotide complexes. An extender sequence can be a barcode, universal sequence, linker sequence, or random sequence.

[0056] In the methods, compositions and kits provided herein, the extender sequence in an oligonucleotide sequence as described herein can comprise any of the four canonical nucleotide bases (i.e., thymine, adenine, cytosine, or guanine) found in DNA. The nucleotide base composition of a nucleotide in an extender sequence can be selected such that the nucleotide base composition can differ from the nucleotide base composition at an end of a polynucleotide to which the extender sequence can be introduced, whereby the presence of one or more nucleotides from the extender sequence at the end of the polynucleotide can serve to increase sequence read length as well as nucleotide base diversity. The increase based diversity can aid identification of individual clusters as well as maintain accurate label detection in NGS sequencing systems that use cluster based sequencing by synthesis (e.g., Illumina).

[0057] The methods, compositions, and kits provided herein can comprise a pool of oligonucleotide sequences as described herein. The pool of oligonucleotide sequences can be a pool of adaptors comprising the oligonucleotide sequence as described herein. The pool of adaptors comprising the oligonucleotide sequence can be introduced to one or both ends of a plurality of polynucleotides through ligation (e.g., blunt end or cohesive end). The pool of oligonucleotide sequences can be a pool of primers comprising the oligonucleotide sequence as

IDT Exhibit 1020.3712

described herein. The pool of primers comprising the oligonucleotide sequence can be introduced to one or both ends of a plurality of polynucleotides through amplification (e.g., isothermal or PCR). The pool of oligonucleotide sequences can comprise sets of oligonucleotide sequences such that each set comprises either no extender sequence or an extender sequence of a length that differs from the length of an extender sequence in a separate set of oligonucleotide sequences within the pool of oligonucleotide sequences. The length of an extender sequence in a set of oligonucleotide sequences in a pool of oligonucleotide sequences can be 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more nucleotides.

[0058] In some cases, a pool of oligonucleotide sequences comprises a first set of oligonucleotide sequences that do not comprise an extender sequence, a second set of oligonucleotide sequences that comprise an extender sequence comprising a single nucleotide, a third set of oligonucleotide sequences that comprise an extender sequence comprising two nucleotides, and a fourth set of oligonucleotide sequences that comprise an extender sequence comprising three nucleotides. Each oligonucleotide sequence within the pool of oligonucleotide sequences can comprise the same sequence complementary to a sequencing primer as described herein. The nucleotide base composition of each nucleotide in an extender sequence comprising more than one nucleotides can be identical or different. The nucleotide base composition of each nucleotide in an extender sequence comprising more than one nucleotide can be known or randomly assigned as provided herein.

[0059] In some cases, the methods, compositions, and kits provided herein are useful for generating bisulfite converted nucleic acid libraries with increased sequence diversity at the ends of nucleic acid inserts within the library. The bisulfite converted libraries can be reduced representation bisulfite sequencing (RRBS) libraries. Such libraries can be useful, for example, for determining the methylation status across a genome, or alternatively, for determining the methylation status at given genomic loci. The RRBS libraries generated by the methods, compositions, and kits provided herein can be useful for determining the methylation status of CpG islands in polynucleotides derived from whole genome samples. The whole genome samples can be from eukaryotic cells (e.g., mammalian cells). The methods can produce RRBS libraries that can be more efficiently sequenced by current NGS sequencers (e.g., Illumina sequencers), and in some embodiments, without the need for spiking in a genomic, higher-diversity sample nucleic acids (e.g. PhiX control library). The method for generating an RRBS library can comprise enzymatically fragmenting a nucleic acid (polynucleotide) such that a plurality of nucleic acids fragments comprising substantially similar nucleotide compositions at the ends or termini of each fragment are generated. In some cases, the method comprises

IDT Exhibit 1020.3713

fragmenting the nucleic acid with a methylation sensitive restriction enzyme such as, for example, MspI. In some cases, a pool of adaptors is ligated to the 5' ends of each nucleic acid fragment in the plurality of nucleic acid fragments such that each nucleic acid fragment comprises double stranded sequence and the ligation generates a plurality of adaptor-ligated nucleic acid fragments. The pool of adaptors can comprise adaptors that share a sequence complementary to a sequencing primer as well as a first set of adaptors comprising no extender sequence, a second set comprising a single nucleotide extender sequence, a third set comprising a dinucleotide extender sequence, and a fourth set comprising a trinucleotide extender sequence. In some cases, the nucleotide base composition of the extender sequences in a pool of adaptors comprising extender sequence are chosen such that the base composition differs for the nucleotide base composition at the end of the nucleic fragment to which the extender sequence is appended. In some cases, the single nucleotide in first set of adaptors can comprises a thymine, guanine, or adenine base, while the each nucleotide in the dinucleotide of the second set can comprise a thymine, guanine, or adenine base and two of the three nucleotides in the trinucleotide of the third set can be a thymine, guanine, or adenine base, while the third nucleotide can be guanine or adenine. As provided herein, the extender sequence can be used to increase sequence diversity between the sequencing primer and the end of a library nucleic acid insert. In some cases, the adaptor-ligated nucleic acid fragments are subjected to an extension reaction as provided herein followed by denaturation and bisulfite treatment, whereby the bisulfite treatment generates adaptor-ligated nucleic acid fragment comprising non-complementary adaptor sequence on opposite ends. The non-complementary adaptor sequences can subsequently be utilized to amplify and/or sequence each adaptor-ligated nucleic acid fragment. The amplification can be solution phase and/or solid-state amplification. In some cases, the plurality of adaptor-ligated nucleic acid fragments are PCR amplified followed by solid-state amplification to generate clusters such that each cluster comprises thousands of copies of an adaptor-ligated nucleic acid fragment form the plurality of adaptor-ligated nucleic acid fragments. The presence of extender sequence in the plurality of adaptor-ligated nucleic acid fragments can increase the ability of NGS sequencers (e.g., Illumina) to successfully identify and discern between neighboring clusters formed by solid-state amplification during sequencing of the clusters (e.g., as illustrated in **FIG. 31**).

[0060] Provided herein are methods for generating adaptors comprising extender sequences for use in the generation of bisulfite converted nucleic acid libraries (e.g., RRBS libraries). Adaptors comprising extender sequences can also be referred to as diversity adaptors.

IDT Exhibit 1020.3714

[0061] Provided herein are *in silico* methods for filtering and trimming the extender sequence present in sequences reads produced from sequencing nucleic acids or polynucleotides comprising extender sequence as generated through the methods provided herein. The *in silico* methods can comprise trimming away any oligonucleotide sequence (e.g., adaptor sequence) that may be present on the 3' ends of reads using publically available trimming algorithms (e.g., Trim Galore). In some cases, the trimming script or algorithm can remove adaptor sequence such as adaptor sequence associated with a specific sequencing platform (e.g., Illumina adaptor sequence) and/or any nucleotide bases whose quality score is below a set quality score (Q-score) threshold. In some cases, the Q-score threshold is set at 30, 20, or 10 such that any Q-score below said thresdold is trimmed or removed from analysis by the script or algorithm. Once the oligonucleotide sequence (e.g., adaptor sequence) is trimmed away, a custom python script can remove any reads that do not contain the known end sequence of the library insert. In RRBS libraries generated through fragmentation via MspI digestion, the python script can remove any reads that do not contain an MspI site signature, YGG (Y= cytosine of thymine bases), at the 5' end. For paired end data an MspI site signature can be required at 5' ends of both sequences. Following filtering and trimming, all reads should begin with YGG, where Y is C or T, 5 bases can be trimmed from every read (6 bases are trimmed for paired-end to prevent alignment issues). If YGG is not found in the first 6 bases, the read is discarded.

[0062] Provided herein are methods for identifying and removing duplicate sequence reads following NGS sequencing. In some cases, an oligonucleotide sequence introduced to one or both ends of a polynucleotide comprises additional sequence elements upstream of a sequence complementary to a sequencing primer relative to an end of a polynucleotide to which the oligonucleotide sequence has been appended. In some cases, the additional sequence elements comprise a barcode and a stretch of random nucleotides immediately adjacent thereto. In some cases, the stretch of random nucleotides is 6 nucleotides in length. The methods provided herein can utilize an *in silico* analysis to discriminate between true duplicates produced during amplification (e.g., PCR). The analysis can utilize *in silico* algorithms in order to utilize information provided by the unique stretch of random nucleotides to discriminate between true PCR duplicates and independent adaptor ligated-nucleic acid fragments. In some cases, the algorithm for removing duplicate sequence reads uses sequence reads recovered from an index read. The index read can provide sequence data or reads comprising the barcode(s) and the random sequence located adjacent to the barcode(s). In order to read the inserted random sequence adjacent to the barcode(s) during an index read, the cycle number of the index read can be increased by the number of nucleotide bases present in the inserted random sequence. For

IDT Exhibit 1020.3715

example, the barcode can be 6 nucleotides and the random sequence can be 6 nucleotides and the index read can be a total of 12 cycles.

II. Adaptors

[0063] In one aspect provided herein is a method for increasing sequence diversity at the ends of polynucleotides that comprise conserved sequence at one or both ends with other polynucleotides using adaptors comprising extender sequence as described herein. The adaptor comprising an extender sequence can be used in the methods described herein to increase the sequence diversity at one or both ends of a polynucleotide by appending the adaptor comprising the extender sequence to one or both ends of the polynucleotide. The increased sequence diversity can aid an NGS sequencer in determining the sequence of each polynucleotide in a group of polynucleotides that share identical or substantially similar nucleotide base compositions at their end or ends when conducting NGS sequencing using NGS methods and systems that generate and identify clusters of individual polynucleotides during sequencing (e.g., Illumina, SOLiD). As a result, an adaptor comprising extender sequence as described herein can be referred to as a diversity adaptor and each nucleotide and its corresponding nucleotide base can be referred to as a diversity base.

[0064] **FIG. 35** illustrates the DNA sequence derived from a bisulfite converted library constructed using diversity adaptors such that following introduction of the diversity adaptors to the ends of DNA polynucleotides, the library comprises polynucleotides with either no diversity bases (i.e., 0 diversity bases added) derived from the diversity adaptor, one diversity base (i.e., 1 diversity base added) derived from the diversity adaptor, two diversity bases (i.e., 2 diversity bases added) derived from the diversity adaptor, or three diversity bases (i.e., 3 diversity "3" bases added) derived from the diversity adaptor. The number of diversity bases added can be variable and can be tailored to the type of library being constructed (e.g., 16S rRNA libraries, RRBS libraries, etc.). The diversity bases in a diversity adaptor can be "D" bases such that a "D" base can be an adenine, thymine, or guanine such as in FIGs. 2A-B, 3, 4A-B, 5A-B, 6A-B, 7A-B, 8, 9A-B, 18A-B and 35. The diversity bases in a diversity adaptor can be "R" bases such that an "R" base can be an adenine or guanine such as in FIG. 18B and 35. The complement of a "D" or an "R" base can be an "H" base. In some cases, an "H" base is a complement of a "D" base such that the "H" base can be a thymine when the "D" base is an adenine, an adenine when the "D" base is a thymine, or a cytosine when the "D" base is guanine. In some cases, an "H" base is a complement of an "R" base such that the "H" base can be a thymine when the "R" base is an adenine or a cytosine when the "R" base is guanine. In some cases, each diversity base in a diversity adaptor is a "D" base. In some cases, each base in a diversity adaptor is an "R" base.

IDT Exhibit 1020.3716