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

Materials and Methods

Step 1) Nebulization

Materials:

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

Procedure:

Mixed 5 µl (5 µg) of PhiX174 DNA with 45 µl of TE and 700 µl 1 of nebulization buffer. Mixed 10 µl (5 µg) of BAC DNA with 40 µl of TE and 700 µl of nebulization buffer. Chilled DNA solutions were each fragmented in a nebulizer on ice for 6 minutes under 32 pounds per square inch (psi) of pressure. The recovered volumes were each purified with a Qiagen PCR purification kit column and eluted in 30 µl of EB.

Step 2) End-repair

Materials:

- Nebulized DNA (from Step 1)
- Water

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

Procedure:

End repair mix was assembled as follows:

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

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

Step 3) A- tailing Reaction

Materials:

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

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

Procedure:

The following reaction mix was assembled:

End repaired DNA	30µl
Water	2ul
NEB buffer 2	5µl
dATP	10µl
Klenow fragment (3' to 5' exo minus)	3µl
	50µl total

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

Step 4) Annealed Adaptors

Materials:

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

Procedure:

The oligos were mixed together to a final concentration of 15uM each, in 10mM Tris/10mM NaCl pH 7.0. 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.

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

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

Step 5) Ligation

Materials:

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

Procedure:

Reaction mix was assembled as follows:

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

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

Step 6) Gel purification

Materials:

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

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

Procedure:

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

Step 7) Exonuclease I Treatment of PCR Primers

Materials:

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

Procedure:

DNA Primers with a phosphorothioate at the n-1 position (5 x 85 µl of each Primer (approx 25 µM)) were aliquoted 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. Subsequently, 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_2O was added to the columns and the columns respun. The filtered solutions were placed into a SpeedVac® and evaporated until each was at 20 µl, and the fractions combined. The pooled fractions were injected into a reverse phase HPLC system, and the main peak was collected. The collected fractions were evaporated to dryness in a SpeedVac®, 50 µl of water was added and the fraction was subjected again to evaporation to dryness. The resulting pellets were dissolved in 50 µl of water, pooled and a UV measurement taken to determine the concentration of the oligonucleotide.

Step 8) PCR

Materials:

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

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

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

Procedure:

The PCR reaction was prepared as follows:

BAC Library	
Gel purified BAC DNA	1µl
Phusion mastermix	25µl
Universal PCR primer	1µl
Indexing PCR primer 1	1µl
Water	22µl
	50µl total

PhiX174 Library	
Gel purified PhiX174 DNA	1µl
Phusion mastermix	25µl
Universal PCR primer	1µl
Indexing PCR primer 2	1µl
Water	22µl
	50µl total

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

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

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

mixing and amplification on a surface amplification platform.

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

Universal LPX primer 5′
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCxT (SEQ ID No. 7) 3′
Lpx primer 5′
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCxT (SEQ ID No. 8)
Tag1 primer (CGTGAT) 5′
CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTC (SEQ ID No. 9) 3′
Tag2 primer (ACATCG) 5′
CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTC (SEQ ID No. 10) 3′
Tag3 primer (GCCTAA) 5′
CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTC (SEQ ID No. 11) 3′
Tag4 primer (TGGTCA) 5′
CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTC (SEQ ID No. 12) 3′
Tag5 primer (CACTGT) 5′
CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTC (SEQ ID No. 13) 3′
Tag6 primer (ATTGGC) 5′
CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTC (SEQ ID No. 14) 3′

Tag7 primer (GATCTG) 5′
CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTC (SEQ ID No. 15)
3′
Tag8 primer (TCAAGT) 5′
CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTC (SEQ ID No. 16)
3′
Tag9 primer (CTGATC) 5′
CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTC (SEQ ID No. 17)
3′
Tag10 primer (AAGCTA) 5′
CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTC (SEQ ID No. 18)
3′
Tag11 primer (GTAGCC) 5′
CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTC (SEQ ID No. 19)
3′
Tag12 primer (TACAAG) 5′
CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTC (SEQ ID No. 20)
3′

The two universal primers contain a phosphorothioate group (x) and were exonucleases treated and purified prior to amplification. The amplification conditions were as described above, and the 6 tagged human samples, and 6 tagged Phi-X sample were pooled together for sequencing on clusters.

Validation of Libraries by Conventional Sanger Sequencing

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

cloned fragments sequenced by conventional Sanger sequencing.

16 clones from BAC Library (ATC tag)

1 (204bp) Insert: E.coli 85bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTACTGATTTCATTGCAGCCAAAGGCAAACTTTGGCTGCATCGTTTTA
CAGTCGCCATAAGCCTTTCCTCTGTTAAACCGCCTTCTGAGATCGGAAGAGC
GGTTCAGCAGGAATGCCGAGACCGATCTCGTATGCCGTCTTCTGCTTG
(SEQ ID No. 21)

2 (214bp) Insert: BAC 95bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTATCAATATTGTGAAAATGACCATACTGCCAAAAAAAAACTACAAATT
CAATGCAATTTTCATCAAAATACCATCATCATTCTTCACAATATTGATAGATC
GGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCGTATGCCGTCTTCT
GCTTG (SEQ ID No. 22)

3 (215bp) Insert: BAC 96bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTCTCACTCCTGGCAGAGGGACGTGTGACTAGCCATGGGCCCCTAG
GTCTCCAGTTCCTGGGTAGCTTGTATTTTTGAACATCTCCTGTATATTAGTTA
GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCGTATGCCGT
CTTCTGCTTG (SEQ ID No. 23)

4 (147bp) Insert: BAC 28bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTAGTGTAGTTGAGATCTGCCTTAGCAGCAAGATCGGAAGAGCGGTT
CAGCAGGAATGCCGAGACCGATCTCGTATGCCGTCTTCTGCTTG (SEQ ID No. 24)

5 (183bp) Insert: BAC 64bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTAACACATTTCAAAGTTTGGGGCCCTCCTCCTCCCCAAAAAACAAAC
CACAAAAAACAAACAAAAAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
ACCGATCTCGTATGCCGTCTTCTGCTTG (SEQ ID No. 25)

6 (170bp) Insert: BAC 59bp

GGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTG
AATGCCTTTTATAGCATTTAATTTTTCCTAAGTATAATTACCAAATAAAAATTG
TATAAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCGTATG
CCGTCTTCTGCTTG (SEQ ID No. 26)

7 (180bp) Insert: BAC 61bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTTGGGCCCGGGAGGAGTTTGCCGGGGAGGAGTGGGTTTGGAATCG
GGGTTAAAGGAAAGAGAAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
ACCGATCTCGTATGCCGTCTTCTGCTTG (SEQ ID No. 27)

8 (190bp) Insert: BAC 73bp

TGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC
GATCTAAGATCTATTTCAAATGGACTGTAGATCTAAGTATAAAAGGTAAGAGA
ATAATTATTCTAGAAAGTAAATGTAAGATCGGAAGAGCGGTTCAGCAGGAAT
GCCGAGACCGATCTCGTATGCCGTCTTCTGCTTG (SEQ ID No. 28)

9 (192bp) Insert: BAC 74bp

AATGATACGGCGACCACCAGATCTACACTCTTTCCCTACACGACGCTCTTCC
GATCTGGGAGGCCAAGGTGGGTGGATCACCTGAGATCAGGAGTTCGAGAC
CAGCTGGCCAACATGATGAAACTCTGTCTAGATCGGAAGAGCGGTTCAGCA
GGAATGCCGAGACCGATCTCGTATGCCGTCTTCTGCTTG (SEQ ID No.
29)

10 (185bp) Insert: BAC 66bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTTGACCATTGTAACCATTAATGTAGACTGCAATGATATGCACTATTTA
CAACCTTTTTTAAGACTCTAGATCGGAAGAGCGGTTCAGCAGGAATGCCGA
GACCGATCTCGTATGCCGTCTTCTGCTTG(SEQ ID No. 30)

11 (199bp) Insert: BAC 80bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTCTTTGAAGAGCTGGCAGTAGAAGATAAACAGGCTGGGGAAGAAGA
GAAAGTGCTCAAGGAGAAGGAGCAGCAGCAGCAGC
AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCGTATGCCGT
CTTCTGCTTG(SEQ ID No. 31)

12 (212bp) Insert: BAC 93bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTAGTATTCAACAAGTCTGTCTTTTCCAAGTGTCTTTAAAGACCAGAAA
TACCTGTTTTTAACACACAGGGTTGCAAAATTCAGAGGAGATTGGCAGATCG
GAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCGTATGCCGTCTTCTG
CTTG(SEQ ID No. 32)

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

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTCTTGAGATGAGTGATGACGGCGCGCTGGAAGTTGCTCGTCGCGCT
CGCGGTACGCCGCGCATTGCCAACCGTCTGCTGCGTCGAGTGCGTGATTTC
GCCGAAGTGAAGCACGATGGCACCATCTCAAGAGATCGGAAGAGCGGTTCA
GCAGGAATGCCGAGACCGATCTCGTATGCCGTCTTCTGCTTG(SEQ ID No.
33)

14 (202bp) Insert: BAC 83bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTGGGGTTGGTGGAACCCAGATGCCTCCCAGGATTGGTGGGCCCTG

TGGCACTTGTACCTGCTGTTGCTGTTGCTGCTGCTGCTGAGATCGGAAGAG
CGGTTCAGCAGGAATGCCGAGACCGATCTCGTATGCCGTCTTCTGCTTG(SE
Q ID No. 34)

15 (166bp) Insert: BAC 47bp
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTATGATAAGGAGCAGGTTTACAGATCATAAGTGCAAAAGCGGGCGA
GAAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCGTATGC
CGTCTTCTGCTTG(SEQ ID No. 35)

16 (147bp) Insert: BAC 31bp
GATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG
ATCTCTGATACTGTTGTAACCACCCAATTGGTTCAAGATCGGAAGAGCGGTT
CAGCAGGAATGCCGAGACCGATCTCGTATGCCGTCTTCTGCTTG(SEQ ID
No. 36)

16 clones from PhiX174 Library (CGA tag)

1 (183bp) Insert: PhiX 67bp
AATGATACGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGA
TCTTAAAGCTACCAGTTATATGGCTGTTGGTTTCTATGTGGCTAAATACGTTA
ACAAAAAGTCAGATATGAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGA
CCGCGATCGTATGCCGTCTTCTGCTTG(SEQ ID No. 37)

2 (174bp) Insert: PhiX 58bp
GATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG
ATCTCGACCACTCGCGATTCAATCATGACTTCGTGATAAAAGATTGAGTGTG
AGGTTATAACAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGCGAT
CGTATGCCGTCTTCTGCTTG(SEQ ID No. 38)

3 (179bp) Insert: PhiX 62bp

TGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC
GATCTTCTATGAAGGATGTTTTCCGTTCTGGTGATTCGTCTAAGAAGTTTAAG
ATTGCTGAGGGTCAAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACC
GCGATCGTATGCCGTCTTCTGCTTG (SEQ ID No. 39)

4 (169bp) Insert: PhiX 50bp
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTGGTTGCGCCGCCAAAACGTCGGCTACAGTAACTTTTCCCAGCCTC
AATCTAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGCGATCGTAT
GCCGTCTTCTGCTTG (SEQ ID No. 40)

5 (166bp) Insert: PhiX 50bp
TGATACGGCGACCACCGAGATCTACATCTTTCCCTACACGACGCTCTTCCGA
TCTAACAACTGAACGGACTGGAAACACTGGTCATAATCATGGTGGCGAATAA
GAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGCGATCGTATGCC
GTCTTCTGCTTG (SEQ ID No. 41)

6 (164bp) Insert: PhiX 45bp
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTAAATACGTTAACAAAAAGTCAGATATGGACCTTGCTGCTAAAGGTA
GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGCGATCGTATGCCGT
CTTCTGCTTG (SEQ ID No. 42)

7 (194bp) Insert: PhiX 75bp
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTGTCAATATAACCAGTAGTGTTAACAGTCGGGAGAGGAGTGGCATTA
ACACCATCCTTCATGAACTTAATCCACTGAGATCGGAAGAGCGGTTCAGCAG
GAATGCCGAGACCGCGATCGTATGCCGTCTTCTGCTTG (SEQ ID No. 43)

8 (198bp) Insert: PhiX 79bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTGATAGTTTGACGGTTAATGCTGGTAATGGTGGTTTTCTTCATTGCAT
TCAGATGGATACATCTGTCAACGCCGCTAATCAGATCGGAAGAGCGGTTCA
GCAGGAATGCCGAGACCGCGATCGTATGCCGTCTTCTGCTTG (SEQ ID No. 44)

9 (179bp) Insert: PhiX 60bp
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTTGAAACCAACATAAACATTATTGCCCGGCGTACGAGGAAGGACGT
CAATAGTCACACAGTAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACC
GCGATCGTATGCCGTCTTCTGCTTG (SEQ ID No. 45)

10 (198bp) Insert: PhiX 79bp
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTCCAGCTTCTTCGGCACCTGTTTTACAGACACCTAAAGCTACATCGT
CAACGTTATATTTTGATAGTTTGACGGTTAATAAGATCGGAAGAGCGGTTCA
GCAGGAATGCCGAGACCGCGATCGTATGCCGTCTTCTGCTTG (SEQ ID No. 46)

11 (151bp) Insert: PhiX 32bp
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTCAGTTTTGCCGCAAGCTGGCTGCTGAACGCCCAGATCGGAAGAGC
GGTTCAGCAGGAATGCCGAGACCGCGATCGTATGCCGTCTTCTGCTTG (SEQ ID No. 47)

12 (173bp) Insert: PhiX 54bp
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTTTTGCTGATGAACTAAGTCAACCTCAGCACTAACCTTGCGAGTCAT
TTCTTTGAAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGCGATCG
TATGCCGTCTTCTGCTTG (SEQ ID No. 48)

13 (151bp) Insert: PhiX 32bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTAATAGACGCAACGCGAGCAGTAGACTCCTTCTAGATCGGAAGAGC
GGTTCAGCAGGAATGCCGAGACCGCGATCGTATGCCGTCTTCTGCTTG (SEQ ID No. 49)

14 (168bp) Insert: PhiX 49bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTGGAGGCCTCCAGCAATCTTGAACACTCATCCTTAATACCTTTCTTTT
TGAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGCGATCGTATGC
CGTCTTCTGCTTG (SEQ ID No. 50)

15 (181bp) Insert: PhiX 62bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTACTCCTAAGCAGAAAACCTACCGCGCTTCGCTTGGTCAACCCCTCA
GCGGCAAAAATTAAAAAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGAC
CGCGATCGTATGCCGTCTTCTGCTTG (SEQ ID No. 51)

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

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTGAGCCGACATCGAGGTGCCAAACACCGCCGTCGATATGAACTCTT
GGGCGGTATCAGCCTGTTAGATCGGAAGAGCGGTTCAGCAGGAATGCCGA
GACCGCGATCGTATGCCGTCTTCTGCTTG (SEQ ID No. 52)

These results confirm that the library preparation method produces 2 libraries of 'sequenceable' DNA templates. Each library contained a plurality of genomic inserts, each of which was flanked by the two adaptors (AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID No. 53) and AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGNNNTCGTATGCCGTCTTCTGCTT G (SEQ ID No. 54)), required for cluster formation and SBS

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

Example of Cluster preparation using the mixed PHiX and BAC libraries

Example 2): Acrylamide coating of glass chips

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

Chips were washed as follows: neat Decon for 30 minutes, milliQ H_2O for 30 minutes, NaOH 1N for 15 minutes, milliQ H_2O for 30 minutes, HCl 0.1N for 15 minutes, milliQ H_2O for 30 minutes.

Polymer solution preparation:

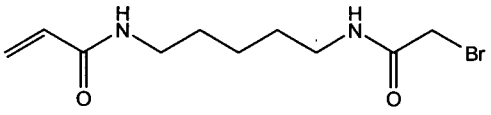
For 10 ml of 2% polymerisation mix.

- 10 ml of 2% solution of acrylamide in milliQ H2O
- 165 µl of a 100mg/ml *N*-(5-bromoacetamidylpentyl) acrylamide (BRAPA) solution in DMF (23.5 mg in 235µl DMF)
- 11.5 µl of TEMED
- 100 µl of a 50 mg/ml solution of potassium persulfate in milliQ H_2O (20mg in 400µl H_2O)

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

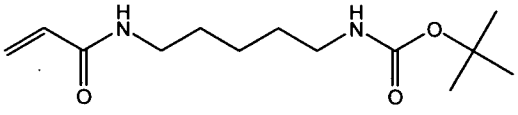
acrylamide solution. The mixture was then quickly vortexed and immediately used. Polymerization was then carried out for 1h 30 minutes at RT. Afterwards the channels were washed with milliQ H_2O for 30 minutes and filled with 0.1 M potassium phosphate buffer for storage until required.

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



(**1**)

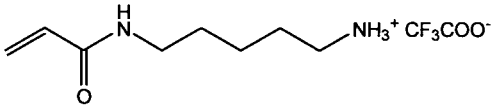
N-Boc-1,5-diaminopentane toluene sulfonic acid was obtained from Novabiochem. The bromoacetyl chloride and acryloyl chloride were obtained from Fluka. All other reagents were Aldrich products.



(**2**)

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

yield 2.56 g (9.98 mmol, 71 %) of product **2** as a beige solid. ^{1}H NMR (400 MHz, d_6-DMSO) : 1.20-1.22 (m, 2H, CH_2), 1.29-1.43 (m, 13H, tBu, 2xCH_2), 2.86 (q, 2H, J = 6.8 Hz and 12.9 Hz, CH_2), 3.07 (q, 2H, J = 6.8 Hz and 12.9 Hz, CH_2), 5.53 (dd, 1H, J = 2.3 Hz and 10.1 Hz, CH), 6.05 (dd, 1H, J = 2.3 Hz and 17.2 Hz, CH), 6.20 (dd, 1H, J = 10.1 Hz and 17.2 Hz, CH), 6.77 (t, 1H, J = 5.3 Hz, NH), 8.04 (bs, 1H, NH). Mass (electrospray+) calculated for $C_{13}H_{24}N_2O_3$ 256, found 279 (256+Na^+).



(**3**)

Product **2** (2.56g, 10 mmol) was dissolved in trifluoroacetic acid:dichloromethane (1:9, 100 ml) and stirred at room temperature. The progress of the reaction was monitored by TLC (dichloromethane:methanol 9:1). On completion, the reaction mixture was evaporated to dryness, the residue co-evaporated three times with toluene and then purified by flash chromatography (neat dichloromethane followed by a gradient of methanol up to 20%). Product **3** was obtained as a white powder (2.43 g, 9 mmol, 90%). ^{1}H NMR (400 MHz, D_2O): 1.29-1.40 (m, 2H, CH_2), 1.52 (quint., 2H, J = 7.1 Hz, CH_2), 1.61 (quint., 2H, J = 7.7 Hz, CH_2), 2.92 (t, 2H, J = 7.6 Hz, CH_2), 3.21 (t, 2H, J = 6.8 Hz, CH_2), 5.68 (dd, 1H, J = 1.5 Hz and 10.1 Hz, CH), 6.10 (dd, 1H, J = 1.5 Hz and 17.2 Hz, CH), 6.20 (dd, 1H, J = 10.1 Hz and 17.2 Hz, CH). Mass (electrospray+) calculated for $C_8H_{16}N_2O$ 156, found 179 (156+Na^+).

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

bromoacetyl chloride (2.07 ml, 1.1eq), through a pressure equalized dropping funnel, over a one hour period and at -60°C (cardice and isopropanol bath in a Dewar). The reaction mixture was then stirred at room temperature overnight and the completion of the reaction was checked by TLC (dichloromethane:methanol 9:1) the following day. The salts formed during the reaction were filtered off and the reaction mixture evaporated to dryness. The residue was purified by chromatography (neat dichloromethane followed by a gradient of methanol up to 5%). 3.2 g (11.55 mmol, 51 %) of the product **1** (BRAPA) were obtained as a white powder. A further recrystallization performed in petroleum ether:ethyl acetate gave 3g of the product **1**. ^{1}H NMR (400 MHz, d_6-DMSO) : 1.21-1.30 (m, 2H, CH_2), 1.34-1.48 (m, 4H, 2xCH_2), 3.02-3.12 (m, 4H, 2xCH_2), 3.81 (s, 2H, CH_2), 5.56 (d, 1H, J = 9.85 Hz, CH), 6.07 (d, 1H, J = 16.9 Hz, CH), 6.20 (dd, 1H, J = 10.1 Hz and 16.9 Hz, CH), 8.07 (bs, 1H, NH), 8.27 (bs, 1H, NH). Mass (electrospray+) calculated for $C_{10}H_{17}BrN_2O_2$ 276 or 278, found 279 (278+H^+), 299 (276+Na^+).

Example 4) Grafting primers onto surface of SFA coated chip

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

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

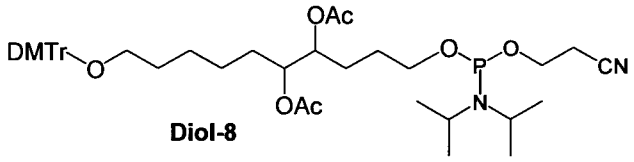
The primers are typically 5'-phosphorothioate oligonucleotides incorporating any specific sequences or modifications required for cleavage. Their sequences and suppliers vary according to the experiment for which they are to be used, and in this case were complementary to the 5'-ends of the template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

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

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

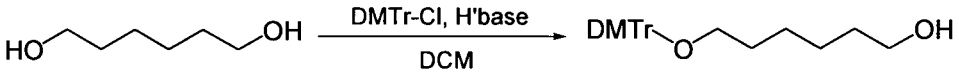
P5diol: 5' PS-TTTTTTTTTT-diol-AATGATACGGCGACCACCGA (SEQ ID No. 55)
P7: 5' PS-TTTTTTTTTTCAAGCAGAAGACGGCATACGA (SEQ ID No. 56)

Example 5: Preparation of diol-phosphoramidite for DNA coupling



Diol-8

Step 1:



Diol-1

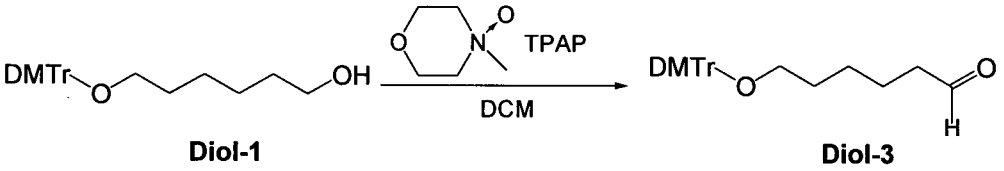
1,6-Hexanediol (Sigma Aldrich 99 %) (14.6 g, 124 mmol), *N,N*-diisopropylethylamine (Hünig's base; Sigma Aldrich; redistilled) (21.6 mL, 124 mmol) is dissolved in anhydrous DCM/ DMF (250/ 50 mL) under N_2. The solution is cooled to 0°C and the first portion of 4, 4'-dimethoxytrityl chloride (DMTr-Cl; Sigma-Aldrich 95 %) (10.5 g, 31 mmol) is added. The reaction mixture is then warmed to room temperature. After stirring for 1 h, the reaction mixture is cooled to 0°C again and the second portion of DMTr-Cl (10.5 g, 31 mmol) is added and then allowed to stir at room temperature for another 2 hours. TLC (EtOAc: petroleum ether 4:6) analysis indicates *ca.* 95% consumption of starting material derivative (DMTr-OH). The reaction is concentrated under reduced pressure and Aq. $NaHCO_3$ (sat.) solution (500 mL) is poured into the residue. The resulting mixture is extracted with petroleum ether/ EtOAc (2:1) (3 x 1000 mL). The

combined organic layers are dried over $MgSO_4$, and concentrated under vacuum. The residue is co-evaporated with xylene (2 x 100 mL) to remove DMF. The reaction mixture, is pre-absorbed on silica gel and subjected to flash chromatography using solvents containing 1 % Et_3N petroleum ether to petroleum ether/ EtOAc (7:3) as eluent. The yield of pale yellow oil is 16.58 g, 64 %, with a further 7.8g (17%) of the bis-tritylated by-product.

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

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

Step 2:



Diol-1 Diol-3

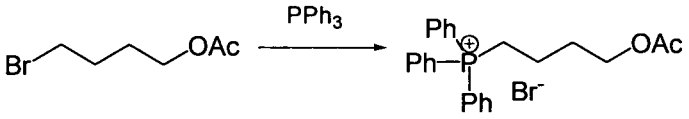
To a solution of Diol-1 (16.6 g, 39.5 mmol) in anhydrous DCM (200 mL), tetrapropylammonium perruthenate (TPAP; Sigma Aldrich 97 %) (277 mg, 0.79 mmol) is added under N_2 atmosphere. The solution is cooled to 0 °C and *N*-methylmopholine *N*-oxide (Sigma Aldrich 97 %) (2.7 g, 23 mmol) is added. The reaction is warmed to room temperature. After 1 hour, the other three portions of *N*-methylmopholine *N*-oxide (3 x 2.0 g, 51.2 mmol) are added within a period of

four hours. TLC (EtOAc: petroleum ether 4:6) indicates the reaction goes to completion. The reaction is quenched with aq. $NaHCO_3$ (sat.) (1000 mL) and extracted to CH_2Cl_2 (4 x 1000 mL). The combined organic layers are dried over $MgSO_4$. The solution is concentrated under reduced pressure. Diol-3, 9.9 g, 60 %, is isolated by flash chromatography using solvents containing 1% Et_3N from petroleum ether to petroleum ether/ EtOAc (6:4) as eluent, as a pale yellow oil.

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

^1H NMR (400MHz, $CDCl_3$): δ 1.30-1.37 (m, 2H, CH_2), 1.48-1.57 (m, 4H, 2x CH_2), 2.34 (td, J = 1.7 and 7.4 Hz, 2H, CH_2CHO), 2.97 (s, 2H, CH_2O), 3.72 (s, 6H, 2x MeO), 6.73-6.76 (m, 4H, Ph), 7.10-7.26 (m, 7H, Ph), 7.34-7.36 (m, 2H, Ph), 9.67 (t, J = 1.7, 1H, CHO).

Step 3:



Diol-2

A solution of triphenylphosphine (Sigma-Aldrich 99 %, ReagentPlus™). (39.3 g, 150 mmol) and 4-bromobutyl acetate (Sigma-Aldrich)(26 mL, 180 mmol) in anhydrous toluene (300 mL) is heated under reflux for 36 hours under N_2 in an oil-bath (140 °C). During the reflux, oil is precipitated out. The reaction mixture is cooled to room temperature. TLC (petroleum ether/ EtOAc 7:3) analysis of the toluene

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

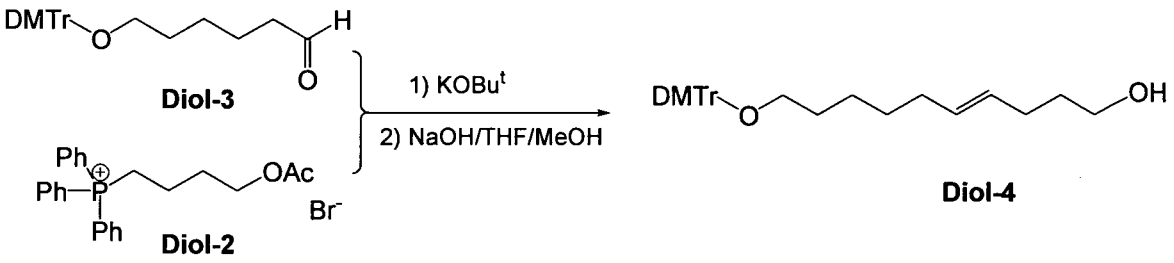
TLC: R_f: 0.0 (petroleum ether/ EtOAc 7:3).

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

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

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

Step 4:



Diol-2 (10.34 g, 22.7 mmol) is weighed into a round-bottomed flask and dissolved with DCM (20 mL). The solution is then evaporated under reduced pressure until it gives a white foam. The flask is then subjected to high vacuum for 24 h. To this flask, anhydrous THF (180 mL) is added under N_2. The resulting suspension is cooled to -78 °C with an acetone-dry ice bath. With vigorous stirring, KOBu^t (3.3 g, 29.5 mmol) is added under N_2. Slowly the colour of the suspension turns orange, and white solids are gradually precipitated out. To this suspension, a solution of diol-3 (dried at 60 °C under high vacuum for 1 h before the reaction),(9.5 g, 22.7 mmol) in THF (50 mL) is added drop wise over half an hour. Acetone-dry ice bath is then removed. The reaction mixture is slowly warmed to room temperature and stirred for another hour. The colour of the reaction mixture turns yellow after the addition of diol-3. The reaction mixture is concentrated down under reduced pressure. The resulting residue is partitioned between DCM (800 mL) and aq. NaCl (sat.) (800 mL). The aqueous layer is extracted with an additional DCM (2 x 800 mL). The organic extractions are combined, dried over $MgSO_4$, filtered, and evaporated under reduced pressure to give yellow oil. The oil is dissolved in THF/MeOH (125/100 mL) and cooled to 0°C. To this solution, NaOH (1M in H_2O, 25 mL) is added. After allowing the reaction to stir for 1 hour, TLC analysis

indicates full consumption of starting material. The reaction mixture is neutralized with acetic acid (1.5 mL). The reaction mixture is concentrated down under reduced pressure. The resulting residue is partitioned between DCM (800 mL) and aq. $NaHCO_3$ (sat.) (800 mL). The aqueous layer is extracted with additional DCM (2 x 800 mL). The organic extractions are combined, dried over $MgSO_4$, filtered, and evaporated to give a pale yellow oil. Diol-4, 6.45 g, 60 % is isolated by flash chromatography using solvents containing 1% Et_3N from petroleum ether to petroleum ether/ EtOAc (5:5) as eluent, as a light yellow oil.

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

^1H NMR (400 MHz, $CDCl_3$) δ 1.24-1.32 (m, 4H, 2x CH_2), 1.54-1.57 (m, 4H, 2x CH_2), 1.93-1.96 (m, 2H, CH_2), 2.02-2.07 (m, 2H, CH_2), 2.96 (t, J = 6.6 Hz, 2H, CH_2O), 3.54-3.59 (m, 2H, CH_2OH), 3.72 (s, 6H, 2x MeO), 5.29-5.32 (m, 2H, 2x =CH), 6.73-6.77 (m, 4H, Ph), 7.11-7.27 (m, 7H, Ph), 7.36-7.38 (m, 2H, Ph).

Step 5:



Diol-4 Diol-5

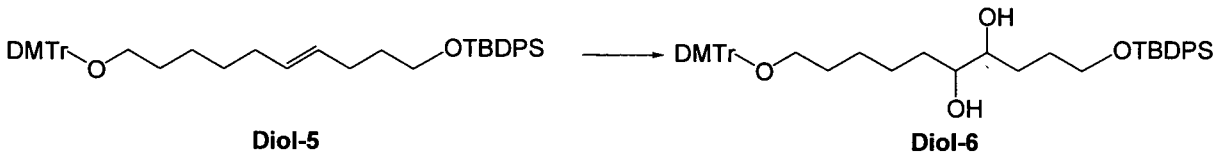
To a solution of Diol-4 (5.68 g, 12 mmol) and imidazole (Sigma Aldrich, 99 %), (1.63 g, 24 mmol) in anhydrous DMF (100 mL), *t*-butyldiphenylsilyl chloride (Sigma Aldrich, 98

%), (4.05 mL, 15.6 mmol) is added drop wise under N_2 atmosphere at room temperature. The reaction is stirred for 1 hour. TLC (petroleum ether/ EtOAc 8:2) indicates that the starting material is fully consumed. A saturated aq. $NaHCO_3$ solution (500 mL) is added to quench the reaction. The resulting mixture is extracted with petroleum ether/EtOAc (2:1) (3 x 500 mL). The organic layers are combined, dried over $MgSO_4$, filtered, and evaporated to give a yellow oil. Diol-5, 8.14 g, 95 % is isolated by flash chromatography using solvents containing 1% Et_3N from petroleum ether to petroleum ether/ EtOAc (9:1) as eluent, as a colourless oil.

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

^1H NMR (400 MHz, $CDCl_3$): δ 0.97 (s, 9H, 3xMe), 1.19-1.30 (m, 4H, 2x CH_2), 1.48-1.55 (m, 4H, 2x CH_2), 1.91-1.95 (m, 2H, CH_2), 2.01-2.06 (m, 2H, CH_2), 2.95 (t, J = 6.6 Hz, 2H, CH_2O), 3.58 (t, J = 6.3 Hz, 2H, CH_2O), 3.70 (s, 6H, 2x MeO), 5.24-5.27 (m, 2H, 2x =CH), 6.72-6.75 (m, 4H, Ph), 7.11-7.37 (m, 15H, Ph), 7.57-7.60 (m, 4H, Ph).

Step 6:



Diol-5 Diol-6

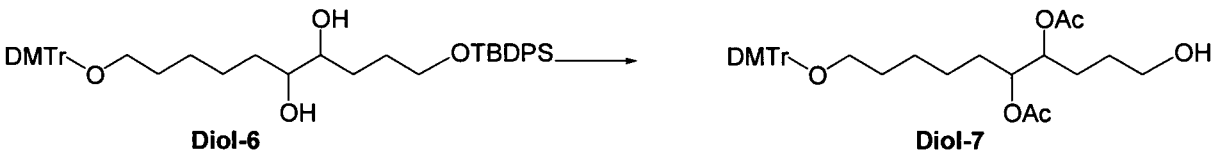
A mixture of diol-5 (9.27 g, 13 mmol), AD-mix-α (Sigma Aldrich), (18.2 g), methanesulfonamide (Sigma Aldrich, 97 %), (1.23 g, 13 mmol), t-BuOH (65 mL) and water (65 mL) is stirred together vigorously at 55 °C for 14 h. The TLC

analysis (petroleum ether : EtOAc 6:4) indicates *ca.* 95 % consumption of the starting material. The reaction mixture is cooled to room temperature, treated with sodium sulfite (15.3 g, 12 mmol), then further stirred for 30 minutes. A saturated aq. $NaHCO_3$ solution (500 mL) is added to the reaction. The resulting mixture is extracted with EtOAc (3 x 500 mL). The organic layers are combined, dried over $MgSO_4$, filtered, and evaporated to give yellow oil. Diol-6, 7.96 g, 82 %, is isolated by flash chromatography (silica gel, Fluka, 70-230 mesh) using solvents containing 1% Et_3N from petroleum ether to petroleum ether/ EtOAc (1:1) as elutant, as a white solid.

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

^1H NMR (400 MHz, $CDCl_3$) δ 1.07 (s, 9 H, 3xMe), 1.41-1.7 (m, 12 H, 6x CH_2), 1.94 (d, *J* = 4.3 Hz, 1H, OH), 2.94-2.95 (m, 1H, OH), 3.06 (t, *J* = 6.6 Hz, 2H, CH_2O), 3.61-3.63 (m, 2 H, 2x C*H*OH), 3.73 (t, *J* = 5.6 Hz, 2H, CH_2O), 3.81 (s, 6H, 2x MeO), 5.24-5.27 (m, 2H, 2x =CH), 6.82-6.85 (m, 4H, Ph), 7.21-7.47 (m, 15 H, Ph), 7.57-7.60 (m, 4 H, Ph).

Step 7:



Diol-6 Diol-7

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

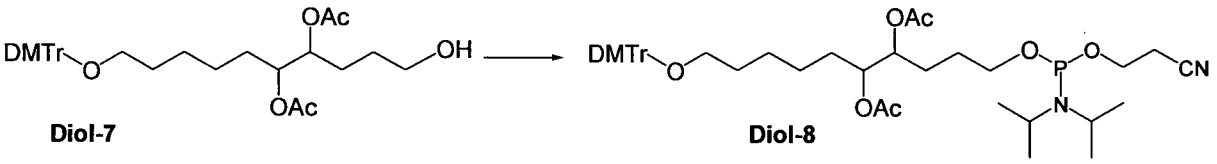
mixture of pyridine (15 mL) and DCM (30 mL), acetic anhydride (Fluka 99%), (2.5 mL, 26.68 mmol) is added at room temperature. TLC analysis (petroleum ether:EtOAc 6:4) indicates full consumption of the starting material after 1 h. The reaction is quenched by saturated aq. $NaHCO_3$ solution (500 mL). After 5 minutes, the mixture is extracted with DCM (3 x 500 mL). The organic layers are combined, dried over $MgSO_4$, filtered, and evaporated. The residue is co-evaporated with toluene (2 x 100 mL). The resulting yellow oil is subjected to a plug of silica gel (50 g, Fluka, 70-230 mesh) to remove DMAP using eluents containing 0.1% Et_3N from petroleum ether to petroleum ether/ EtOAc (7:3) (250 mL each). The combined fractions of product are concentrated to dryness. The resulting colourless oil is dissolved in THF (100 mL) and treated with TBAF (Sigma-Aldrich; 5% wt water), (1 M in THF, 15 mL) at 0 °C. The reaction solution is slowly warmed to room temperature and stirred for 2 hours. TLC analysis (petroleum ether:EtOAc 6:4) indicates that desilylation is completed. The volatile solvent (THF) is evaporated under reduced pressure at low temperature. A saturated aq. $NaHCO_3$ solution (500 mL) is added to the residue. The resulting mixture is extracted with EtOAc (3 x 500 mL). The organic layers are combined, dried over $MgSO_4$, filtered, and evaporated to give yellow oil. Diol-7, 4.2 g, 66 %, is isolated by flash chromatography using solvents containing 1% Et_3N from petroleum ether to petroleum ether/ EtOAc (1:1) as eluent, as a white solid.

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

^1H NMR (400 MHz, $CDCl_3$) δ 1.29-1.33 (m, 4H, 2x CH_2), 1.47-1.63 (m, 8 H, 4x CH_2), 1.99, 2.01 (2s, 6H, 2 MeC(O)), 3.00

(t, *J* = 6.5 Hz, 2H, CH_2O), 3.60-3.64 (m, 2 H, CH_2O), 3.75 (s, 6H, 2x MeO), 4.92-4.97 (m, 2H, 2x C*H*OAc), 6.76-6.80 (m, 4H, Ph), 7.15-7.29 (m, 7 H, Ph), 7.38-7.40 (m, 2 H, Ph).

Step 8:



Diol-7 Diol-8

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

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

^1H NMR (400 MHz, $CDCl_3$) δ 1.09, 1.10, 1.11, 1.12 (4 s, 12 H, $N(CHMe_2)_2$), 1.26-1.31 (m, 4H, 2x CH_2), 1.45-1.56 (m, 8 H, 4x CH_2), 1.95, 1.969, 1.971, 1.98 (4 s, 6H, 2 MeCO), 2.56 (t, *J* = 6.5 Hz, 2H, CH_2CN), 2.95 (t, *J* = 6.5 Hz, 2H, CH_2O), 3.49-3.55 (m, 4 H, CH_2O), 3.72 (s, 6H, 2x MeO), 4.89-4.92 (m, 2H,

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

^{31}P NMR (162 MHz, $CDCl_3$): 148.67, 148.69 ppm.

Example 6) Preparation of clusters by isothermal amplification:

Step 1: Hybridisation and amplification

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

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

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

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

Table 1. Template Hybridization and 1st Extension

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

Table 2. Isothermal Amplification

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

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

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

Step 2: Blocking extendable 3'-OH groups

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

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

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

	wait				
5	Pump wash buffer	20	300	15	75

Example 7: Linearisation and hybridization of a sequencing primer

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

To linearize the nucleic acid within the clusters formed within the flow cell channels, 300 µL per channel of linearisation mix is flowed in at 15 µL/minutes at 20 °C; followed by 75 µL of water at the same flow rate.

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

5'-ACACTCTTTCCCTACACGACGCTCTTCCGATC (SEQ ID No. 57)

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

Step	Description	T (°C)	Time (sec)	Flow rate (µl/min)	Pumped V (µl)
1	Pump 0.1 M NaOH	20	300	15	75
2	Pump TE	20	300	15	75
3	Pump Primer mix	20	300	15	75

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

Example 8: Sequencing from the target fragment

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

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

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

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

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

Heat to 60 °C.

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

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

Cool to 20 °C.

Wash with Enzymology buffer.

Wash with 5XSSC/0.05% Tween 20.

Prime with Imaging buffer.

Scan in 4 colors at RT.

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

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

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

Cool to 20 °C.

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

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

Prime with imaging buffer, 100 µL/channel; 60 µL/minutes

Scan in 4 colors at RT.

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

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

Example 9) Sequencing the tag

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

5'-AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCG (SEQ ID No. 58)

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

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

Step 1) Grafting primers onto surface of SFA coated chip

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

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

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

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

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

Step 2) Preparation of clusters by isothermal amplification:

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

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

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

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

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

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

To prepare the Amplification pre-mix (of sufficient volume for the first extension and 35 cycles of isothermal amplification), 35 mL of H_20 (milliQ), 7 mL of Cluster buffer (200 mM Tris, 100 mM Ammonium Sulfate, 20 mM Magnesium sulfate, 1% Triton, 1.3% DMSO, pH 8.8), and 28 mL of Amplification additive (5 M betaine solution) are mixed to achieve a final volume of 70 mL.

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

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

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

Table 1. Template hybridization and first extension.

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

Isothermal amplification at 60 °C using formamide as denaturant

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

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

Table 2. Isothermal amplification

Step	Description	T (°C)	Time (sec)	Flow rate (µl/min)	Pumped V (µl)
(1) This sequence 35 times	Pump Formamide	60	56	30	28
	Pump Amplification pre-mix	60	56	30	28
	Pump Bst mix	60	72	30	36
2	Pump wash buffer	60	280	30	140
3	Pump Storage Buffer	20	380	15	95

Wash buffer = 0.3 x SSC / 0.1% Tween

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

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

Storage Buffer = 5X SSC.

Step 3) Preparation of clusters for first sequencing read

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

surface immobilised oligonucleotides was achieved by incubation with USER enzyme mix (cocktail of Uracil DNA Glycosylase and Endonuclease VIII, NEB #M5505, 10 U/ml, 10 mM KCl, 20 mM Tris pH 8.8, 10 mM (HN4)2SO4, 2 mM MgSO4, 0.1% Triton X-100, 37 °C, 30 minutes). All exposed 3'-OH termini of DNA, either from the extended template or unextended surface oligonucleotides were blocked by dideoxy chain termination using a cocktail of terminal transferase (0.25 U/μl) and a modified polymerase (SBS polymerase as described below) (0.015 mg/ml, 100 μM ddNTP, 50 mM tris, 50 mM NaCl, 6 mM MgSO4, 1 mM EDTA, 0.05 % Tween 20). Blocking was achieved in a two stage protocol, initial incubation at 37 °C for 30 minutes followed by a ramping to 60 °C and incubating the flowcell for a further 15 minutes). Linearised and blocked clusters were washed with 0.3X SSC and storage buffer prior to denaturation with 0.1N NaOH. Denatured clusters were neutralised with TE buffer (10 mM Tris pH 8.0, 1 mM EDTA) and washed with storage buffer. The read 1 specific sequencing primer

(5'- ACACTCTTTCCCTACACGACGCTCTTCCGATC -3'(SEQ ID No. 61), 0.5 μM in hybridisation buffer) was annealed to the clusters by incubation at 60 °C for 15 minutes, followed by a 0.3X SSC wash at 40 °C (ramp rate 1 °C / sec). The flowcell was finally flushed with storage buffer (at 20 °C). Processed flowcells were transferred to the Illumina Genome Analyser for sequencing read 1.

Step 4) Sequencing from the target fragment

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

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

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

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

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

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

Cool to 20 °C.

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

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

Prime with imaging buffer, 100 µL/channel; 60 µL/minutes

Scan in 4 colors at RT.

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

Heat to 60 °C.

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

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

Cool to 20 °C.

Wash with Enzymology buffer.

Wash with 5XSSC/0.05% Tween 20.

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

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

Step 5) Sequencing of the Tag

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

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

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

Following the successful completion sequencing of two reads on the Genome Analyser, flowcells remained mounted and were prepared for paired read 2 *in situ*, using the Illumina Paired End module. Temperature control was achieved by using the Genome Analyser peltier. All flow rates were 60 µl/min and 75 µl per lane unless otherwise stated. Clusters were denatured with 0.1 M NaOH to remove the extended sequencing primer from the tag sequencing. Clusters were 3'-dephosphorylated using T4 polynucleotide kinase (Invitrogen #18004-010, 200 U/ml, 50 mM imidazole pH 6.4, 12 mM MgCl2, 70 µM ADP, 1 mM 2-mercaptoethanol, 37 °C, 30 minutes), prior to re-synthesis of the A-strand achieved using 15 cycles of 60 °C isothermal amplification (same reagents and conditions as described for cluster creation except conducted at 30 µl/min). Clusters were washed before and after resynthesis with 0.3X SSC (150 µl and 245 µl respectively). Linearisation of the B-strand of the re-synthesised clusters was achieved by the excision of 8-oxoguanine from the B-type oligo using Fpg (formamidopyrimidine DNA glycosylase, NEB #M0240, 80 U/ml, 10 mM Bis Tris propane pH 7.0, 10 mM MgCl2, 1mM dithiothreitol, 37 °C, 30 minutes). Blocking was performed as described for read 1 using the same blocking cocktail. Linearised and blocked clusters were denatured prior to hybridisation of the read 2 specific sequencing primer

(5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC-3') (SEQ ID No. 63), 0.5 µM in hybridisation buffer) as described for read 1.

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

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

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

Claims:

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

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

(b) ligating a universal adaptor to the ends of each target fragment of each sample to generate adaptor-target-adaptors of each sample, wherein each of said adaptor-target-adaptors comprises a target fragment flanked by universal adaptor sequences and said universal adaptor comprises a region of double stranded nucleic acid and at least one region of single stranded nucleic acid;

(c) amplifying adaptor-target-adaptors of each sample with two or more sample specific amplification primers to generate amplified nucleic acids, wherein one of said amplification primers comprises a sample specific tag sequence, and wherein amplified nucleic acids of each sample comprise said sample specific tag sequence and said sample specific tag sequence differentiates amplified nucleic acids originating from different samples;

(d) pooling the amplified nucleic acids of different samples;

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

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

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

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

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

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

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

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

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

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

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

(b) ligating a sample specific tagged adaptor to the ends of each target fragment of each sample to generate adaptor-target-adaptors of each sample, wherein each of said adaptor-target-adaptors comprises a target fragment flanked by tagged adaptor sequences and said tagged adaptor comprises a region of double stranded nucleic acid and at least one region of single stranded nucleic acid and a sample specific tag that differentiates adaptor-target-adaptors originating from different samples, and wherein the sample specific tag sequence is attached directly to the target fragment in each adaptor-target-adaptor with no intervening nucleotides between the sample specific tag sequence and target fragment sequences;

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

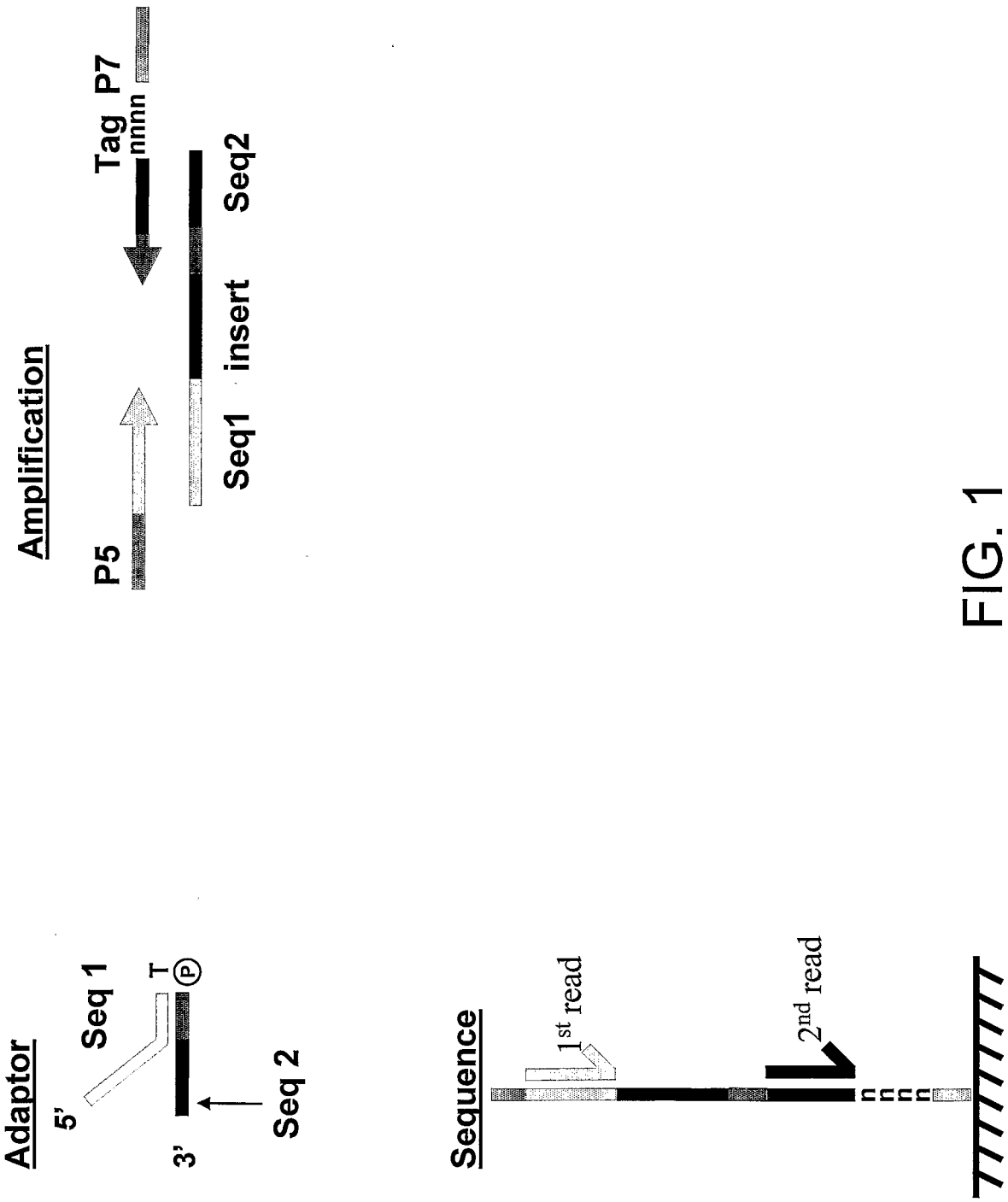

Amplification
P5
Tag P7
nnnn
Seq1
insert
Seq2
Adaptor
5'
Seq 1
T
P
3'
Seq 2
Sequence
1st read
2nd read
n
n
n
n

FIG. 1

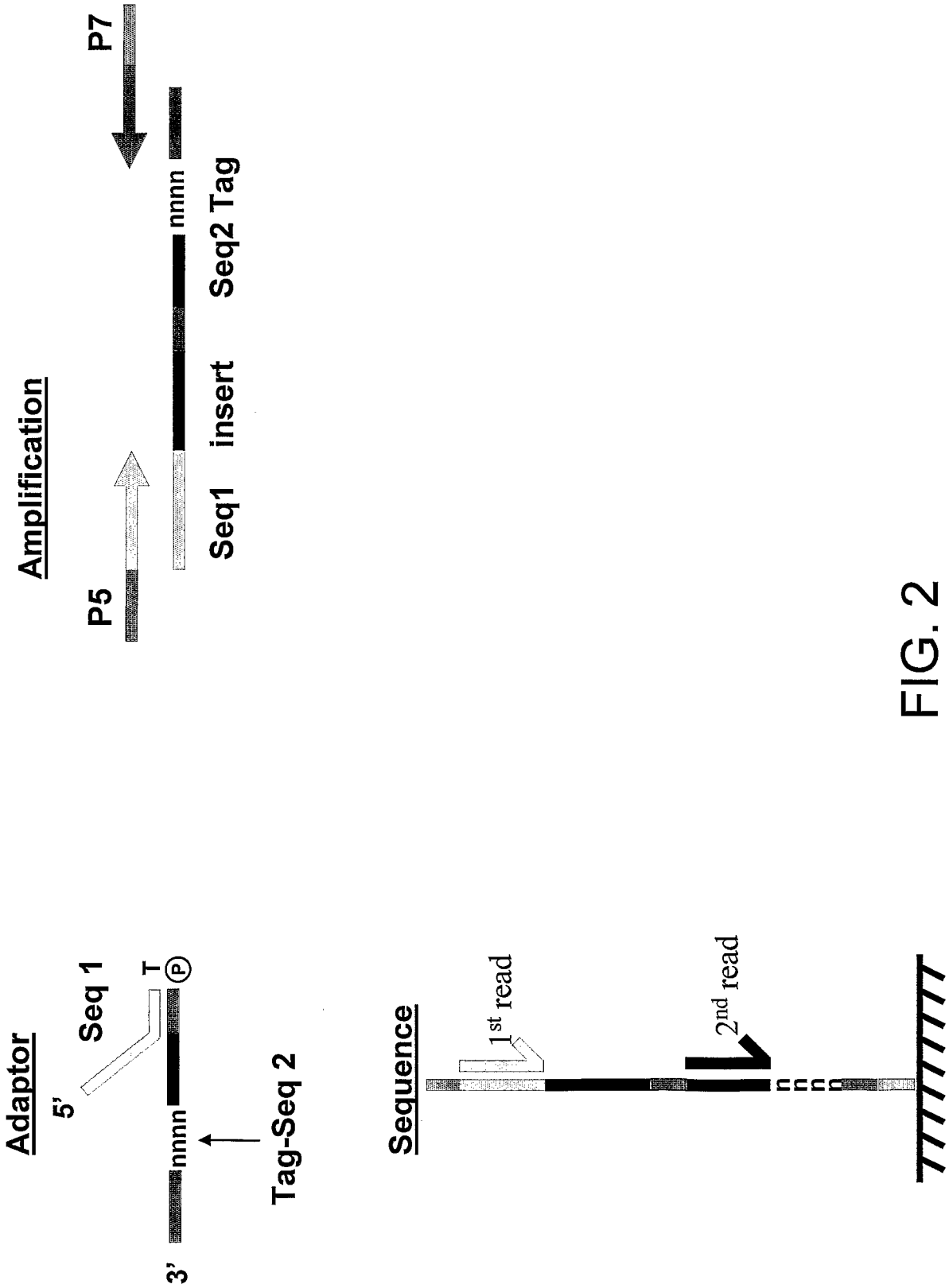

Amplification
P7
P5
nnnn
Seq1 insert Seq2 Tag
Adaptor
5'
Seq 1
T
P
nnnn
3'
Tag-Seq 2
Sequence
1st read
2nd read
n
n
n
n

FIG. 2

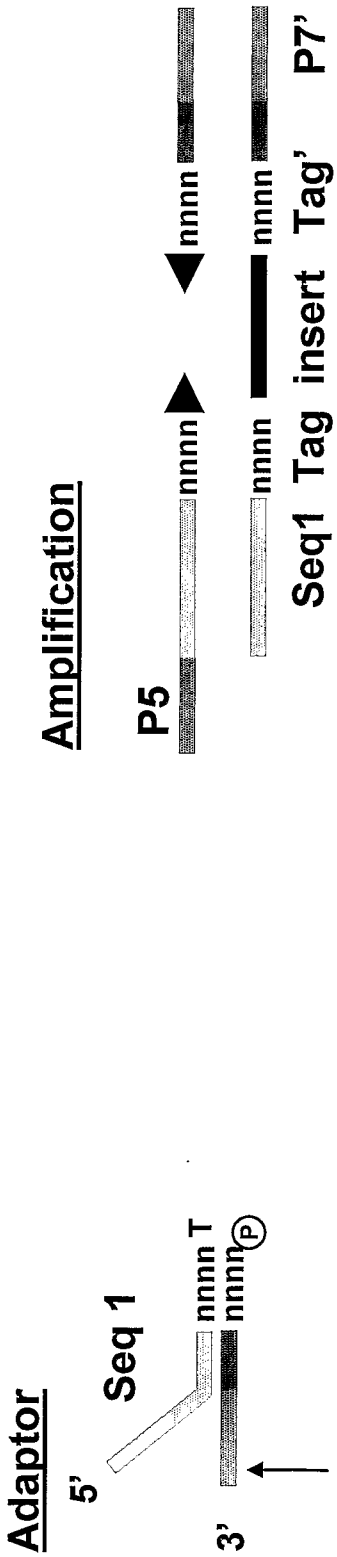

Amplification
P5
nnnn
nnnn
nnnn
Seq1 Tag insert Tag' P7'
Adaptor
5'
Seq 1
nnnn T
nnnn P
3'
P7rev
Sequence
1 read
nnnn
nnnn

FIG. 3

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

Reference Sequences	% of Clusters Aligning to Reference
PhiX	68
BAC	29
E.coli	3

	PhiX TAG	BAC TAG	Other TAGs
PhiX	98 %	0.1 %	2 %
BAC	0.1 %	98 %	2 %

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

FIG. 4

IDT Exhibit 1020.1995

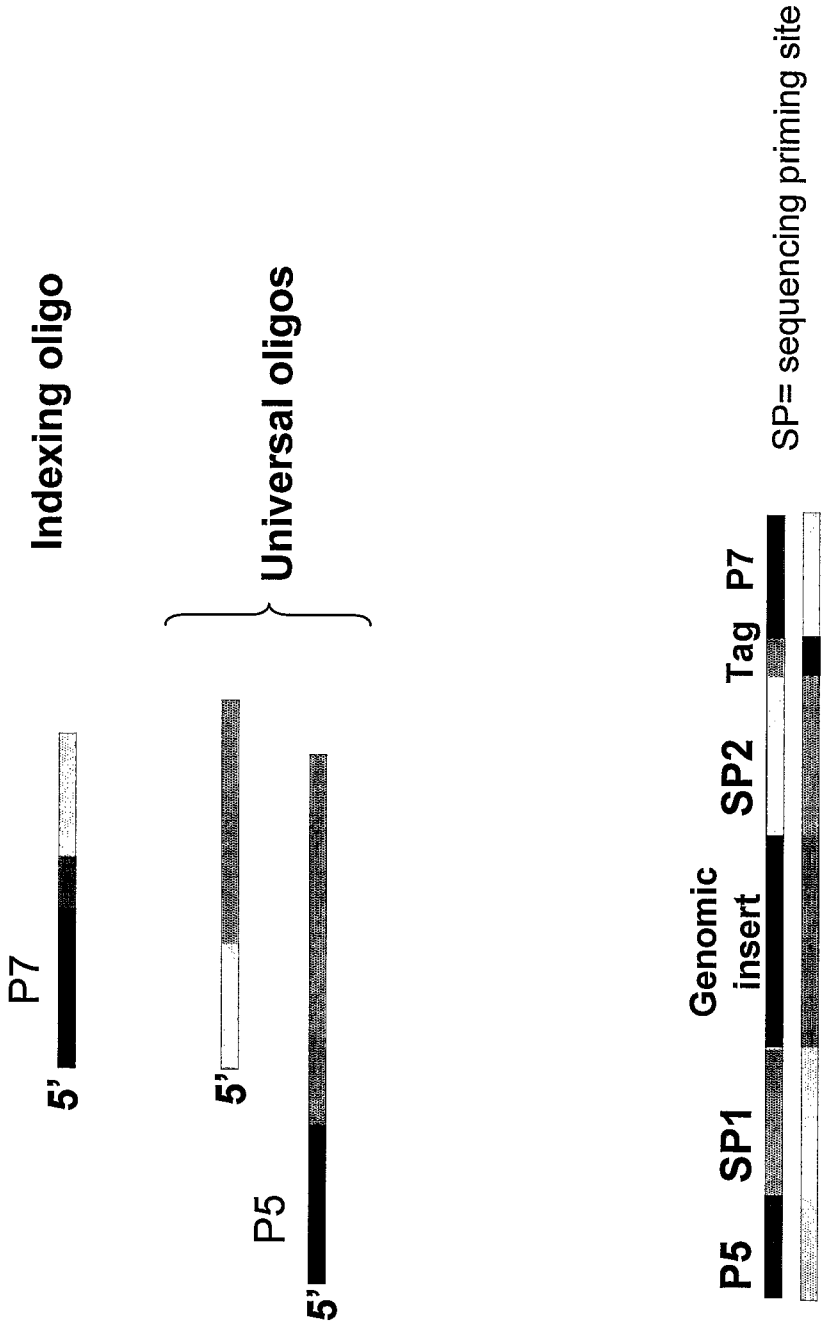

Indexing oligo
Universal oligos
P7
5'
5'
P5
5'
SP= sequencing priming site
P5
SP1
Genomic insert
SP2
Tag
P7

FIG. 5

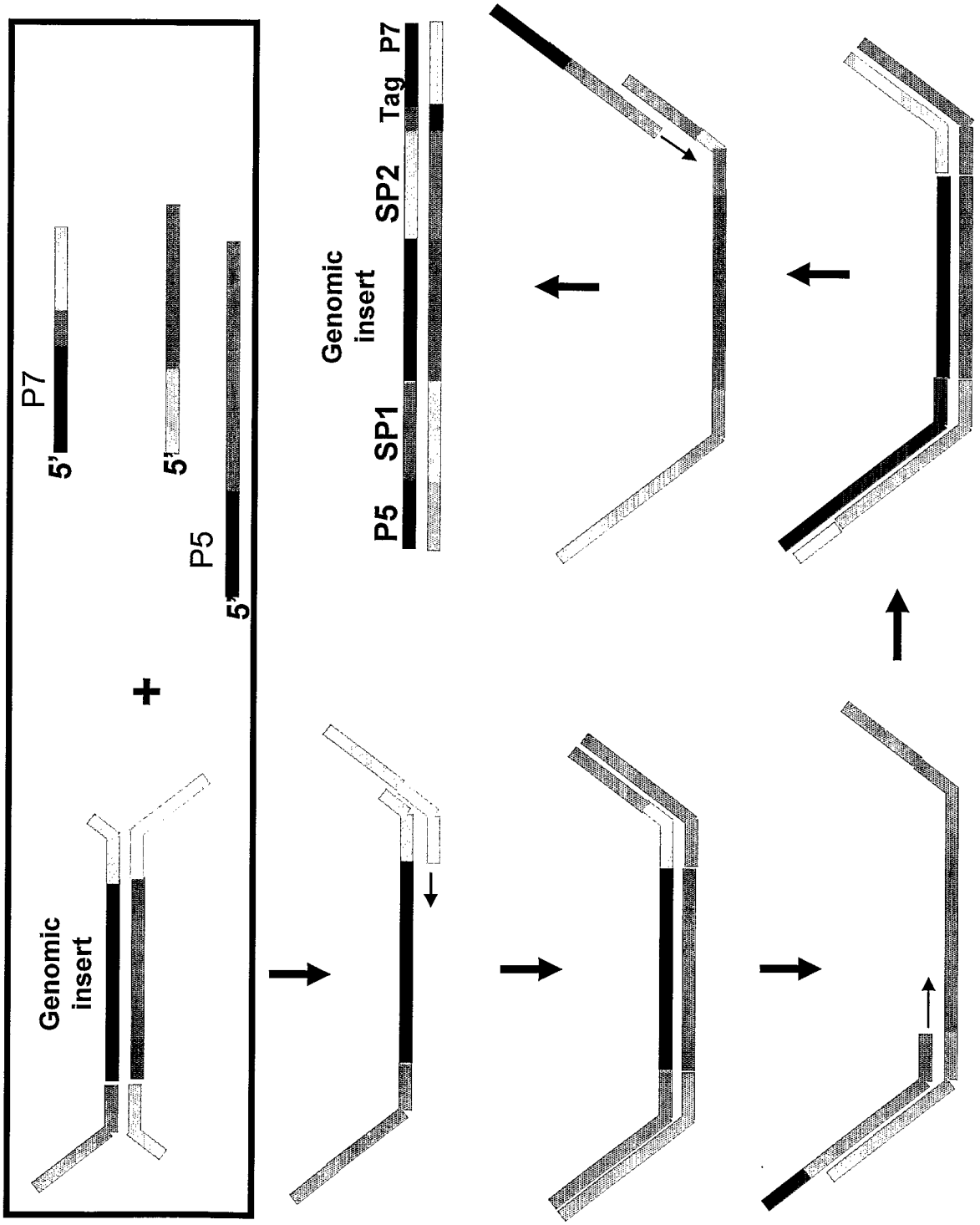


FIG. 6a

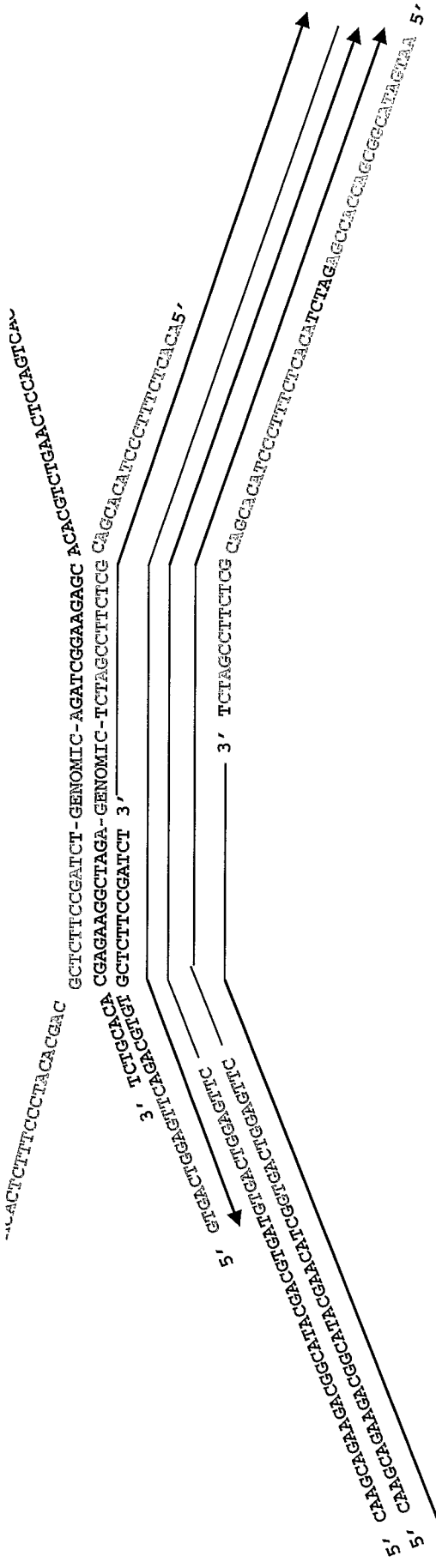

GCTCTTCCGATCT-GENOMIC-AGATCGGAAGAGC ACACGTCTGAACTCCAGTCAC
CGAGAAGGCTAGA-GENOMIC-TCTAGCCTTCTCG CAGCACATCCCTTTCTCACA 5'
3' TCTGCACA GCTCTTCCGATCT 3'
GTGACTGGAGTTCAGACGTGT
5' GTGACTGGAGTTC
3' TCTAGCCTTCTCG CAGCACATCCCTTTCTCACATCTAGAGCCACCAGCGCCATAGTAA 5'
5' CAAGCAGAAGACGGCATACGAGCGTGATGTGACTGGAGTTC
5' CAAGCAGAAGACGGCATACGAGCATCGGTGACTGGAGTTC

FIG. 6b

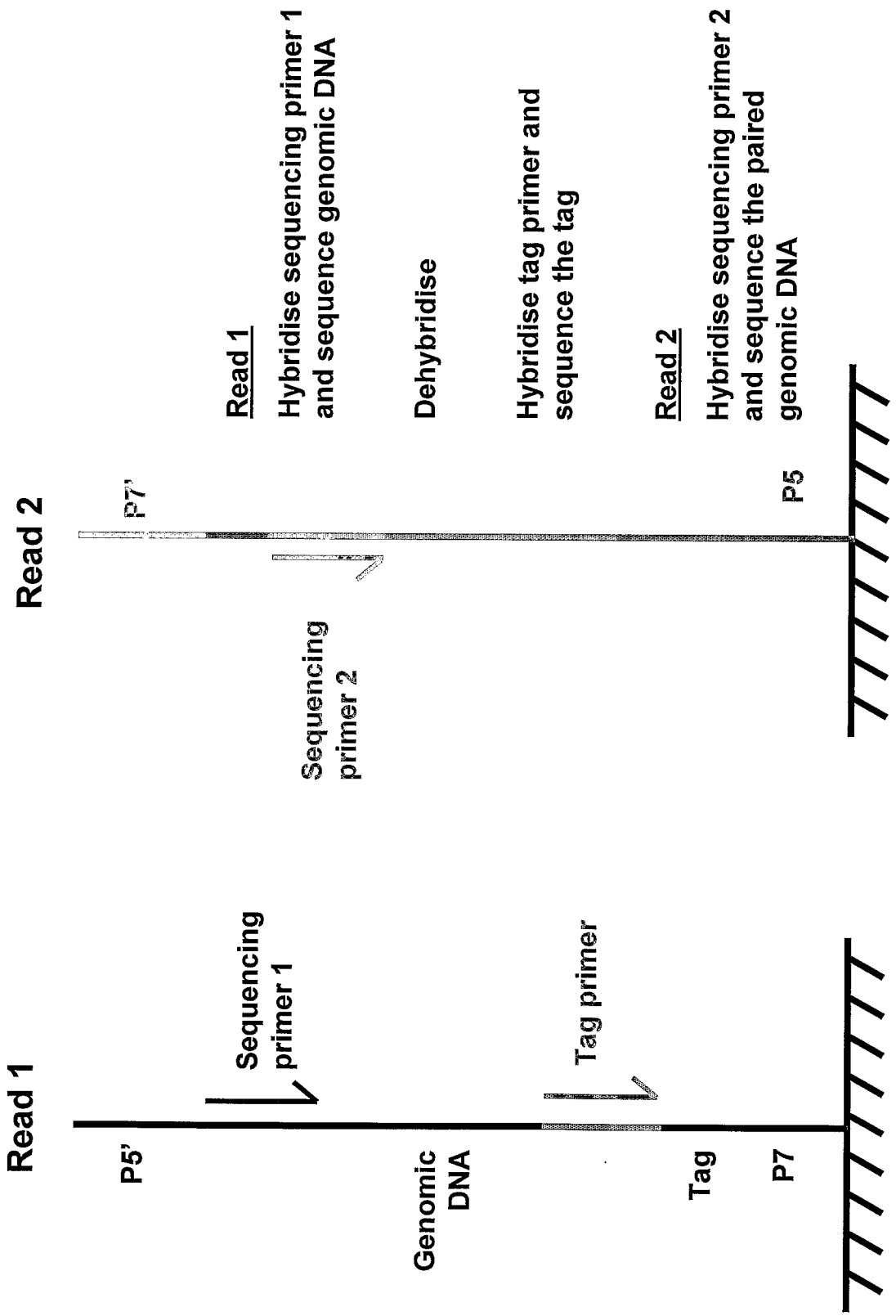

Read 1
P5'
Sequencing primer 1
Genomic DNA
Tag primer
Tag
P7
Read 2
P7'
Sequencing primer 2
Read 1
Hybridise sequencing primer 1 and sequence genomic DNA
Dehybridise
Hybridise tag primer and sequence the tag
Read 2
Hybridise sequencing primer 2 and sequence the paired genomic DNA
P5

FIG. 7

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

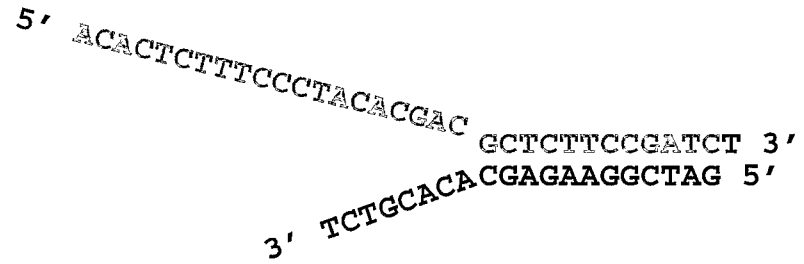
Expected Tag	Observed Tag	Genome	Clus No	BAC R1	BAC R2	PhiX R1	PhiX R2
TGACCA	TGACAA	PhiX	8	0%	0%	100%	88%
GCCAAT	TCCAAT	BAC	34	97%	94%	0%	0%
GATCAG	TATCAG	BAC	28	100%	96%	0%	0%

FIG. 8

Forked Adaptor Primers SBS3/SBS12 (indexing/paired ends)

Adaptor oligo 2 5′ ACACTCTTTCCCTACACGACGCTCTTCCGATCT 3′

Adaptor oligo 1 (short) 5′ P-GATCGGAAGAGCACACGTCT 3′



PCR Primer

Universal LPX primer	5′	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCxT 3′
Lpx primer	5′	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCxT
Tag1 primer (CGTGAT)	5′	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTC 3′
Tag2 primer (ACATCG)	5′	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTC 3′
Tag3 primer (GCCTAA)	5′	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTC 3′
Tag4 primer (TGGTCA)	5′	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTC 3′
Tag5 primer (CACTGT)	5′	CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTC 3′
Tag6 primer (ATTGGC)	5′	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTC 3′
Tag7 primer (GATCTG)	5′	CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTC 3′
Tag8 primer (TCAAGT)	5′	CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTC 3′
Tag9 primer (CTGATC)	5′	CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTC 3′
Tag10 primer (AAGCTA)	5′	CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTC 3′
Tag11 primer (GTAGCC)	5′	CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTC 3′
Tag12 primer (TACAAG)	5′	CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTC 3′

Sequencing primers

SBS12	5′	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC 3′
SBS12rev	5′	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC 3′
SBS3	5′	ACACTCTTTCCCTACACGACGCTCTTCCGATC 3′

FIG. 9

IDT Exhibit 1020.2001

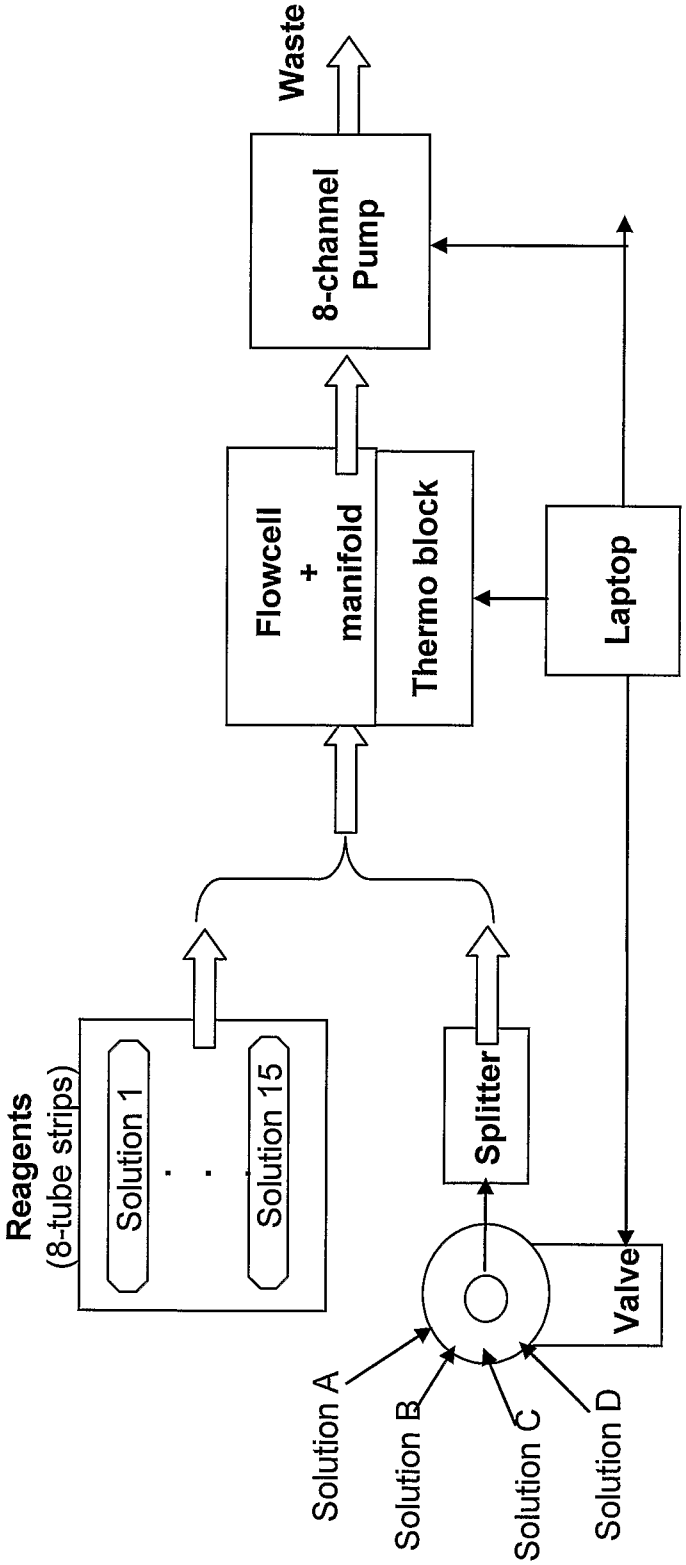

Waste
8-channel Pump
Flowcell + manifold
Thermo block
Laptop
Reagents (8-tube strips)
Solution 1
Solution 15
Splitter
Valve
Solution A
Solution B
Solution C
Solution D

FIG. 10

Espacenet

Bibliographic data: WO2010030683 (A1) — 2010-03-18

METHODS OF GENERATING GENE SPECIFIC LIBRARIES

Inventor(s): RAYMOND CHRISTOPHER K [US] ± (RAYMOND, CHRISTOPHER, K)

Applicant(s): ROSETTA INPHARMATICS LLC [US]; RAYMOND CHRISTOPHER K [US] ± (ROSETTA INPHARMATICS LLC, ; RAYMOND, CHRISTOPHER, K)

Classification: - international: *C12P19/34*
- cooperative: C12N15/1093; C12N15/66; C12Q1/6855; C12Q1/6869; C40B40/06; C40B40/08; C40B50/06
more

Application number: WO2009US56380 20090909 Global Dossier

Priority number (s): US20080095579P 20080909

Also published as: CN102203273 (A) EP2334802 (A1) EP2334802 (A4)

Abstract of WO2010030683 (A1)

The invention provides compositions and methods for generating a target enriched, sequencing ready library for resequencing at least one target region of interest from a nucleic acid containing sample.

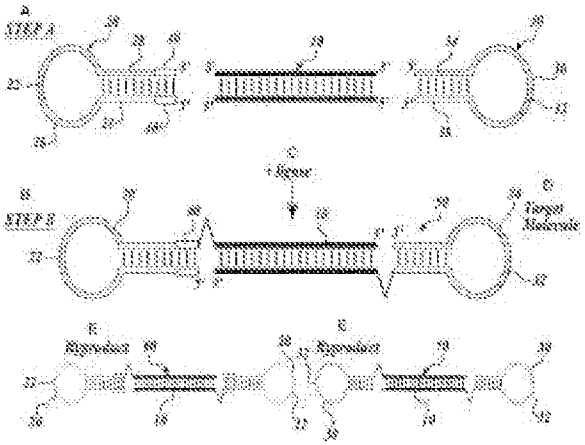


Fig.1.

IDT Exhibit 1020.2003

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 March 2010 (18.03.2010)

PCT

(10) International Publication Number
WO 2010/030683 A1

(51) International Patent Classification:
C12P 19/34 (2006.01)

(21) International Application Number:
PCT/US2009/056380

(22) International Filing Date:
9 September 2009 (09.09.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/095,579 9 September 2008 (09.09.2008) US

(71) Applicant *(for all designated States except US)*: **ROSETTA INPHARMATICS LLC** [US/US]; 401 Terry Avenue North, Seattle, WA 98109 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **RAYMOND, Christopher, K.** [US/US]; 401 Terry Avenue North, Seattle, WA 98109 (US).

(74) Agent: **QUINTON, Tineka, J.**; Christensen O'connor Johnson Kindness Pllc, 1420 Fifth Avenue, Suite 2800, Seattle, WA 98101 (US).

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

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

Published:

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

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

(54) **Title**: METHODS OF GENERATING GENE SPECIFIC LIBRARIES

(57) **Abstract**: The invention provides compositions and methods for generating a target enriched, sequencing ready library for resequencing at least one target region of interest from a nucleic acid containing sample.

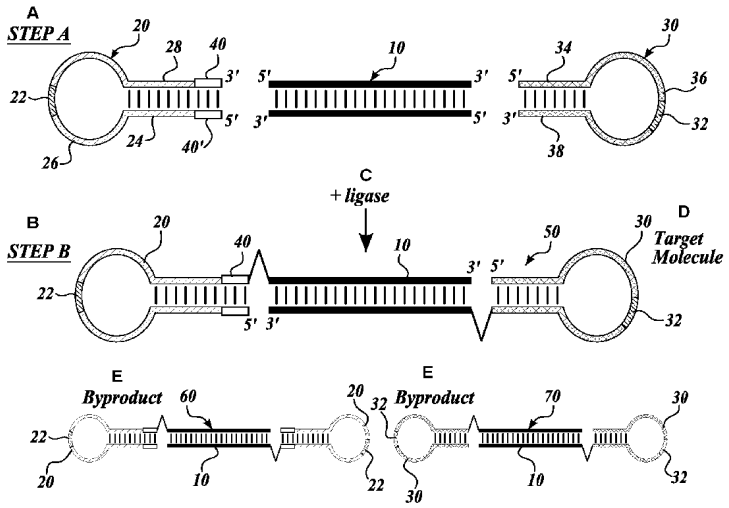


Fig.1.

WO 2010/030683 A1

METHODS OF GENERATING GENE SPECIFIC LIBRARIES

BACKGROUND

The ability to sequence deoxyribonucleic acid (DNA) accurately and rapidly is revolutionizing biology and medicine. The pharmacogenomics challenge is to comprehensively identify the genes and functional polymorphisms associated with the variability in drug response. Screens for numerous genetic markers performed for populations large enough to yield statistically significant data are needed before associations can be made between a given genotype and a particular disease.

The study of complex genomes and, in particular, the search for the genetic basis of disease in humans, requires genotyping on a massive scale, which is demanding in terms of cost, time, and labor. Such costly demands are even greater when the methodology employed involves serial analysis of individual DNA samples, i.e., separate reactions for individual samples. Resequencing of polymorphic areas in the genome that are linked to disease development will contribute greatly to the understanding of diseases such as cancer and therapeutic development. Thus, there is a need for accurate, high-throughput methods for generating nucleic acid libraries for selective resequencing of target regions of the genome and/or transcriptome for pharmacogenetics applications and genetic disease association studies.

SUMMARY

This summary is provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. This summary is not intended to identify key features of the claimed subject matter, nor is it intended to be used as an aid in determining the scope of the claimed subject matter.

In one aspect, the present invention provides a method of generating a population of DNA molecules, each DNA molecule comprising a nucleic acid insert region flanked by a first primer binding region and a second primer binding region, the method comprising (a) fragmenting a starting population of DNA molecules into a population of fragmented insert DNA molecules; (b) combining in a ligation reaction, the population of fragmented insert DNA molecules of step (a) with (i) a plurality of first stem-loop linker oligonucleotides comprising a sequence that is complementary to a first primer binding region, and (ii) a plurality of second stem-loop linker oligonucleotides comprising a sequence that is complementary to a second primer binding region; (c) contacting the

ligation reaction of step (b) with a polymerase under conditions suitable to synthesize the complementary strands corresponding to the first and second stem-loop linkers, thereby generating a plurality of double-stranded DNA molecules, each DNA molecule comprising an insert region flanked by a first primer binding region and a second primer binding region; and (d) performing a polymerase chain reaction on the double-stranded molecules of step (c) with a plurality of first PCR primers that bind to the first primer binding region and a plurality of second PCR primers that bind to the second primer binding region to selectively amplify the population of DNA molecules comprising an insert fragment flanked by a first stem-loop linker oligonucleotide and a second stem-loop linker oligonucleotide. The methods according to this aspect of the invention are useful, for example, to generate sequencing-ready libraries of DNA molecules that may be used as templates in a high-throughput sequencing platform.

In another aspect, the present invention provides a method of enriching a library for target nucleic acid regions of interest. The method according to this aspect of the invention comprises (a) contacting a library of DNA molecules comprising a subpopulation of nucleic acid target insert sequences of interest flanked by a first primer binding region and a second primer binding region within a larger population of nucleic acid insert sequences flanked by the first primer binding region and the second primer binding region with a set of capture probes, the set of capture probes comprising a plurality of capture oligonucleotides, each comprising a first target sequence-specific binding region and a second capture reagent binding region, under conditions that allow binding between the capture oligonucleotides and the nucleic acid target regions of interest, to form a plurality of complexes between target regions of interest and capture probes; (b) contacting the mixture of step (a) with a capture reagent and separating the capture reagent bound complex from the mixture; and (c) eluting the target regions of interest flanked by the first primer binding region and the second primer binding region from the capture reagent bound complex. In some embodiments, the method further comprises amplifying the eluted target regions of interest flanked by the first primer binding region and the second primer binding region with a forward PCR primer and a reverse PCR primer that bind to the first and second primer binding regions to generate a library that is enriched for target regions of interest.

In another aspect, the invention provides a method of generating a target enriched, sequencing ready library for resequencing at least one target region of interest from a

IDT Exhibit 1020.2006

nucleic acid containing sample. The method according to this aspect of the invention comprises (a) providing a library comprising fragmented nucleic acid molecules flanked by a first primer binding region and a second primer binding region; and (b) enriching the library for target sequences with a set of capture probes comprising a plurality of capture oligonucleotides, each comprising a first target-specific binding region and a second capture reagent binding region, thereby generating an enriched sequencing ready library for resequencing at least one target region of interest.

The methods of the invention can be used to create populations of nucleic acid molecules (also referred to in the art as "libraries" of nucleic acid molecules) useful for a variety of purposes, such as resequencing a target region of interest.

DESCRIPTION OF THE DRAWINGS

The foregoing aspects and many of the attendant advantages of this invention will become more readily appreciated as the same become better understood by reference to the following detailed description, when taken in conjunction with the accompanying drawings, wherein:

FIGURE 1 illustrates an embodiment of a method for generating a population of DNA molecules comprising a nucleic acid insert region flanked by a first primer binding region and a second primer binding region, as described in Example 1;

FIGURE 2A shows the densities for groups of bar codes (rows) for each amplicon of five genes (columns), demonstrating that the bar-coded population of DNA molecules comprising a nucleic acid insert region flanked by a first primer binding region and a second primer binding region generated sequence that was equivalent to non-bar-coded population of DNA molecules, as described in Example 1;

FIGURE 2B shows both the expected and observed distribution of sequencing reads, demonstrating the accurate association of bar-coded sequence results with the correct samples in accordance with an embodiment of a method for generating a population of DNA molecules comprising a nucleic acid insert region flanked by a first primer binding region and a second primer binding region, as described in Example 1;

FIGURE 3 is a flowchart showing the steps of a method of generating a sequence ready library from a starting population of DNA molecules, with the optional steps of enrichment of the library for target sequences using solution-based capture methods, in accordance with various embodiments of the methods of the invention;

IDT Exhibit 1020.2007

FIGURE 4 illustrates an embodiment of a method for enriching a population of DNA molecules comprising a nucleic acid insert region flanked by a first primer binding region and a second primer binding region for target regions of interest using capture probes comprising a capture binding region that directly binds to a capture reagent, as described in Example 3;

FIGURE 5 illustrates an embodiment of a method for enriching a population of DNA molecules for target regions of interest using capture probes comprising a capture binding region that indirectly binds to a capture reagent, as described in Example 4;

FIGURE 6 is a flow chart of the steps of solution-based capture in accordance with various embodiments of the methods of the invention;

FIGURE 7 illustrates the sequencing read depth for the exons in the exemplary gene target PIK3CA obtained from a library that was enriched using indirect solution capture with capture oligos that were complementary to these exons, demonstrating high read densities (e.g., 1,000 reads) along all of the targeted exons, as described in Example 4;

FIGURE 8 illustrates the sequencing read depth for the exons in the exemplary gene target AKT 1 gene in a 77-gene experiment, as described in Example 5;

FIGURE 9 graphically illustrates the percent of target bases sequenced at a specific sequencing read depth from a library that was enriched with three rounds of solution-based capture in accordance with an embodiment of the methods of the invention, as described in Example 5;

FIGURE 10A illustrates a read density map for determining the copy number variation of a region on a chromosome from sequence analysis of a sequence ready library generated according to an embodiment of a method of the invention, as described in Example 6;

FIGURE 10B shows the results of an experiment carried out to measure the copy number variation from a region of chromosome 14 in a normal human subject using the sequence ready library generated according to an embodiment of a method of the invention, as described in Example 6;

FIGURE 11A shows the results of transcriptional analysis of a cardiovascular risk locus on a 1500 Kb region of chromosome 9p21 containing two identified SNPs, (SNPA and SNPB) showing plus strand transcription that includes the associated SNPA and

IDT Exhibit 1020.2008

SNPB appears to span approximately 800 Kb, with the arrows showing potential transcription units, as described in Example 7; and

FIGURE 11B shows the generation of a sequencing-ready library generated from not-so-random primer amplified from the whole transcriptome and enriched for the risk associated locus encompassing SNPA and SNPB shown in FIGURE 11A with capture probes (arrows), as described in Example 7.

DETAILED DESCRIPTION

This section presents a detailed description of the many different aspects and embodiments that are representative of the inventions disclosed herein. This description is by way of several exemplary illustrations of varying detail and specificity. Other features and advantages of these embodiments are apparent from the additional descriptions provided herein, including the different examples. The provided examples illustrate different components and methodology useful in practicing various embodiments of the invention. The examples are not intended to limit the claimed invention. Based on the present disclosure, the ordinary skilled artisan can identify and employ other components and methodology useful for practicing the present invention.

I. DEFINITIONS

Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by one of ordinary skill in the art to which this invention belongs. Practitioners are particularly directed to Sambrook et al., "Molecular Cloning: A Laboratory Manual," 2d ed., Cold Spring Harbor Press, Plainsview, New York (1989); and Ausubel et al., "Current Protocols in Molecular Biology," (Supplement 47), John Wiley & Sons, New York (1999), for definitions and terms of the art.

It is contemplated that the use of the term "**about**" in the context of the present invention is to connote inherent problems with precise measurement of a specific element, characteristic, or other trait. Thus, the term "about," as used herein in the context of the claimed invention, simply refers to an amount or measurement that takes into account single or collective calibration and other standardized errors generally associated with determining that amount or measurement. For example, a concentration of "about" 100 mM of Tris can encompass an amount of 100 mM ± .5 mM, if .5 mM represents the collective error bars in arriving at that concentration. Thus, any measurement or amount referred to in this application can be used with the term "about"

IDT Exhibit 1020.2009

if that measurement or amount is susceptible to errors associated with calibration or measuring equipment, such as a scale, pipetteman, pipette, graduated cylinder, etc.

The use of the word "**a**" or "**an**" when used in conjunction with the term "comprising" in the claims and/or the specification may mean "one," but it is also consistent with the meaning of "one or more," "at least one," and "one or more than one."

As used herein, the term "**nucleic acid molecule**" encompasses both deoxyribonucleotides and ribonucleotides and refers to a polymeric form of nucleotides including two or more nucleotide monomers. The nucleotides can be naturally occurring, artificial, and/or modified nucleotides.

As used herein, an "**isolated nucleic acid**" is a nucleic acid molecule that exists in a physical form that is non-identical to any nucleic acid molecule of identical sequence as found in nature; "isolated" does not require, although it does not prohibit, that the nucleic acid so described has itself been physically removed from its native environment. For example, a nucleic acid can be said to be "isolated" when it includes nucleotides and/or internucleoside bonds not found in nature. When instead composed of natural nucleosides in phosphodiester linkage, a nucleic acid can be said to be "isolated" when it exists at a purity not found in nature, where purity can be adjudged with respect to the presence of nucleic acids of other sequences, with respect to the presence of proteins, with respect to the presence of lipids, or with respect to the presence of any other component of a biological cell, or when the nucleic acid lacks sequence that flanks an otherwise identical sequence in an organism's genome, or when the nucleic acid possesses sequence not identically present in nature. As so defined, "isolated nucleic acid" includes nucleic acids integrated into a host cell chromosome at a heterologous site, recombinant fusions of a native fragment to a heterologous sequence, recombinant vectors present as episomes, or as integrated into a host cell chromosome.

As used herein, "**subject**" refers to an organism or to a cell sample, tissue sample, or organ sample derived therefrom, including, for example, cultured cell lines, biopsy, blood sample, or fluid sample containing a cell. For example, an organism may be an animal, including but not limited to, an animal such as a cow, a pig, a mouse, a rat, a chicken, a cat, a dog, etc., and is usually a mammal, such as a human.

As used herein, the term "**specifically bind**" refers to two components (e.g., target-specific binding region and target) that are bound (e.g., hybridized, annealed, complexed) to one another sufficiently that the intended capture and enrichment steps can

IDT Exhibit 1020.2010

be conducted. As used herein, the term "specific" refers to the selective binding of two components (e.g., target-specific binding region and target) and not generally to other components unintended for binding to the subject components.

As used herein, the term "**high stringency hybridization conditions**" means any condition in which hybridization will occur when there is at least 95%, preferably about 97% to 100%, nucleotide complementarity (identity) between the nucleic acid sequences of the nucleic acid molecule and its binding partner. However, depending upon the desired purpose, the hybridization conditions may be "medium stringency hybridization," which can be selected that require less complementarity, such as from about 50% to about 90%, (e.g., 60%, 70%, 80%, 85%). The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm of Karlin and Altschul (*Proc. Natl. Acad. Sci. USA 87*:2264-2268 (1990)), modified as in Karlin and Altschul (*Proc. Natl. Acad. Sci. USA 90*:5873-5877 (1993)). Such an algorithm is incorporated into the NBLAST and XBLAST programs of Altschul et al. (*J. Mol. Biol. 215*:403-410 (1990)).

As used herein, the term "**complementary**" refers to nucleic acid sequences that are capable of base-pairing according to the standard Watson-Crick complementary rules. That is, the larger purines will base pair with the smaller pyrimidines to form combinations of guanine paired with cytosine (G:C) and adenine paired with either thymine (A:T) in the case of DNA, or adenine paired with uracil (A:U) in the case of RNA.

As used herein, the term "**target**" refers to a nucleic acid molecule or polynucleotide whose presence and/or amount and/or sequence is desired to be determined and that has an affinity for a given target capture probe. Examples of targets include regions of genomic DNA, PCR amplified products derived from RNA or DNA, DNA derived from RNA or DNA, ESTs, cDNA, and mutations, variants, or modifications thereof.

As used herein, the term "**resequencing**" refers to a technique that determines the sequence of a genome of an organism using a reference sequence that has already been determined. It should be understood that resequencing may be performed on both the entire genome/transcriptome of an organism or a portion of the genome/transcriptome large enough to include the genetic change of the organism as a result of selection. Resequencing may be carried out using various sequencing methods, such as any

IDT Exhibit 1020.2011

sequencing platform amenable to producing DNA sequencing reads that can be aligned back to a reference genome, and is typically based on highly parallel technologies such as, for example, dideoxy "Sanger" sequencing, pyrosequencing on beads (e.g., as described in U.S. Patent No. 7,211,390, assigned to 454 Life Sciences Corporation, Brandord, Connecticut), ligation based sequencing on beads (e.g., Applied Biosystems Inc,/Invitrogen), sequencing on glass slides (e.g., Illumina Genome Analyzer System, based on technology described in WO 98/44151 (Mayer, P., and Farinelli L.), microarrays or fluorescently labeled micro-beads.

II. ASPECTS AND EMBODIMENTS OF THE INVENTION

In accordance with the foregoing, in one aspect, the invention provides a method of generating a population of DNA molecules (i.e., a library) that may be used for resequencing analysis. Each DNA molecule in the population of DNA molecules comprises a nucleic acid insert region flanked by a first primer binding region and a second primer binding region. The method comprises (a) fragmenting a starting population of DNA molecules into a population of fragmented insert DNA molecules; and (b) combining in a ligation reaction, the population of fragmented insert DNA molecules of step (a) with (i) a plurality of first stem-loop linker oligonucleotides comprising a sequence that is complementary to a first primer binding region, and (ii) a plurality of second stem-loop linker oligonucleotides comprising a sequence that is complementary to a second primer binding region; (c) contacting the ligation reaction of step (b) with a polymerase under conditions suitable to synthesize the complementary strands corresponding to the first and second stem-loop linkers, thereby generating a plurality of double-stranded DNA molecules, each DNA molecule comprising an insert region flanked by a first primer binding region and a second primer binding region; and (d) performing a polymerase chain reaction on the double-stranded molecules of step (c) with a plurality of first PCR primers that bind to the first primer binding region and a plurality of second PCR primers that bind to the second primer binding region to selectively amplify the population of DNA molecules comprising an insert fragment flanked by a first stem-loop linker oligonucleotide and a second stem-loop linker oligonucleotide.

The methods of this aspect of the invention can be used to generate a library suitable for genomic or transcriptome analysis such as, for example, resequencing analysis of the fragmented inserts.

IDT Exhibit 1020.2012

FIGURE 1, step D (PCR products) illustrates exemplary DNA molecules 50A, 50B generated according to the methods of this aspect of the invention comprising an insert fragment 10 flanked by a first stem-loop linker oligonucleotide 20 and a second stem-loop linker oligonucleotide 30.

FIGURE 3 illustrates an exemplary embodiment of the method of generating a sequencing-ready library 600 comprising a plurality of DNA molecules 50A, 50B according to this aspect of the invention. As shown in FIGURE 3, at step 610, a starting population of DNA molecules containing one or more target sequence(s) of interest is fragmented. At step 620, a plurality of first stem-loop linker oligonucleotides, each comprising a sequence that is complementary to a first primer binding region, and a plurality of second stem-loop linker oligonucleotides, each comprising a sequence that is complementary to a second primer binding region, are ligated to the ends of the DNA fragments (inserts). At step 630, the ligation mixture is filled in and PCR amplified with primers that bind to the first and second primer binding regions to generate a population of double-stranded DNA molecules, each DNA molecule comprising an insert region flanked by a first primer binding region and a second primer binding region (i.e., a library). At step 640, the library can be optionally sequenced or may be enriched for the target sequences of interest according to steps 650-670 shown in FIGURE 3, FIGURE 6, and further described herein.

STARTING POPULATIONS OF NUCLEIC ACID MOLECULES

Examples of starting populations of nucleic acid molecules containing one or more target sequence(s) of interest for use in the methods of this aspect of the invention include genomic DNA, mRNA, tRNA, rRNA, cRNA, oligonucleotides, DNA derived from RNA or DNA, ESTs, cDNA, cDNA generated from not-so-random primed total RNA (e.g., as described in Example 7), PCR amplified products derived from RNA or DNA, microRNA, shRNA, siRNA, and mutations, variants, or modifications thereof.

The starting nucleic acid molecules may be isolated from a subject, such as a cell sample, tissue sample, or organ sample derived therefrom, including, for example, cultured cell lines, biopsy, blood sample, or fluid sample containing a cell. The subject may be an animal, including but not limited to, an animal such as a cow, a pig, a mouse, a rat, a chicken, a cat, a dog, etc., and is usually a mammal, such as a human.

As used herein, the term "**target nucleotide**" refers to a nucleic acid molecule or polynucleotide in a starting population of nucleic acid molecules having a target sequence

IDT Exhibit 1020.2013

whose presence and/or amount and/or nucleotide sequence is desired to be determined and that has an affinity for a given target capture probe.

As used herein, the term "**target sequence**" refers generally to a nucleic acid sequence on a single strand of nucleic acid. The target sequence may be a portion of a gene, a regulatory sequence, genomic DNA, cDNA, RNA including mRNA and rRNA, or others. The target sequence may be a target sequence from a sample or a secondary target such as a product of an amplification reaction.

In some embodiments, the starting population of nucleic acid molecules comprises PCR products amplified from a plurality of target-specific amplicons from a nucleic acid containing sample, as described in Example 1. In other embodiments, the starting population of nucleic acid molecules comprises total genomic DNA, as described in Example 2. In some embodiments, the starting population of nucleic acid molecules represents the whole transcriptome, as described in Example 7.

The starting population of nucleic acid molecules is fragmented into a population of fragmented insert DNA molecules of one or more specific size range(s). In one embodiment, for mammalian-sized genomes, fragments are generated from at least about 1 genome-equivalent of starting DNA, such as at least about 10 genome-equivalents of DNA, such as at least about 100 genome-equivalents of DNA, such as at least about 1,000 genome-equivalents of DNA, such as at least about 10,000 genome-equivalents of DNA, such as at least about 100,000 genome-equivalents of DNA, such as at least about 300,000 genome-equivalents of DNA.

This fragmentation may be accomplished by methods known in the art, including chemical, enzymatic, and mechanical fragmentation. In one embodiment, the fragments are from about 10 to about 10,000 nucleotides in length. In another embodiment, the fragments are from about 50 to about 2,000 nucleotides in length. In another embodiment, the fragments are from about 10-1,000, 10-800, 10-500, 50-500, 50-250, 50-150 nucleotides in length. In another embodiment, the fragments are less than 500 nucleotides in length, such as less than 400 nucleotides, less than 300 nucleotides, less than 200 nucleotides, or less than 150 nucleotides in length. In one embodiment, the fragmentation is accomplished mechanically through the use of sonication. In one embodiment, the fragmentation is accomplished by digestion with DNase I, which induces random double-stranded breaks in DNA in the absence of Mg^{++} and in the presence of Mn^{++}, as described in Example 1. In some embodiments, the method may

IDT Exhibit 1020.2014

include the step of size selecting the fragments via standard methods such as column purification or isolation from an agarose gel.

In some embodiments, the fragmented DNA molecules are blunt-end polished prior to ligation to the stem-loop linkers. The blunt-end polishing step may be accomplished by incubation with a suitable enzyme, such as T4 polymerase (which has both 3' to 5' exonuclease activity and 5' to 3' polymerase activity). The fragmented DNA molecules may be optionally phosphorylated, for example, using T4 polynucleotide kinase, prior to ligation to the stem-loop linkers.

STEM-LOOP OLIGONUCLEOTIDE LINKERS

As shown in FIGURE 1, step A, the first stem-loop linker oligonucleotide 20 comprises a 5' region 24 with a sequence that is complementary to a sequence located in the 3' region 28 that forms a stem structure, and an intervening region 26 between the 5' and 3' region that forms a loop structure. Also located in the first stem-loop linker oligonucleotide 20 is a sequence 22 that is complementary to a first primer binding region 82 that may be positioned in the intervening region 26 or in the stem region. Under non-denaturing conditions, the 5' region 24 and 3' region 28 hybridize together, resulting in the stem-loop linker oligonucleotide 20 structure with a double-stranded stem 24 and 28 with an intervening region 26 that forms a loop structure.

Similarly, as further shown in FIGURE 1 at step A, the second stem-loop linker oligonucleotide 30 comprises a 5' region 34 having a sequence that is complementary to a sequence located in the 3' region 38 that forms a stem structure, and an intervening region 36 between the 5' and 3' region that forms a loop structure. Also located in the second stem-loop linker oligonucleotide 30 is a sequence 32 that is complementary to a second primer binding region 92 that may be positioned in the intervening region 36 or in the stem region. Under non-denaturing conditions, the 5' region 34 and 3' region 38 hybridize together, resulting in the stem-loop linker oligonucleotide 30 structure with a double-stranded stem 34 and 38 with an intervening region 36 that forms a loop structure.

The length of each stem-loop linker 20, 30 is typically at least 40 nucleotides, such as at least 45 nucleotides, at least 50 nucleotides, at least 55 nucleotides, at least 60 nucleotides, at least 65 nucleotides, at least 70 nucleotides, up to a maximum length of about 200 nucleotides. In some embodiments of the methods, the stem-loop linkers are each from about 45 nucleotides to about 70 nucleotides in length.

IDT Exhibit 1020.2015

The 5' complementary region 24 and the 3' complementary region 28 in the first stem-loop linker 20, and the 5' complementary regions 34 and the 3' complementary region 38 in the second stem-loop linker 30, can be from about 5 nucleotides to 100 nucleotide or greater, such as 10 nucleotides, 15 nucleotides, 20 nucleotides or more in length, and may be designed using a variety of different sequences that result in hybridization between the complementary regions on each stem-loop linker, resulting in a local region of double-stranded DNA (i.e., a stem). For example, stem sequences may 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 ~45°C.

The intervening loop regions 26, 36 in the first and second stem-loop linkers can be from about 10 nucleotides in length, 20 nucleotides, 30 nucleotides, 40 nucleotides or more in length. In order to facilitate subsequent PCR amplification and sequencing, in some embodiments the intervening loop region 26, 36 includes a nucleic acid sequence 22, 32 ranging in size from about 10 nucleotides to about 30 nucleotides that is complementary to a first and second PCR primer binding sequences 82, 92. The regions complementary to a first and second primer binding sequence may be contained within any other part of the stem-loop linker.

The first 82 and second 92 PCR primer binding regions contain sequences that are distinct from one another and designed for providing a universal first primer binding site and a universal second primer binding site in the plurality of DNA molecules in a sequence-ready library, for binding to a first and second PCR primer to enable PCR amplification of an intervening insert sequence.

In some embodiments, the stem-loop linker oligonucleotides further comprise one or more additional features such as a restriction enzyme site and/or an anchor probe binding site for attachment to a sequencing platform, such as a flow cell for massive parallel sequencing (e.g., Illumina, Inc.). For example, the Illumina Genome Analyzer System is based on technology described in WO 98/44151, hereby incorporated by reference, wherein DNA molecules are bound to a sequencing platform (flow cell) via an anchor probe binding site (otherwise referred to as a flow cell binding site) and amplified *in situ* on a glass slide. The DNA molecules are then annealed to a sequencing primer and sequenced in parallel base-by-base using a reversible terminator approach. The Illumina Genome Analyzer System utilizes flow-cells with 8 channels, generating

IDT Exhibit 1020.2016

sequencing reads of 18 to 36 bases in length, generating >1.3 Gbp of high quality data per run (see http://www.illumina.com).

In some embodiments, the first 20 and second 30 stem-loop linkers, each contains an anchor probe binding site for binding to a sequencing platform (e.g., a flow cell as described above). In some embodiments, the first 82 and second 92 PCR primer binding sites comprise a sequence that is also used as an anchor probe binding site for binding to a sequencing platform. In some embodiments, at least one of the first 20 or second 30 stem-loop linker oligonucleotides further comprises a sequence for annealing to a sequencing primer. In some embodiments, the first 20 stem-loop linker oligonucleotide comprises a sequence for annealing to a sequencing primer.

STEM-LOOP LINKER OLIGONUCLEOTIDES COMPRISING MOLECULAR BAR CODES

In some embodiments, at least one of the stem-loop linker oligonucleotides (e.g., either 20 or 30) further comprises one or more molecular bar code sequences (e.g., a nucleotide tag with a length of 1, 2, 3, 4 or more nucleotides) that can be utilized to identify the origin of insert sequences 10 in mixtures of bar-coded samples. In some embodiments, the molecular bar code sequences are used to create groups of polynucleotides that share a common feature. For example, such features can include the source/sample of origin, the processing conditions used to generate the polynucleotide, etc., as further described in Example 1.

LIGATION OF STEM-LOOP LINKERS TO INSERT FRAGMENTS

In accordance with the methods of this aspect of the invention, the double-stranded nucleic acid fragments 10 are combined with the first 20 and second 30 stem-loop linker oligonucleotides in a ligation reaction with a suitable enzyme, such as T4 DNA ligase. As shown in FIGURE 1, Step A, the stem region of each stem-loop linker 20, 30 forms a blunt-ended, double-stranded DNA segment suitable for ligation to the blunt-ended, double-stranded nucleic acid fragments 10, resulting in a ligated structure having the 3' end of a stem-loop linker 20 or 30 covalently joined to the 5' end of the double-stranded DNA insert 10. A pre-PCR fill-in reaction with a suitable polymerase, such as Taq polymerase, is used to copy the sequence information from the ligated insert:stem-loop linker to the complementary strand, resulting in the fill-in ligation products shown in FIGURE 1, step C.

IDT Exhibit 1020.2017

As shown in FIGURE 1, Step C, the ligation reaction results in a mixture of ligation products including the target ligation products comprising inserts 10 flanked on each end by a pair of heterogeneous stem-loop linkers 20, 30 in a first orientation 50A and a second orientation 50B, as well as ligation byproducts comprising inserts 10 flanked on each end by a pair of homogenous stem-loop linkers 20, 20 shown as ligation byproducts 60 or 30, 30, shown as ligation byproducts 70.

SUPPRESSION PCR TO SELECTIVELY AMPLIFY TARGET LIGATION PRODUCTS

As shown in FIGURE 1, Step C, the initial population of ligation products includes a mixture of inserts flanked by heterogeneous linker ends 50A, 50B and inserts flanked by homogenous linker ends 60, 70. A phenomenon referred to as suppression PCR (P.D. Siebert et al., *Nucleic Acids Res. 23*:1087-1088 (1995)) is used to selectively enrich for the inserts flanked by heterogeneous linker ends 50A, 50B. As demonstrated in Example 1, it is difficult to amplify an extended stem-loop structure (e.g., greater than 40 nucleotides) because the double-stranded stem occludes the binding of PCR primers. Accordingly, as shown in FIGURE 1, Step D, the unwanted ligation byproducts 60, 70 are refractive to PCR amplification because the first stem-loop linker oligonucleotide and second stem-loop linker oligonucleotide are greater than 40 nucleotides. Therefore, as shown in FIGURE 3, step 630, the ligation mixture is amplified in a polymerase chain reaction (PCR) with a first PCR primer 52 that hybridizes to the first PCR primer binding site 82 and a second PCR primer 54 that hybridizes to the second PCR primer binding site 92 to generate a sequencing-ready library comprising a plurality of nucleic acid molecules 50A, 50B containing a plurality of inserts that are derived from the starting population of DNA molecules (as shown in FIGURE 1, step D "PCR Products").

Polymerase chain reaction (PCR) is a technique that is well known and involves the use of primer extension combined with thermal cycling to amplify a target sequence. In general, the greater the number of amplification cycles during the polymerase chain reaction, the greater the amount of amplified DNA product is obtained. In some embodiments, a desirable number of amplification cycles for use in the suppression PCR amplification (see FIGURE 3) step 630 is from 2 to 60 cycles, such as from 10 to 30 cycles, such as about 20 cycles.

The resulting amplification product comprises a library of a plurality of double-stranded nucleic acid molecules 50A, 50B, each comprising a nucleic acid insert

IDT Exhibit 1020.2018

region flanked by a first primer binding region and a second primer binding region. The plurality of nucleic acid insert regions in the library includes one or more target sequences and can include enough different nucleic acid sequences to cover (i.e., represent) part or all of a source nucleic acid including, without limitation, a genome of an organism, a genomic locus, a cDNA library, a whole transcriptome of an organism, and the like. For example, such a library of double-stranded nucleic acid molecules may cover at least about 50%, or at least about 60%, or at least about 70%, or at least about 80%, or at least about 90%, or at least about 95% up to about 100% of the source nucleic acid.

Such libraries generated according to the methods of the invention may be applied directly to a flow cell sequencing platform, such as an Illumina Genome Analyzer, for sequence analysis or sequenced using other standard methods and are therefore referred to as "sequencing-ready" libraries.

In one embodiment, the methods of the invention are used to generate a sequencing-ready library for sequence analysis using the Illumina Genome Analyzer System and at least one of the linkers 20, 30 includes at least one anchor probe binding site (otherwise referred to as a flow cell binding site) and a sequence for annealing to a sequencing primer. Prior to sequence analysis, the library is denatured (i.e., in 0.2 M NaOH) for 5 minutes at room temperature) and bound to the flow cell.

Such sequence-ready libraries can be analyzed separately or, if modified to contain molecular bar codes, a plurality of libraries can be combined as a mixture into a single pool of libraries and analyzed. When a reaction is performed on a pooled bar-coded library, the reaction need only be performed once. When analyzed as a pool of libraries, the analysis can include detection (such as sequencing) of the molecular bar codes.

As shown in FIGURE 3, a library or pool of libraries made according to the methods of the invention can be sequenced at step 640 or may be further enriched for target sequences of interest (as shown in FIGURE 3, steps 650-670) using solution-based capture methods and analyzed as described in detail below.

SOLUTION-BASED CAPTURE TO ENRICH A LIBRARY FOR TARGET SEQUENCES OF INTEREST

In another aspect, the present invention provides a method of enriching a library for target nucleic acid regions of interest. The method according to this aspect of the

IDT Exhibit 1020.2019

invention comprises (a) contacting a library of DNA molecules comprising a subpopulation of nucleic acid target insert sequences of interest flanked by a first primer binding region and a second primer binding region within a larger population of nucleic acid insert sequences flanked by the first primer binding region and the second primer binding region with a set of capture probes, the set of capture probes comprising a plurality of capture oligonucleotides, each comprising a first target sequence-specific binding region and a second capture reagent binding region, under conditions that allow binding between the capture oligonucleotides and the nucleic acid target regions of interest, to form a plurality of complexes between target regions of interest and capture probes; (b) contacting the mixture of step (a) with a capture reagent and separating the capture reagent bound complex from the mixture; and (c) eluting the target regions of interest flanked by the first primer binding region and the second primer binding region from the capture reagent bound complex.

Any library of DNA molecules comprising a subpopulation of nucleic acid target insert sequences of interest flanked by a first primer binding region and a second primer binding region within a larger population of nucleic acid insert sequences flanked by the first primer binding region and the second primer binding region may be enriched for target sequences using the methods of this aspect of the invention. In one embodiment of the method, a library of DNA molecules comprising a subpopulation of nucleic acid target insert sequences of interest flanked by a first primer binding region and a second primer binding region within a larger population of nucleic acid insert sequences flanked by the first primer binding region and the second primer binding region, generated using the methods of the invention, as shown in FIGURE 3 (steps 610-630) and described, *supra*, is enriched using the methods of this aspect of the invention. The use of solution-based capture to enrich a library allows for the efficient creation of resequencing samples (sequence-ready libraries) that are largely composed of target sequences, as demonstrated in Examples 3-7.

TARGET CAPTURE PROBES

As shown in FIGURE 4, in one embodiment, the sense 100 or antisense 100' target capture probes, each comprises a target sequence-specific binding region 102, 102' and a capture reagent binding region 104 attached to a moiety 110 for binding to a capture reagent 400. In operation, as shown in FIGURE 4, step B, the target-specific binding region 102 of sense 100 or antisense 100' target capture probes bind to a

IDT Exhibit 1020.2020

complementary or substantially complementary nucleic acid sequence contained in an insert region 10 or 10' of a nucleic acid molecule 50 in the library. The moiety 110 (e.g., biotin) attached to the capture probe 100, 100' is then contacted with a capture reagent 400 (e.g., a magnetic bead) having a binding region 410 (e.g., streptavidin coating) and the complex is pulled out of solution with a sorting device 500 (e.g., a magnet) that binds to the capture reagent 400.

The length of a capture probe is typically in the range of from 10 nucleotides to about 200 nucleotides, such as from about 20 nucleotides to about 150 nucleotides, such as from about 30 nucleotides to about 100 nucleotides, and such as from about 40 nucleotides to about 80 nucleotides.

The target-specific binding region 102,102' of the target capture probe is typically from about 25 to about 150 nucleotides in length (e.g., 50 nucleotides, 100 nucleotides) and is chosen to specifically hybridize to a target sequence of interest. In one embodiment, the target-specific binding region comprises a sequence that is substantially complementary (i.e., at least 90% identical, at least 95% identical, at least 96% identical, at least 97% identical, at least 98% identical, at least 99% identical, or 100% identical) to a target sequence of interest.

In one embodiment, the capture probe is about 70 nucleotides in length, comprising a target-specific region of about 35 nucleotides in length.

One of skill in the art can use art-recognized methods to determine the features of a target binding region that will hybridize to the target with minimal non-specific hybridization. For example, one of skill can determine experimentally the features such as length, based composition, and degree of complementarity that will enable a nucleic acid molecule (e.g., the target-specific binding region of a target capture probe) to specifically hybridize to another nucleic acid molecule (e.g., the nucleic acid target) under conditions of selected stringency, while minimizing non-specific hybridization to other substances or molecules. For example, for an exon target of interest, a target gene sequence is retrieved from a public database such as GenBank, and the sequence is searched for stretches of from 25 to 150 bp with a complementary sequence having a GC content in the range of 45% to 55%. The identified sequence may also be scanned to ensure the absence of potential secondary structure and may also be searched against a public database (e.g., a BLAST search) to ensure a lack of complementarity to other genes.

IDT Exhibit 1020.2021

The capture oligonucleotides may be designed to bind to a target region at selected positions spaced across the target region at various intervals. The capture oligo design and target selection process may also take into account genomic features of the target region such as genetic variation, G:C content, predicted oligo Tm, and the like.

In some embodiments, the methods of the invention are used to capture and sequence a modified or mutated target, such as to determine the presence of a particular single nucleotide polymorphism (SNP) or deletion, addition, or other modification. In accordance with such embodiments, the set of target capture probes are typically designed such that there is a very dense array of capture probes that are closely spaced together such that a single target sequence, which may contain a mutation, will be bound by multiple capture probes that overlap the target sequence. For example, capture probes may be designed that cover every base of a target region, on one or both strands, (i.e., head to tail) or that are spaced at intervals of every 2, 3, 4, 5, 10, 15, 20, 40, 50, 90, 100 or more bases across a sequence region.

As another example, the selection of the target capture probes over a target region of interest is based on the size of the target region. For example, for a target region of less than 100 nucleotides in length, capture probes (either sense, antisense, or both) are typically designed to hybridize to target sequences spaced apart by from 0 to 100 nucleotides, such as every 45 nucleotides. As another example, for a target region greater than 200 nucleotides, capture probes (either sense, antisense, or both) are typically designed to hybridize to target sequences spaced apart by from 0 to 200 nucleotides, such as at 45 to 65 nucleotide intervals. In one embodiment, for a target region greater than 200 nucleotides (e.g., a 200,000-nucleotide target region), a set of sense and antisense capture probes are designed that are each about 35 nucleotides in length and are spaced about 45 nucleotides apart across the target region (alternating sense/antisense) in order to saturate the region (e.g., "tile" across the region of interest).

In some embodiments of the method, a set of capture probes is designed to specifically bind to a plurality of target regions, such as the exons of a single gene or multiple genes, such as at least 5 genes, at least 10 genes, at least 20 genes, at least 50 genes, at least 75 genes, or more.

In some embodiments of the method, a set of capture probes is designed to specifically bind to target sequences across a genomic location, such as across a

IDT Exhibit 1020.2022

chromosomal region, and the capture probes are contacted with nucleic acid molecules from a total genomic library.

In some embodiments of the method, a set of capture probes is designed to specifically bind to target sequences across a genomic location, such as across a chromosomal region, and the capture probes are contacted with nucleic acids in a whole-transcriptome library in order to analyze the whole transcriptome across the chosen genomic locus, as described in Example 7.

In some embodiments of the method, a set of capture probes is designed to specifically bind to a genomic locus known to be associated with a clinical outcome or disease, or disease risk, for example, as described in Example 8.

As shown in FIGURE 4, in one embodiment, the target capture probe 100, 100' comprises a capture reagent binding region 104 attached to a moiety 110 for binding to a capture reagent 400. As will be understood by those of skill in the art, the solution-based capture method utilizes a binding interaction between a moiety 110 attached (directly or indirectly) to a capture probe 100, 100' and a capture reagent 400 to enable the selective separation of captured sequences (bound to the capture probe) from the bulk solution of captured and uncaptured DNA molecules. The moiety 110 and capture reagent 400 may be any suitable binding partners such as, for example, biotin/streptavidin; epitope/antibody, or DNA hybridizing partners.

In one embodiment, the moiety 110 is biotin and the capture reagent 400 is a streptavidin-coated bead 400, which is sorted with a magnetic sorting device 500. Although the moiety 110 shown in FIGURE 4 is located at the 5' end of the capture probe, it will be understood by those of skill in the art that the moiety may alternatively be positioned at the 3' end of the target capture probe 100.

As another example, the moiety 110 and capture reagent 400 may be an epitope/antibody pair, such as a digoxin moiety that is bound by digoxin antibodies or a fluoroscein moiety that is bound by fluoroscing antibodies, or other small epitope/antibody configurations.

As another example, the moiety 110 and capture reagent 400 may be DNA hybridization partners. For example, the moiety 110 on the capture probe may be a sequence that is complementary to an oligonucleotide affixed to beads 400.

As shown in FIGURE 5, in another embodiment of the method of this aspect of the invention, the capture probes 200 comprise a target-sequence specific binding

IDT Exhibit 1020.2023

region 202, 202' and a capture reagent binding region 204 that hybridizes to a universal adaptor oligonucleotide 300 comprising a moiety 310 that binds to a capture reagent 400. In operation, as shown in FIGURE 5, step B, the target-specific binding region 202 of sense 200 or antisense 200' target capture probes bind to a substantially complementary nucleic acid sequence contained in an insert region 10 or 10' of a nucleic acid molecule 50 in the library. The universal adaptor oligonucleotide 300 is present at an equal concentration as the capture probes 200, and hybridize to the capture reagent binding region 204. The moiety 310 (e.g., biotin) attached to the universal oligo adaptor 300 is then contacted with a capture reagent 400 (e.g., a magnetic bead) having a binding region 410 (e.g., streptavidin coating) and the complex is pulled out of solution with a sorting device 500 (e.g., a magnet) that binds to the capture reagent 400.

As shown in FIGURE 6, the methods of solution-based capture 650 include the step 652 of providing a library of nucleic acid molecules comprising nucleic acid target insert sequences of interest flanked by a first primer binding region on one end and a second primer binding region on the other end (e.g., produced as shown in step 630, from FIGURE 3).

At step 654, the library of nucleic acid molecules 50A, 50B is annealed with a set of capture probes, each capture probe comprising a region that hybridizes to a target sequence contained in a library insert. In one embodiment, the capture probes 100 comprise a moiety 110 (e.g., biotinylated) for binding to a capture reagent 400 (e.g., streptavidin-coated beads). In another embodiment, the library of nucleic acid molecules 50A, 50B is annealed with a combination of a set of capture probes 200, each comprising a region 204 that hybridizes to a universal adaptor oligo 300 and an equimolar amount of universal adaptor oligos 300 comprising a moiety 310 for binding to a capture reagent 400.

The annealing step 654 is carried out by mixing a molar excess of capture probes (or capture probes plus universal adaptor oligos) with the library (or pool of bar-coded libraries) in a high salt solution comprising from 100mM to 2 M NaCl (osmolarity = 200 to 4,000 molar). An exemplary high salt solution for annealing is 10 mM Tris pH 7.6, 0.1 mM EDTA, 1 M NaCl (osmolarity = 2,000 molar). The nucleic acid molecules in the mixture are then denatured (i.e., by heating to 94 degrees) and allowed to cool to room temperature. In one embodiment, the annealing step is carried out in a high salt solution

IDT Exhibit 1020.2024

comprising from 100 mM to 2 M NaCl with the addition of 0.1% triton X100 (or tween or NP40) nonionic detergent.

At step 655, an amount of capture reagent is added to the annealed mixture sufficient to generate a plurality of complexes, each containing a nucleic acid molecule, a capture probe (or a capture probe and a universal adaptor oligo), and a capture reagent. This step is carried out in a high salt solution comprising from 100 mM to 2 M NaCl (osmolarity = 200 to 4,000 molar). An exemplary high salt solution for anneal is 10 mM Tris pH 7.6, 0.1 mM EDTA, 1 M NaCl (osmolarity = 2,000 molar). The mixture is incubated at room temperature with mixing for about 15 minutes.

At step 656, the complexes formed in step 655 are isolated or separated from solution with a sorting device 500 (e.g., a magnet) that pulls or sorts the capture reagent 400 out of solution.

At step 658, the sorted complexes bound to the capture reagent 400 are washed with a low salt wash buffer (less than 10 mM NaCl, and more preferably no NaCl) to remove non-target nucleic acids. An exemplary low salt wash buffer is 10 mM Tris pH 7.6, 0.1 mM EDTA (osmolarity = 10 millimolar). In some embodiments, the low salt wash optionally contains from 15% to 30% formamide, such as 25% formamide (osmolarity = 6.3 molar). For each wash step, the capture reagent 400 bound to the complexes (i.e., magnetic beads) are resuspended in the low salt wash buffer and rocked for 5 minutes, then sorted again with the sorting device (magnet). The wash step may be repeated 2 to 4 times.

At step 660, the nucleic acid molecules containing the target sequences are eluted from the complexes bound to the capture reagent as follows. The washed complexes bound to the capture reagent 400 are resuspended in water, or in a low salt buffer (i.e., osmolarity less than 100 millimolar), heated to 94°C for 30 seconds, the capture reagent (i.e., magnetic beads) are pulled out using a sorting device (i.e., magnet), and the supernatant (eluate) containing the target nucleic acid molecules is collected.

At step 670, the eluate is amplified in a PCR reaction with a first PCR primer that binds to the first primer binding site in the first linker and a second PCR primer that binds to the second primer binding site in the second linker, producing a once-enriched library that can be optionally sequenced at step 680.

As shown in FIGURE 6, the once-enriched library may be further processed according to steps 654-670 using the same set of capture probes in each round of

IDT Exhibit 1020.2025

enrichment to generate a library that is twice-enriched or three-times enriched, etc., for the target sequences of interest prior to sequence analysis.

In one embodiment, the ratio of the concentration of the DNA target in the first and second round of enrichment to the concentration of capture oligo is a concentration of about 500 ng/ml DNA target to a concentration in the range of from about 1 nM to 10 nM of capture oligo. In one embodiment, the ratio of the concentration of DNA target in the third round of enrichment to concentration of capture oligo is a concentration of about 500 ng/ml of the twice-enriched library to a concentration of about 1 nM of capture oligo.

In one embodiment, the first round of enrichment (steps 654-670 shown in FIGURE 6) is carried out with a first set of capture probes designed to target a first set of targets, followed by a second round of enrichment that is carried out with a second set of capture probes designed to target a second set of targets.

In one embodiment, the capture reagent (400) comprises streptavidin coated magnetic beads, each bead having a binding capacity of approximately 50 pmol of biotinylated double-stranded DNA/50 μl of beads. In one embodiment, at step 655, about 50 μl of the streptavidin coated magnetic beads are added to about 5 μg of the annealed nucleic acids (e.g., in the first and second rounds of enrichment). In one embodiment, at step 655, about 5 μl of the streptavidin coated magnetic beads are added to about 5 μg of the annealed nucleic acids (e.g., in the third round of enrichment).

As described in Examples 3-5, the solution-based capture methods according to the various embodiments described herein may be used to produce a level of target fragment specific enrichment in the range of 500- to 900-fold in the first round of enrichment, with a 50-fold higher level of enrichment in the second round (i.e., 25,000- to 45,000-fold total enrichment levels).

In one embodiment, the final round of enrichment may be carried out with a limiting amount of capture probe to library, in order to allow for the normalizing or leveling of target gene sequences in the enriched library, such that there will be a broad distribution in the frequency of amplified targets.

OLIGONUCLEOTIDE SYNTHESIS

DNA synthesis of the various oligonucleotides of the invention (e.g., stem-loop linkers, capture probes, and universal adaptor oligonucleotides) can be carried out by any art-recognized chemistry, including phosphodiester, phosphotriester, phosphate triester,

IDT Exhibit 1020.2026

or N-phosphonate and phosphoramidite chemistries (see e.g., Froehler et al., *Nucleic Acid Res. 14*:5399-5407, 1986; McBride et al., *Tetrahedron Lett. 24*:246-248, 1983). Methods of oligonucleotide synthesis are well known in the art and generally involve coupling an activated phosphorous derivative on the 3' hydroxyl group of a nucleotide with the 5' hydroxyl group of the nucleic acid molecule (see, e.g., Gait, "Oligonucleotide Synthesis: A Practical Approach," IRL Press, 1984).

In some embodiments, capture probes 100, 100' are synthesized to include RNA residues (i.e., DNA/RNA hybrid molecules) and/or unnatural bases such as inosine that have altered base pairing and/or have modified backbone sequences such as thiophosphate.

The following examples merely illustrate the best mode now contemplated for practicing the invention but should not be construed to limit the invention.

EXAMPLE 1

This example describes the use of a PCR-based approach to generate a sequencing-ready library of the exon amplicons of 5 genes of interest, with an optional further modification to include the use of molecular bar code sequences.

Rationale

One application of highly-parallel sequencing technology such as the Illumina sequencing platform (Illumina, Inc., San Diego, California) is the targeted resequencing of particular regions of a sequenced genome, such as the human genome. In this example, the targeted regions were the coding exons of 5 human genes--AKT1, KRAS, PIK3CA, PTEN, and TP53. PCR was used to retrieve 52 exonic regions derived from these 5 genes and methods are described herein for converting these DNA amplicons into fragmented samples flanked by linkers containing primer binding sites suitable for sequencing. The output of sequences from a system such as the Illumina platform is of sufficient quantity that it is conceivable to sequence several samples at once. To analyze samples simultaneously, each sample must be uniquely tagged. One method for tagging, validated in this example, is to append a specific sequence of nucleotides, with each attached sequence unique to each sample, between the sequencing initiation site and the fragmented library segments that are to be sequenced. In this way, the first few bases of sequence uniquely identify the sample while the remaining sequence will be derived from the target regions that are being analyzed in that sample. In this example, molecular bar

IDT Exhibit 1020.2027

code tags of 3 nucleotides were attached to unique sequencing libraries, and all 64 possible combinations of these codes were combined into a single sequencing library. Analysis of the output sequences confirmed that each code was uniquely associated with the appropriate library sequences. By extension, varying the code length to n bases makes it possible to generate $4^{(n)}$ codes.

This example demonstrates that the all of the regions included in pooled PCR fragments that were sequenced were successfully converted into fragments flanked by linkers that generated sequence information. Moreover, this example also shows that molecular bar codes can be used to multiplex samples into a single sequencing reaction from which sequence information unique to each sample can be subsequently extracted by computational analysis.

Selection and Initial Evaluation of Primer Pairs for Exon Amplification.

A. Selection of Primer Pairs

PCR primer pairs for the following 5 genes--AKT1, KRAS, PIK3CA, PTEN, and TP53--were selected using an exon primer selection software entitled "Exon Primer," available on the UCSC Genome Bioinformatics browser at http://genome.ucsc.edu/. Five pairs of PCR primers per exon were initially selected for evaluation for PCR amplification of each exon in the 5 gene set.

PCR primers were chosen using the following criteria:

(1) A minimum distance of 35 bp between the primer and the exon/intron boundary, (resulting in a primer region of 70 bp).

(2) A maximal target exon size of 500 bp, with an overlap of 50 bp, such that exons larger than the maximal target size were divided into two primer sets. In the case where the introns were small, the primers were chosen to amplify across more than 1 exon.

(3) A target primer annealing temperature of 60°C, with a GC clamp, which is comprised of one or more G:C base pairs at the 3' primer terminus and is intended to stabilize the termini of the primer:template duplex.

(4) A primer length of from 17 nt to 27 nt, such as from 24 to 27 nt.

(5) A maximum length of a mononucleotide repeat (e.g., AAAA) of 4 nt.

IDT Exhibit 1020.2028

(6) Primer sequences were also masked against common repeat elements found in the human genome such that primer pairs with a potential to amplify multiple segments of the genome were removed.

Using the above criteria, an initial set of PCR primer was selected and tested as described below. Primers were delivered in 10 individual 96 well plates as 100 µl of a 100 µM stock. Stock primers were diluted 1:50 in water to create working primers that were 2 µM. Stock primers and working primers were stored at -20°C.

B. PCR Amplification of Exons

PCR reactions were carried out using the candidate set of primers as described below and the reactions were evaluated on agarose gel to determine if the correct sized PCR product was generated.

PCR Reaction Conditions:

3.5 µl H_2O

2 µl 5X buffer (supplied by the manufacturer with Expand High Fidelity PLUS, Roche Applied Sciences, Indianapolis, IN)

2 µl Forward exon-specific primer (2 µM)

2 µl Reverse exon-specific primer (2 µM)

0.2 µl genomic DNA (100 ng/µl)

0.2 µl dNTP (10 mM)

0.1 µl enzyme (Expand High Fidelity PLUS)

10 µl total

PCR cycling conditions:

1 cycle

94°C 2 minutes

10 cycles:

94°C for 30 sec

60°C for 30 sec

72°C for 1 min

25 cycles:

94°C for 30 sec

60°C for 30 sec

72°C for 1 min + 10 sec/cycle

IDT Exhibit 1020.2029

72°C for 7 min

4°C hold

Results

The results were analyzed on agarose gels for the presence of a PCR product of the expected size, and amount of product. The results are summarized below in TABLE 1.

TABLE 1: Summary of Initial Results Using Candidate PCR Primer Pairs Tested for Exon Amplification

Target Gene	Genbank Ref No.	Total Number of Exon Amplicons	Total number of primer pairs Tested	Range of expected PCR product sizes	Percentage of exons that were amplified by at least one primer pair of 5 tested
AKT1	NM_001014431	14	70	224-701bp	11/14 = 78%
KRAS	NM_004985	5	25	294-410bp	3/5 = 60%
PIK3CA	NM_006218	21	93	213-656bp	1/21 = 5 %
PTEN	NM_000314	9	45	200-1000bp	2/9 = 22 %
TP53	NM_000546	11	50	257-677bp	6/11 = 54%
Total: 5 Genes		Total: 60 Exon Amplicons	Total: 283 primer pairs tested	Size Range of amplicon: 200 bp to 1000bp	Average exon amplification success rate per gene tested: 44%

As summarized in TABLE 1, for many exons, all PCR primer pairs attempted failed in the first PCR reactions attempted. For some exons, only one PCR primer pair or a few PCR primer pairs gave any PCR product. Therefore, it was concluded that the PCR reaction conditions needed to be varied in order to increase the success rate and robustness of the reaction.

C. Methods of Increasing Yield and Specificity of PCR Products

Methods

PCR reaction conditions were varied to test the effect of $MgCl_2$ concentration (1.5 mM or 3.0 mM) DMSO (5%) and Betaine (1.5M) on exon PCR product yield and specificity using the candidate set of primer pairs for the target gene AKT1, which were designed as described above.

10 µl PCR reactions were set up as described in TABLE 2 below. PCR Cycling conditions were as shown above, with a 55°C degree annealing temperature. For the

IDT Exhibit 1020.2030

sample set shown in TABLE 2, 5 primer pairs were tested that previously produced results that ranged from good to no product from the AKT1 primers, as summarized in TABLE 1.

TABLE 2: PCR Conditions and Results for AKT Exon-Specific Amplification

Sample Set	gDNA (template)	$MgCl_2$	DMSO (5%)	Betaine (1.5M)	Results: Product amplified (5 primer pairs tested)	Results: Product yield
1	4 ng (low)	1.5 mM (low)	-	-	no product (all 5 primer pairs)	no product (all 5 primer pairs)
2	4 ng	1.5 mM	+	-	yes- single band (all 5 primer pairs)	low-medium (all 5 primer pairs)
3	4 ng	1.5 mM	-	+	no product (all 5 primer pairs)	no product (all 5 primer pairs)
4	20 ng (high)	1.5 mM	-	-	no product (all 5 primer pairs)	no product (all 5 primer pairs)
5	20 ng	1.5 mM	+	-	yes- single band (all 5 primer pairs)	medium-high (all 5 primer pairs)
6	20 ng	1.5 mM	-	+	no product (all 5 primer pairs)	no product (all 5 primer pairs)
7	4 ng	3.0 mM (high)	-	-	yes, but with multiple bands (4 of 5 primer pairs)	low- medium (4 of 5 primer pairs)
8	4 ng	3.0 mM	+	-	yes - single band (all 5 primer pairs)	medium-high (all 5 primer pairs)
9	4 ng	3.0 mM	-	+	yes -single band (4 of 5 primer pairs)	Medium (4 of 5 primer pairs)
10	20 ng	3.0 mM	-	-	yes, but multiple bands (all 5 primer pairs)	Medium (all 5 primer pairs)
11	20 ng	3.0 mM	+	-	yes- single band (all 5 primer pairs)	High (all 5 primer pairs)
12	20 ng	3.0 mM	-	+	yes- single band (all 5 primer pairs)	High (all 5 primer pairs)

Results

The PCR reactions described above in TABLE 2 were analyzed on a 2% agarose gel with respect to the expected product size, presence of single or multiple (non-specific) bands, and the amount of product. As shown in TABLE 2, it was observed that at low concentrations of $MgCl_2$ (e.g., 1.5 mM), only the PCR reactions containing DMSO produced product, regardless of the amount of template. At the higher concentration of

IDT Exhibit 1020.2031

$MgCl_2$ tested (3.0 mM), all PCR reactions produced product; however, non-specific products were observed in PCR reactions without additives (DMSO or Betaine), which were suppressed in the presence of either 5% DMSO or 1.5 M Betaine. Therefore, it was concluded that DMSO was the most reliable additive and enhanced yield and specificity of the product. $MgCl_2$ at 3.0 mM also increased yield in combination with DMSO. Therefore, high $MgCl_2$ (3.0 mM) and either 1.5 M Betaine or 5% DMSO was chosen as the best combination for exon amplification.

The same set of primers that generated the results summarized in TABLE 1 were used in PCR reactions with high $MgCl_2$ (3.0 mM) and 5% DMSO with 20 ng template and 55°C annealing temperature. Under these conditions, at least one of the 5 primer pairs for each exon produced a single PCR band of the expected size (>98% success rate).

The set of PCR primer pairs that was determined to successfully amplify the 60 exon amplicons of the 5 target genes is provided below in TABLE 3.

TABLE 3: Exon-Specific Primer Pairs and PCR Conditions Used to Amplify Exon Amplicons

Gene	Exon	Left Primer 5' to 3'	Right Primer 5' to 3'	Size	DMSO (5%)		Betaine (1.5M)	
					55° C	60° C	55° C	60 °C
AKT	1	CATCCATCAGAGGGGAAATG (SEQ ID NO:1)	CCAGCCGTTGGACAAATC (SEQ ID NO:2)	668			X	
AKT	2	AGCTTAGAGGGATGGCAGC (SEQ ID NO:3)	CAGACCCTGGGGCTACTACC (SEQ ID NO:4)	334				X
AKT	3	GACGGGTAGAGTGTGCGTG (SEQ ID NO:5)	AGCCAGTGCTTGTTGCTTG (SEQ ID NO:6)	304			X	
AKT	4	GAGAGAGGAAGAGATGGGGC (SEQ ID NO:7)	GAGGATGGCTACAGGCAGAG (SEQ ID NO:8)	352				X
AKT	5	GCTCCTGATCTGGTACAGGC (SEQ ID NO:9)	TCCTCTGACATGGGGAAGAG (SEQ ID NO:10)	378		X		
AKT	6-7	GTAGGCCACCAGGTGTGAAG (SEQ ID NO:11)	CATCGTCCCCTAGAGACAGC (SEQ ID NO:12)	481			X	
AKT	8	GCGACGTGGTATCAAGCAG (SEQ ID NO:13)	GTCTGGTGCCATGGAGAGTAG (SEQ ID NO:14)	264			X	
AKT	9-10	TCTGAGACTTTCCAGGCCC (SEQ ID NO:15)	CTTGGGGCACAGAGAGGAC (SEQ ID NO:16)	561				X

IDT Exhibit 1020.2032

Gene	Exon	Left Primer 5' to 3'	Right Primer 5' to 3'	Size	DMSO (5%)		Betaine (1.5M)	
					55° C	60° C	55° C	60 °C
AKT	11	CCTTGATGCCGAGTCCTG (SEQ ID NO:17)	CCTGAGGCTTTGGAGATC AG (SEQ ID NO:18)	407				X
AKT	12	GGCCCTACATCACAGGA GG (SEQ ID NO:19)	AGTGTGGATATGTGGGGA GC (SEQ ID NO:20)	230			X	
AKT	13	ATCCAGGTGCTTTGAAG GTC (SEQ ID NO:21)	CATTGGCACTCTCCAAAA GG (SEQ ID NO:22)	331		X		
AKT	14	GTCCCTGTGTCAATCTGT GG (SEQ ID NO:23)	CTGGCTGACAGAGTGAGG G (SEQ ID NO:24)	598				X
KRA S	1	GGAACGCATCGATAGCT CTG (SEQ ID NO:25)	CCCTAATTCATTCACTCGC C (SEQ ID NO:26)	347			X	
KRA S	2	CGATACACGTCTGCAGTC AAC (SEQ ID NO:27)	TGGTCCTGCACCAGTAAT ATG (SEQ ID NO:28)	342		X		
KRA S	3	TTGTCCGTCATCTTTGGA GC (SEQ ID NO:29)	TGCATGGCATTAGCAAAG AC (SEQ ID NO:30)	410		X		
KRA S	4	GTTGTGGACAGGTTTTGA AAG (SEQ ID NO:31)	GGATTAAGAAGCAATGCC CTC (SEQ ID NO:32)	387			X	
KRA S	5	CTGTACACATGAAGCCA TCG (SEQ ID NO:33)	CAGTCTGCATGGAGCAGG (SEQ ID NO:34)	294		X		
PIK3 CA	1	GCTGAGGTGTCGGGCTG (SEQ ID NO:35)	GAGTCTCCGGCACCCAC (SEQ ID NO:36)	292				X
PIK3 CA	2	TTTCTGCTTTGGGACAAC C (SEQ ID NO:37)	TTTAAGATTACGAAGGTA TTGGTTTAG (SEQ ID NO:38)	530		X		
PIK3 CA	3	AATCTACAGAGTTCCCTG TTTGC (SEQ ID NO:39)	TCAGTATAAGCAGTCCCT GCC (SEQ ID NO:40)	432		X		
PIK3 CA	4	ACTTGTTGAAATTTCTCC CTTG (SEQ ID NO:41)	CAGAGCCTGCAGTGAGCC (SEQ ID NO:42)	473				X
PIK3 CA	5	TAGAACTACAGTTTCAA AAGTTGACC (SEQ ID NO:43)	CGGAGATTTGGATGTTCTC C (SEQ ID NO:44)	634		X		
PIK3 CA	6	AAGGCAGCAACTAATTT TGG (SEQ ID NO:45)	CTGCTAAACACTAATATA ACCTTTGG (SEQ ID NO:46)	320		X		

IDT Exhibit 1020.2033

Gene	Exon	Left Primer 5' to 3'	Right Primer 5' to 3'	Size	DMSO (5%)		Betaine (1.5M)	
					55° C	60° C	55° C	60 °C
PIK3 CA	7	TGGTTGATCTTTGTCTTC GTG (SEQ ID NO:47)	TTCAATCAGCGGTATAAT CAGG (SEQ ID NO:48)	268		X		
PIK3 CA	8-9	TGGGGAAGAAAAGTGTT TTG (SEQ ID NO:49)	AGAGAAAGTATCTACCTA AATCCACAG (SEQ ID NO:50)	517		X		
PIK3 CA	10	CCTGTCTCTGAAAATAAA GTCTTGC (SEQ ID NO:51)	TGCTGAGATCAGCCAAAT TC (SEQ ID NO:52)	371		X		
PIK3 CA	11	CATGTCAACCTTTTGAAC AGC (SEQ ID NO:53)	TGAGAGAAAACAATTTAA GTGACATAC (SEQ ID NO:54)	214		X		
PIK3 CA	12-13	GGCTCATTCACAACTATC TTTCC (SEQ ID NO:55)	AAACTCTTCCAGCCAAAC ATAAAC (SEQ ID NO:56)	656		X		
PIK3 CA	14	CAGGAACTACCTGAAAC TCATGG (SEQ ID NO:57)	GGGCTTCTAAACAACTCT GCC (SEQ ID NO:58)	424		X		
PIK3 CA	15	CTGCTCTGTGTTGTAGAA ACCC (SEQ ID NO:59)	TCTCAAGATTTTATCCAGA AAAGG (SEQ ID NO:60)	335		X		
PIK3 CA	16	GGTGAAAGTTGTAAATC TTTGTAACAC (SEQ ID NO:61)	AGATAGCTAAATTCATGC ATCATAAG (SEQ ID NO:62)	293		X		
PIK3 CA	17	AAAACCATGTGATGGCG TG (SEQ ID NO:63)	CACTCCAGAGGCAGTAGC AG (SEQ ID NO:64)	393			X	
PIK3 CA	18	GAAAGGCAGTAAAGGTC ATGC (SEQ ID NO:65)	TGTTCTAACTCAGAGGAA TACACAAAC (SEQ ID NO:66)	382		X		
PIK3 CA	19-20	AAATGGAAACTTGCACC CTG (SEQ ID NO:67)	ACAGGCATGAACCACCAC (SEQ ID NO:68)	584				X
PIK3 CA	21	CGAAAGCCTCTCTAATTT TGTG (SEQ ID NO:69)	CCTATGCAATCGGTCTTTG C (SEQ ID NO:70)	452		X		
PTE N	1	GAGTCGCCTGTCACCATT TC (SEQ ID NO:71)	ATCCGTCTACTCCCACGTT C (SEQ ID NO:72)	601		X		
PTE N	2	TTTAGTTTGATTGCTGCA TATTTC (SEQ ID NO:73)	AGTCCATTAGGTACGGTA AGCC (SEQ ID NO:74)	301		X		
PTE N	3	AAACCCATAGAAGGGGTA TTTG (SEQ ID NO:75)	TCTGTGCCAACAATGTTTT ACC (SEQ ID NO:76)	418		X		

IDT Exhibit 1020.2034

Gene	Exon	Left Primer 5' to 3'	Right Primer 5' to 3'	Size	DMSO (5%)		Betaine (1.5M)	
					55° C	60° C	55° C	60 °C
PTEN	4	AAAGATTCAGGCAATGTTTGTTAG (SEQ ID NO:77)	TGACAGTAAGATACAGTCTATCGGG (SEQ ID NO:78)	200		X		
PTEN	5	GGAATCCAGTGTTTCTTTTAAATACC (SEQ ID NO:79)	TCAGATCCAGGAAGAGGAAAG (SEQ ID NO:80)	407		X		
PTEN	6	ATGGCTACGACCCAGTTACC (SEQ ID NO:81)	AAACTGTTCCAATACATGGAAGG (SEQ ID NO:82)	280		X		
PTEN	7	CAGTTAAAGGCATTTCCTGTG (SEQ ID NO:83)	CTCACCAATGCCAGAGTAAGC (SEQ ID NO:84)	359		X		
PTEN	8	GCAACAGATAACTCAGATTGCC (SEQ ID NO:85)	GTCAAGCAAGTTCTTCATCAGC (SEQ ID NO:86)	515		X		
PTEN	9	AGCTTGGCAACAGAGCAAG (SEQ ID NO:87)	CCACAAGTGCAAAGGGGTAG (SEQ ID NO:88)	894				X
TP53	1	CTCCTCCCCAACTCCATTTC (SEQ ID NO:89)	CAAGCTTCCATCCCACTCAC (SEQ ID NO:90)	421			X	
TP53	2-3	CTCAGACACTGGCATGGTG (SEQ ID NO:91)	TGGGTGAAAAGAGCAGTCAG (SEQ ID NO:92)	448		X		
TP53	4	CGTTCTGGTAAGGACAAGGG (SEQ ID NO:93)	AGGAATCCCAAAGTTCCAAAC (SEQ ID NO:94)	485		X		
TP53	5-6	TCACTTGTGCCCTGACTTTC (SEQ ID NO:95)	TCATGGGGTTATAGGGAGGTC (SEQ ID NO:96)	545		X		
TP53	7	TGCTTGCCACAGGTCTCC (SEQ ID NO:97)	AGCAGTAAGGAGATTCCCCG (SEQ ID NO:98)	321		X		
TP53	8-9	TGGTTGGGAGTAGATGGAGC (SEQ ID NO:99)	ACCAGGAGCCATTGTCTTTG (SEQ ID NO:100)	539		X		
TP53	10	AACTTGAACCATCTTTTAACTCAGG (SEQ ID NO:101)	AGCTGCCTTTGACCATGAAG (SEQ ID NO:102)	276		X		
TP53	11	GATTTGAATTCCCGTTGTCC (SEQ ID NO:103)	CAGCATTTCACAGATATGGGC (SEQ ID NO:104)	598		X		

IDT Exhibit 1020.2035

D. DNase I Fragmentation of PCR Exon Amplicon Pool

The 51 exon amplicons were PCR amplified from genomic DNA using the primer pairs and conditions shown in TABLE 3. These PCR products were then pooled and purified over QiaQuick® columns (Qiagen), which removes DNA fragments less than approximately 40 bp. The purified pooled PCR products were present at 50 ng/µl in a size range of approximately 50 bp to 900 bp.

DNase I Digestion

It was determined that bovine pancreatic deoxyribonuclease I (DNase I) induces random double-stranded breaks in DNA in the absence of Mg^{++} and in the presence of Mn^{++} (Anderson, S., *Nucleic Acids Res. 9*(13):3015-3027 (1981); Melgar, E., et al., *J. Biol. Chem. 243*(17):4409-16 (1968)). Therefore, bovine pancreatic DNase I (New England Biolabs Catalog #M0303S) was used to randomly fragment the pool of exon amplicons to generate a sequencing library as described below.

Bovine pancreatic DNase I treatment was tested over a range of concentrations of 0.004U, 0.002U, and 0.001U per µl (in the absence of Mg++ and in the presence of $MnCl_2$) in order to identify the DNase I digestion conditions suitable to result in an average fragment size range of about 50 to about 500 bp from the PCR amplified exon pool.

DNase I Digestion:

2 µl	50 ng DNA (PCR amplified exon pool) per reaction:
1 µl	10X buffer (500 mM Tris pH 7.6, 0.5 mg/ml acetylated BSA)
1.25 µl	40 mM $MnCl_2$
4.75 µl	H_2O
1.0 µl	Bovine Pancreatic DNase I (N.E.B. #M0303S) (2U/µl diluted to 0.004, 0.002, and 0.001U/µl)
10.0 µl	

The Dnase I reaction was incubated at room temperature for 10 minutes, stopped with 0.2 volume of 100 mM EDTA, and run on an agarose gel to determine the size range resulting from the Dnase I digestion.

IDT Exhibit 1020.2036

Results

Agarose gel analysis showed that the range of Dnase I enzyme concentrations tested resulted in digested products ranging in size from a complete digestion (e.g., di- or tri-nucleotides in length) to a slight fragmentation of the exon amplicon pool (e.g., 850 nt in length, data not shown). From this analysis, it was determined that the range of 1:1,000 to 1:1,500 dilutions of Dnase I (2U/µl stock) treatment resulted in the production of DNA fragments in the desired range of about 50 to about 500 bp.

The Dnase I reaction was then scaled up to digest 10 µg total pooled PCR fragments under the conditions described above. The Dnase I digested material was run over a Qiaquick® column (removing fragments smaller than about 50 bp). The purified DNA was then concentrated with Ethanol precipitation, by combing the 200 µl purified DNA, 20 µl of 3M Sodium Acetate, 3 µl of Glyco-blue, and 500 µl 100% ETOH. A total of 4.5 µg DNA was recovered (45 ng/µl in 100 µl total volume).

E. Blunt-End Polishing the Dnase I Digested Fragments

40 µl (1.8 µg) of the purified, Dnase I digested fragment pool was end polished with the Quick Blunting® Kit (New England Biolabs, Catalog #E1201L), according to the manufacturer's instructions. The Quick Blunting® Kit includes a reaction mixture with T4 polymerase (which has both 3' to 5' exonuclease activity and 5' to 3' polymerase activity) and T4 polynucleotide kinase (for phosphorylation of the blunt-ended DNA for subsequent ligation to the stem-loop adaptors), resulting in a final fragment concentration of 40 ng/µl.

Blunt-End Polishing Reaction:

10 µl purified DNase I treated DNA (45 ng/µl)
2 µl 10X blunt buffer (supplied with kit)
5.2 µl H_2O
2 µl 1 mM dNTP
0.8 µl enzyme (mixture of T4 polymerase plus T4 polynucleotide kinase)
20 µl total

The reaction mixture was incubated at room temperature for 30 minutes, then at 70°C for 10 minutes. This blunt-end polished DNA was ligated to the stem-loop adaptors as follows.

IDT Exhibit 1020.2037

F. Ligation of Stem-Loop Linkers to Fragments

Rationale

To facilitate subsequent PCR amplification and sequencing, oligonucleotide linkers containing PCR primer binding sites (referred to as stem-loop linkers) were ligated to blunt-ended library fragments. The oligo linkers were designed as single DNA oligonucleotides capable of self-annealing to form a stem-loop secondary structure. The stem forms a blunt ended dsDNA segment suitable for ligation to the blunt-end library fragments. In this example, stem sequences were utilized that were 15 to 18 nucleotides in length with roughly 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 ~45°C. Moreover, the formation of the ligatable dsDNA stem is a self:self intermolecular reaction that is highly efficient and each adaptor has only one dsDNA termini capable of ligation. In principle, self-annealing stem structures that range in size from 5 nucleotides to >100 nucleotides may be included in the stem loop adaptor.

As shown in FIGURE 1 at step A, a pair of stem-loop linker oligonucleotides, shown as a first stem-loop linker 20 and a second stem-loop linker 30, was designed for ligation to the ends of each DNase I digested and blunt end-polished double-stranded DNA fragment 10. This ligation reaction generated a mixture of ligation products including the target molecules 50A and 50B comprising a plurality of DNA inserts 10 flanked by the first stem-loop linker 20 at one end and the second stem-loop linker 30 at the other end, as well as unwanted byproduct ligation products 60, 70 comprising a plurality of DNA inserts 10 flanked at both ends by either the first stem-loop linker 20 or flanked at both ends by the second stem-loop linker 30, as shown in FIGURE 1 at step D.

As further shown in FIGURE 1 at step A, the first stem-loop linker oligonucleotide 20 comprises a 5' region 24 with a sequence that is complementary to a sequence located in the 3' region 28 and an intervening region 26 between the 5' and 3' region that forms a loop structure. Also located in the first stem-loop linker oligonucleotide 20 is a sequence 22 that is complementary to a first primer binding region 82 that may be positioned in the intervening region 26 or in the stem region. Under non-denaturing conditions, the 5' region 24 and 3' region 28 hybridize together, resulting in the stem-loop linker oligonucleotide 20 structure with a double-stranded stem 24 and 28 with an intervening region 26 that forms a loop structure.

IDT Exhibit 1020.2038

Similarly, as further shown in FIGURE 1 at step A, the second stem-loop linker oligonucleotide 30 comprises a 5' region 34 having a sequence that is complementary to a sequence located in the 3' region 38 and an intervening region 36 between the 5' and 3' region that forms a loop structure. Also located in the second stem-loop linker oligonucleotide 30 is a sequence 32 that is complementary to a second primer binding region 92 that may be positioned in the intervening region 36 or in the stem region.

Under non-denaturing conditions, the 5' region 34 and 3' region 38 hybridize together, resulting in the stem-loop linker oligonucleotide 30 structure with a double-stranded stem 34 and 38 with an intervening region 36 that forms a loop structure.

The sequences 22, 32 are complementary to the first and second primer binding regions 82, 92, which contain primer binding sites for binding to forward and reverse PCR primers, as described in more detail below.

The total length of each stem-loop linker 20, 30 is typically at least 40 nucleotides, such as at least 45 nucleotides, at least 50 nucleotides, at least 55 nucleotides, at least 60 nucleotides, at least 65 nucleotides, at least 70 nucleotides, up to a maximum length of about 200 nucleotides. In some embodiments of the methods described herein, the stem-loop linkers are from about 45 nucleotides in length to about 70 nucleotides in length.

The use of 5' and 3' stem-loop linkers is a key element of the library construction since they provide universal primer binding sites for subsequent PCR and may contain primer binding sites/anchors for sequencing cluster generation and they can be used to introduce bar-codes for sample multiplexing.

As described in more detail below, suppression PCR may be used to prepare a sequencing-ready library enriched for the target molecules 50A and 50B comprising heterogeneous stem-loop adaptors at each end of the insert, as shown in the PCR products in FIGURE 1 at step D.

As further illustrated in FIGURE 1, step A, at least one of the stem-loop linkers (e.g., 20 may optionally include a bar code sequence 40. As shown in FIGURE 1, the bar code sequence 40 may be positioned at the 3' end of the linker 20, so that it is adjacent the insert 10 after ligation. As shown in FIGURE 1, a complementary sequence 40' is present on the 5' end of the linker 20.

An exemplary set of stem-loop linkers 20, 30, shown below, were used in the following experiment.

IDT Exhibit 1020.2039

First Stem-Loop Linker #1 (20)

5'AGATCGGAAGAGCGT*AATGATACGGCGACCACCGA*CACTCTTTCCCTA CACGACGCTCTTCCGATCT3' (SEQ ID NO:105)

SEQ ID NO:105 has a total length of 67 nucleotides, and consists of a 5' 15 nucleotide stem hybridizing region 24 (underlined), a 37 nucleotide intervening loop region 26 and a 3' 15 nucleotide stem hybridizing region 28 (underlined), with a sequence 22 complementary to the first PCR primer binding region 82 shown in italics.

Second Stem-Loop Linker #1 (30)

5'AGATCGGAAGAGCTC*GTATGCCGTCTTCTGCTTG*GAGCTCTTCCGATC T3' (SEQ ID NO:106).

SEQ ID NO:106 has a total length of 49 nucleotides, and consists of a 5' 15 nucleotide stem hybridizing region 34 (underlined), a 19 nucleotide intervening loop region 36 and a 3' 15 nucleotide stem hybridizing region 38 (underlined), with a sequence 32 complementary to the second PCR primer binding region 92 shown in italics.

Second Stem-Loop Linker #2 (30)

5'AGATCGGAAGAGCTC*CAAGCAGAAGACGGCATAC*GAGCTCTTCCGATC T3' (SEQ ID NO:107).

SEQ ID NO:107 is a total length of 49 nucleotides, and consists of a 15 nucleotide stem hybridizing region 34 (underlined), a 19 nucleotide intervening loop region 36, and a 3' 15 nucleotide stem hybridizing region 38 (underlined), with a sequence 32 complementary to the second PCR primer binding region 92 shown in italics.

Following fragmentation of PCR products, a pair of first 20 and second 30 stem-loop linker oligonucleotides were ligated to the blunt end-polished fragments 10 as follows.

Dephosphorylation of Stem-Loop Linkers

A test experiment was carried out to determine the conditions for ligation of stem-loop linkers to a double-stranded DNA fragment with phosphorylated blunt ends.

A test vector pCR2.1 (Invitrogen, Carlsbad California) was digested with PvuII to generate blunt ends. Stem-loop linkers (SEQ ID NO:105 and SEQ ID NO:107) and Antarctic alkaline phosphatase (New England Biolabs, Catalog #M0289S) at a ratio of 30- to 50-fold enzyme to linker, were incubated in dephosphorylation buffer for one

IDT Exhibit 1020.2040

hour at 37°C. The dephosphorylation enzyme was heat inactivated at 65°C for 5 minutes. The PvuII digested plasmid (1 µl) was ligated with the dephosphorylated stem-loop linkers (SEQ ID NO:105 and SEQ ID NO:107) (4 µl) in a 20 µl ligation reaction, the ligations were PCR amplified (25 cycles) and an aliquot of the PCR reaction was examined on an agarose gel.

Results

It was observed that at the highest amount of phosphatase treated linker (8 µg and 4 µg) some linker:dimer PCR band was present (data not shown). However, dephosphorylation of the stem-loop linkers and dilution of the stem-loop linkers prior to ligation completely eliminated the linker:dimer PCR artifact.

G. Ligation of Dephosphorylated Stem-Loop Linkers With Blunt End-Polished DNase I Treated Exon Amplicon Pools

A series of ligation reactions were set up to determine the ability to ligate stem-loop linkers with blunt end-polished DNase I fragmented exon amplicon pools to generate a sequencing library.

It was determined that dephosphorylation of the stem-loop linkers (e.g., SEQ ID NO:105 and SEQ ID NO:107) followed by ligation of the dephosphorylated stem-loop linkers (SEQ ID NOS:105 and 107) to the DNase I digested blunt-end filled-in exon amplicon pool resulted in a ligated configuration with a stem-loop linker oligonucleotide ligated to the 5' end of a first strand of a double-stranded fragment and a stem-loop linker oligonucleotide ligated to the 5' end of a second strand of the double-stranded fragment that is the reverse complement of the first strand, as shown in FIGURE 1 at step B.

Ligation Mixture:

10 µl	2X buffer (N.E.B. Quick ligation Kit #M2200S)
2 µl	genomic DNA, DNAseI treated and blunt-end polished (40 ng/µl)
4 µl	dephosphorylated forward stem-loop adaptor (10 µM) (SEQ ID NO:105)
4 µl	dephosphorylated reverse stem-loop adaptor (10 µM) (SEQ ID NO:107)
1 µl	Quick Ligase enzyme (N.E.B. Quick ligation Kit #M2200S)
20 µl total	

The ligation mixture was incubated for 10 minutes at room temperature, diluted with 180 µl of TEzero (10 mM Tris pH 7.6 and 0.1 mM EDTA), and was used as a template in the suppression PCR reaction described below.

IDT Exhibit 1020.2041

Pre-PCR Fill-In Reaction

As further shown in FIGURE 1 at step B, the first stem-loop linker 20 adds information to the 5' end of the double-stranded insert 10; however, this information is on the wrong strand to be useful in PCR amplification, therefore this information needs to be copied over to the 3' end to create a primer binding site. This was accomplished by a pre-PCR fill-in reaction using Taq polymerase. As described below, the reaction mixture was incubated for 1 minute at 72°C prior to standard PCR in order to transfer the linker information over to the complementary strand, resulting in the fill-in products 50 and 50' shown in FIGURE 1B, step C.

H. Suppression PCR to Selectively Amplify Target Ligation Products

Rationale

One of the primary goals of the adaptor ligation is to enrich the library for target molecules 50A and 50B, as shown in FIGURE 1, step D (PCR products), that possess a different stem-loop linker 20, 30 on each end of the insert 10. During the ligation reaction, the stem-loop linkers attach randomly to library fragments, resulting in an initial population of ligation products where half the ligation products have the same linker termini on each end (homogeneous linker ends) and half the ligation products possess different linker termini (heterogeneous linker ends). The phenomenon of suppression PCR (P.D. Siebert et al., *Nucleic Acids Res. 23*:1087-1088 (1995)) was used in this example to selectively enrich for library fragments with heterogeneous adaptor termini. Briefly described, suppression PCR refers to the phenomenon that DNA segments that contain perfect inverted repeats at their termini longer than 40 nucleotides are poor substrates for amplification by PCR. The conceptual model is that these molecules form spontaneous intramolecular stem-loop structures that occlude PCR primer binding and subsequent amplification. The empirical observation is that molecules with perfect inverted repeat termini ≥40 nt amplify poorly relative to similar DNA fragments with heterogeneous ends. Here, we exploit the fact that our stem loop adaptors add either 50, 67, or 73 nucleotides of additional sequence to the ends of ligated DNA fragments. In molecules with homogenous ends, these added sequences are long enough to evoke suppression PCR effects; hence, molecules with heterogeneous ends (e.g., 50A, 50B) are preferentially amplified and therefore the library is enriched for the sequencing-ready

IDT Exhibit 1020.2042

target molecules 50A, 50B, by the PCR reaction that follows ligation of the stem-loop linkers, resulting in a library enriched for sequencing-ready target molecules.

As further shown in FIGURE 1 at step A, the ligation of first stem-loop linkers 20 (e.g., SEQ ID NO:105) and second stem-loop linkers 30 (e.g., SEQ ID NO:106) to blunt end fragments 10 would be expected to result in the following mixture of ligation products--approximately 50% target molecules 50A, 50B with heterogeneous ends (including 25% first linker-insert-second primer 50A and 25% second primer-insert-first primer 50B); and 50% byproducts (including 25% first primer-insert-first primer 60 and 25% second primer-insert-second primer 70). In order to remove the 50% byproduct ligation products 60, 70 with the same primer sequence at both ends, suppression PCR was used as described below in order to selectively amplify the target 50A, 50B ligation products to generate a library of nucleic acid molecules that are suitable for direct use as sequencing templates (i.e., sequence ready).

It is known that an extended stem loop structure (e.g., greater than 40 nucleotides) is difficult to amplify because the double-stranded stem occludes binding of primers. This phenomenon has been termed "suppression PCR effects." As shown in FIGURE 1 at step D, the unwanted 50% ligation byproducts 60, 70 are refractive to PCR amplification because the first stem-loop linker oligonucleotide (e.g., SEQ ID NO:105) and second stem-loop linker oligonucleotide (e.g., SEQ ID NO:106) are long (i.e., greater than 40 nucleotides) and result in a stem-loop structure with the fragment insert 10 as the intervening region with the stem formed by hybridizing linker regions. Therefore, a post-ligation PCR amplification step is used to selectively enrich the ligation products having the desired target structure 50A, 50B with heterogeneous linker termini (shown as PCR products in FIGURE 1, step D), as follows.

With reference to FIGURE 1, step D, a first PCR primer 52 that hybridizes to the first PCR primer binding site 82 and a second PCR primer 54 that hybridizes to the second PCR primer binding site 92 generated in the second strand during the PCR fill-in reaction of linkers 20 and 30, respectively, are used to selectively amplify the ligation products with the target structure 50A, 50B.

First PCR primer 52: 5'-AATGATACGGCGACCACCGA-3' (SEQ ID NO:109)

Second PCR primer 54: 5'-CAAGCAGAAGACGGCATACG-3' (SEQ ID NO:110)

IDT Exhibit 1020.2043

PCR Reaction Mixture (with 5% DMSO):

10 µl DNA template (ligation mixture from Step G above)

20 µl 5X buffer (supplied by the manufacturer with the EXPAND$^{plus®}$ Kit, Roche)

10 µl 25 mM $MgCl_2$

10 µl 10 µM First PCR primer (SEQ ID NO:109)

10 µl 10 µM Second PCR primer (SEQ ID NO:110)

5 µl DMSO

5 µl dNTPs (10 mM in each dNTP)

30 µl H_2O

1 µl Taq Polymerase (native Taq 5U/µl, Invitrogen)

1 µl EXPAND$^{PLUS®}$ Polymerase (5U/µl, Roche)

100 µl total

PCR Cycling Conditions:

1 cycle:

72°C for 1 minute; 94°C for 2 minutes. (Note: This step copies the sequence from the ligated stem-loop linker to the complementary strand.)

10 cycles:

94°C for 30 sec

60°C for 30 sec

72°C for 1 minute

10 cycles:

94°C for 30 sec

60°C for 30 sec

72°C for 1 minute plus 10 sec/cycle

1 cycle:

72°C for 7 minutes

4°C hold

I. Construction of PCR Pools

Eight unique PCR pools were constructed as follows.

Amplicons from the 5 genes--AKT1, KRAS, PIK3CA, PTEN, and TP53--were generated as described above and pooled in eight unique configurations. As shown below in TABLE 4, each pool in the set of 8 pools has a unique composition of exonic

IDT Exhibit 1020.2044

amplicons. Each of these eight unique pools was fragmented, blunt-ended, and then each pool was itself attached to a set of eight bar-coded stem-loop linkers that were synthesized, using the stem-loop first linker (SEQ ID NO:105), with an additional 3 nucleotide sequence tag (molecular bar code) added to the 3' end of the stem-loop linker. In this way, each of 8 unique pools was attached to a set of eight bar codes, generating the complete set of 64 bar coded samples shown in TABLE 5.

TABLE 4: Compositions of Amplicon Pools

Gene	Exon	Pool 1	Pool 2	Pool 3	Pool 4	Pool 5	Pool 6	Pool 7	Pool 8
AKT1	1	--							
AKT1	3		--						
AKT1	6-7			--					
AKT1	8				--				
AKT1	12					--			
KRAS	1						--		
KRAS	4							--	
PIK3CA	17								--
TP53	1	--							
AKT1	2		--						
AKT1	4			--					
AKT1	9-10				--				
AKT1	11					--			
AKT1	14						--		
PIK3CA	1							--	
PIK3CA	4								--
PTEN	9	--							
PIK3CA	19-20		--						
AKT1	5			--					
AKT1	13				--				
KRAS	2					--			
KRAS	3						--		
KRAS	5							--	
PIK3CA	2								--
PIK3CA	3	--							

IDT Exhibit 1020.2045

Gene	Exon	Pool 1	Pool 2	Pool 3	Pool 4	Pool 5	Pool 6	Pool 7	Pool 8
PIK3CA	5		--						
PIK3CA	6			--					
PIK3CA	7				--				
PIK3CA	8-9					--			
PIK3CA	10						--		
PIK3CA	11							--	
PIK3CA	12-13								--
PIK3CA	14	--							
PIK3CA	15		--						
PIK3CA	16			--					
PIK3CA	18				--				
PIK3CA	21					--			
PTEN	1						--		
PTEN	2							--	
PTEN	3								--
PTEN	4	--							
PTEN	5		--						
PTEN	6			--					
PTEN	7				--				
PTEN	8	--	--	--	--		--	--	--
TP53	2-3	--	--	--	--	--		--	--
TP53	4	--	--	--	--	--	--		--
TP53	5-6	--	--	--	--	--	--	--	
TP53	7		--	--	--	--	--	--	--
TP53	8-9	--		--	--	--	--	--	--
TP53	10	--	--		--	--	--	--	--
TP53	11	--	--	--		--	--	--	--

Note: the symbol "--" indicates the pool is missing this PCR product.

For example, one representative bar-coded forward stem-loop linker oligonucleotide, designated as the first bar code in Pool #1 ("AAA") in TABLE 5 (shown in italics) is added to SEQ ID NO:105, resulting in the following sequence:

IDT Exhibit 1020.2046

5'*TTT*AGATCGGAAGAGCGTAATGATACGGCGACCACCGACACTCTTTC CCTACACGACGCTCTTCCGATCT*AAA*3' (SEQ ID NO:108).

This set of eight unique samples was matched with all 64 3 nucleotide codes in groups of eight, as shown below in TABLE 5.

TABLE 5: Bar Codes

	1st bar code	2nd bar code	3rd bar code	4th bar code	5th bar code	6th bar code	7th bar code	8th bar code
Pool 1 codes	AAA	AGA	CAA	CGA	GAA	GGA	TAA	TGA
Pool 2 codes	AAC	AGC	CAC	CGC	GAC	GGC	TAC	TGC
Pool 3 codes	AAG	AGG	CAG	CGG	GAG	GGG	TAG	TGG
Pool 4 codes	AAT	AGT	CAT	CGT	GAT	GGT	TAT	TGT
Pool 5 codes	ACA	ATA	CCA	CTA	GCA	GTA	TCA	TTA
Pool 6 codes	ACC	ATC	CCC	CTC	GCC	GTC	TCC	TTC
Pool 7 codes	ACG	ATG	CCG	CTG	GCG	GTG	TCG	TTG
Pool 8 codes	ACT	ATT	CCT	CTT	GCT	GTT	TCT	TTT

The 8 pools were made by generating amplicons for the 5 selected genes, as described in Example 1, which were used to make eight unique pools by (1) leaving out one of eight rows of PCR fragments and pooling 20 µl of the remaining samples, (2) adding an additional 100 µl of weakly amplified products (unless they were designated to be left out), and (3) adding 200 µl of a unique PCR fragment to each pool. The pools were purified over four Qiaquick® columns. The samples were eluted in 60 µl of elution buffer per column, yielding about 200 µl. DNA quantitation by nanodrop revealed a DNA concentration range of 120 to 150 ng/µl, with a total yield of 24-30 µg.

Dephosphorylation of stem-loop adaptors with bar code tags

Sixty-four bar-coded stem-loop linker oligonucleotides were synthesized that contained the sequence of SEQ ID NO:105 plus three additional nucleotides added to the 3' end of SEQ ID NO:105 as shown in TABLE 5, for example, the first bar code ("AAA") in Pool #1 = SEQ ID NO:108.

IDT Exhibit 1020.2047

The 64 bar-coded stem-loop linker oligos were suspended in water to 100 μM, then an aliquot of 10 μM was made by adding 20 μl stock oligo to 180 μl water.

Phosphatase Reaction

5 μl (10 μM stem-loop linker)

5 μl 10X dephosphorylation buffer

5 μl phosphatase enzyme

35 μl H_2O

The reaction was incubated at 37°C for one hour followed by incubation at 65°C for 5 minutes.

Preparation of Amplicon Pools for Ligation

Eight pools comprising various combinations of amplicons were mixed as shown in TABLE 4. The amplicon pools were treated with DNase I in the presence of Mn^{++} as described in Example 1 to yield DNase I digested fragments, which were then purified over Qiaquick® columns (Qiagen Corp.) to generate a pool of fragments in the size range averaging in length from about 50 bp to about 500 bp. The purified fragments were then filled in as described in Example 1 with the Quick Blunting® Kit (New England Biolabs, Catalog #E1201L) according to the manufacturer's instructions. The Quick Blunting® Kit includes a reaction mixture with T4 polymerase (which has both 3' to 5' exonuclease activity and 5' to 3' polymerase activity) and T4 polynucleotide kinase (for phosphorylation of the blunt-ended DNA for subsequent ligation to the stem-loop linkers). The reaction was incubated at room temperature for 30 minutes then at 70°C for 10 minutes.

Ligations of stem-loop linkers to amplicon fragment pools

For each pool (of 8 pools), a master mix was first prepared:

20 μl of blunt end (filled-in), fragmented amplicon pool DNA

40 μl of second stem-loop linker #1 (SEQ ID NO:106)

100 μl 2X ligation buffer

160 μl total volume of master mix

16 μl of the master mix was aliquoted into each of a series of 8 bar-coded forward stem-loop linkers:

(e.g., pool 1 = SEQ ID NO:105+ first to eighth bar code sequence shown in TABLE 5). One μl of ligase was then added to each tube and incubated for 10 minutes.

IDT Exhibit 1020.2048

The 20 µl ligation mix was then diluted 10-fold into TEzero and 2 µl of this was added to subsequent 20 µl PCR reactions as follows.

Post-Ligation PCR Reactions to Selectively Amplify the Target Ligation Products (Suppression PCR)

80 x 20 µl PCR reactions were run as follows:

2 µl diluted ligation mix (e.g., for pool 1: SEQ ID NO:105+3 nucleotide linkers first to eighth bar code sequences--pool 1 amplicons--SEQ ID NO:106, opposite orientation, and ligation byproducts)

4 µl 5x buffer (supplied by the manufacturer with the EXPAND$^{plus®}$ kit, Roche)

1.2 µl 25 mM $MgCl_2$

0.4 µl dNTPs (10 mM each dNTP)

1 µl DMSO

2 µl (4 µM) First PCR primer: 5'-AATGATACGGCGACCACCGA-3' (SEQ ID NO:109)

2 µl (4 µM) Second PCR primer: 5'-CAAGCAGAAGACGGCATACG-3' (SEQ ID NO:110)

1.0 µl enzyme (1:1 blend of Roche Expand$^{Plus®}$ and InVitrogen Taq)

7 µl water

20 µl total

PCR Cycling conditions:

1 cycle:

72°C for 1 minute; 94°C for 1 minute. (Note: This step copies the sequence from the ligated stem-loop linker to the complementary strand.)

10 cycles:

94°C 30 sec

55°C or 60°C 30 sec

72°C 30 sec

15 cycles:

94°C 30 sec

55°C or 60°C 30 sec

72°C 30 sec+ 10 sec/cycle

72°C 7 minutes

4°C hold

IDT Exhibit 1020.2049

5 μl of each of the PCR products were analyzed on an agarose gel. 10 μl aliquots of the remaining material was pooled into a single tube, purified over a QiaQuick® column and submitted for sequencing.

Results

Agarose gel analysis showed that the amplicon pools prior to DNase I digestion had discrete banding patterns. As expected, after DNase I digestion and Qiaquick® purification, a smear was observed without a discrete banding pattern, with a cut-off in size of fragments smaller than 40 bp (due to the column purification step). Importantly, it was observed that the additional bar code sequence added to the stem-loop linkers did not change the ligation reaction, as determined from a side-by-side comparison of ligation products on an agarose gel.

The 64 individual samples were pooled and sequenced on an Illumina GA® sequencing instrument. A total of 3,901,100 sequencing reads were obtained that could be uniquely aligned back to the target regions. To determine if the bar codes were accurately associated with the correct samples, the reads were sorted in two dimensions, as shown in FIGURE 2A. In FIGURE 2A, each row corresponds to the sequencing read density (i.e., number of sequencing reads) associated with a particular 3 nucleotide barcode sequence and each column corresponds to the sequencing reads associated with each gene exon region to which the sequence read aligned. In FIGURE 2A, boxes are white (not shaded) if abundant sequencing reads were detected (≥80% of the average read counts over all bar codes) and black (shaded) if few reads were detected (≤10% of the average counts over all bar codes). FIGURE 2A shows both the expected and observed distribution of reads, which exhibited an identical distribution and could therefore be represented in a single figure. Notably, the pattern of abundant and underrepresented reads was perfectly consistent for all of the bar codes associated with pool 1 and pool 2, and, although not shown, for all eight groups of bar codes analyzed. These results demonstrate that all eight barcode sequences ligated to pool 1 or pool 2 DNA exhibited that same pattern of read densities (which was also true of the entire set of 64 codes used in this experiment--data not shown).

The results shown in FIGURE 2A are summarized in FIGURE 2B, in which the expected and observed pattern of read alignment densities for each pool of bar coded samples is shown. To obtain the sequencing read densities for the pools, the data from

IDT Exhibit 1020.2050

the eight bar codes that formed each pool were summed and analyzed relative to the average density of sequencing reads. Here again, the results can be shown as a single FIGURE because the expected and observed results were identical (i.e., the results shown in FIGURE 2A matched the composition of the pools prepared as described in TABLE 4).

As noted above, a small but nonetheless significant percentage of bar coded reads aligned to exonic regions that were not present in pools (~5% of the average read density relative to pools that included the exonic amplicon). This error rate in assigning bar codes is much higher than the intrinsic sequencing error rates observed throughout the data set, suggesting that incorrect read assignment arises from a different source. At present, it is believed that alkaline phosphatase treatment of the stem-loop ligation adaptors removes one or two bases from the 3 nucleotide code at rather significant frequencies. The consequence would be attachment of a truncated code that is subsequently misinterpreted. Thus, one way to reduce the observed error rate in assigning bar codes would be to purify the oligonucleotides to remove partially truncated barcode sequences prior to use using standard techniques well known in the art.

This example describes the use of a 3 nucleotide tag (molecular bar code) for each first stem-loop linker, resulting in 8 pools, each having 8 unique sequence tags per pool, for a total of 64 tagged sources. For example, pool 1 is a pool of the amplicons listed in TABLE 4 that were generated using first adaptor stem loop primer pool 1 codes: (AAA; AGA, CAA, CGA, GAA, GGA, TAA, and TGA). It will be understood by those of skill in the art that alternative arrangements of the length of the nucleotide tag can provide varying levels of complexity. For example, a 1 nucleotide tag provides a 4 plex; a 2 nucleotide tag provides a 16 plex, a 3 nucleotide tag provides a 64 plex; and a 4 nucleotide tag provides a 256 plex, etc.

As stated above, the output of sequence information from, for example, the Illumina GA2® sequencer, far exceeds the data requirements for analysis of individual samples. Multiplexing strategies are required to make full use of these emerging sequencing technologies and to increase the throughput of samples that can be analyzed. The results described in this example validate the feasibility of adding trinucleotide molecular barcodes to individual samples, facilitating the simultaneous analysis of 64 samples. Other configurations of bar code complexity (nucleotide length) can be applied to samples that require greater or lesser sequence coverage.

IDT Exhibit 1020.2051

As shown in FIGURE 3, this example demonstrates a method of generating a sequencing-ready library 600 comprising the steps of fragmenting a starting population of DNA molecules 610, attaching stem-loop linkers with primer binding sides and optional bar codes 620, and suppressing PCR 630 to generate the sequencing-ready library, which can be sequenced 640. In some embodiments of the method, the starting population of DNA molecules is PCR amplified target regions; therefore, the sequence ready library is already enriched for the sequencing target(s) of interest. As further shown in FIGURE 3, in other embodiments of the method of generating a sequencing-ready library 600, the method further comprises the steps of solution-based capture 650 to enrich the library (e.g., a library generated from total genomic DNA or whole amplified transcriptome) for the sequencing targets of interest prior to sequencing, as described in Examples 3-8 and shown in FIGURE 6.

EXAMPLE 2

This example describes the generation of a sequencing-ready library of genomic DNA inserts. Such libraries can be used for solution-based capture targeted resequencing methods as described below, for analysis of sequence-based chromosomal copy number variation or for biomarker screening/discovery.

Rationale

While PCR is ideally suited for the resequencing of a small number of targets in a modest number of samples, the logistical complexity of large-scale resequencing studies becomes unwieldy as target sizes and sample numbers expand. In fact, such experiments expand in size and complexity as a function of the number of amplicons and samples to be analyzed. To accommodate resequencing of hundreds of genes in hundreds of samples, a different experimental approach to targeted resequencing was required in which resequencing targets could be collected from each sample in a single procedure. To perform this procedure, a collection of oligonucleotides complementary to target resequencing regions is annealed to a whole genome fragment library. The collection of sequences bound to these probes can then be characterized by sequencing. The overall procedure, termed "solution-based capture," is an alternative to PCR that can be scaled to very large resequencing regions. This example describes the construction and characterization of genomic DNA libraries to be used in such a procedure.

IDT Exhibit 1020.2052

With reference to FIGURE 3, this example describes an embodiment of the method of generating a sequencing-ready library 600 by fragmenting a starting population of genomic DNA 610, ligating stem-loop linkers to the DNA fragments 620, suppressing PCR to enrich the library for ligation products with heterogeneous linkers 630, and followed by one or more rounds of solution-based capture 650 to enrich the library for sequencing targets of interest.

In this example, genomic DNA was used as the starting material for the library, although cDNA could also be used as starting material to generate a library. With the exception of the starting material used to generate the inserts for the library, the process of generating the library using stem-loop linkers was nearly identical to that described in Example 1.

Methods

A. Genomic Library Construction

Library construction involved the generation of inserts by fragmentation of genomic DNA or cDNA, followed by blunt-end polishing and ligation of 5' and 3' stem-loop linkers to the blunt-end inserts. The 5' and 3' stem-loop linkers are key elements of the library construction because they provide universal anchors for subsequent PCR and optional sequencing cluster generation, they can be used to introduce bar-codes for sample multiplexing, as described in Example 1, and suppressing PCR may be used to enrich for a library containing heterogeneous stem-loop adaptors at each end of the insert, as shown in FIGURE 1 step C, which can be used as templates for sequencing.

Preparation of Stem-Loop Adaptors

The sequence design of the stem-loop linkers is described in Example 1. An exemplary set of stem-loop linkers used in this example are SEQ ID NO:105 (first stem-loop linker #1) and SEQ ID NO:107 (second stem-loop linker #2).

The forward stem-loop linkers (SEQ ID NO:105) were bar coded as follows:

Four bar codes were used in this experiment, which were chosen to represent all four bases in each of the three base positions, and homopolymers were avoided.

In order to reduce the level of primer-dimer background material, prior to ligation, the stem-loop linkers were pre-treated with Antarctic alkaline phosphatase (New England Biolabs Catalog #M0289S), as described in Example 1.

IDT Exhibit 1020.2053

100 μM stem-loop linkers (SEQ ID NO:105 and SEQ ID NO:107) were dephosphorylated and reconcentrated to approximately 10 μM as follows:

20 μl of 100 μM (SEQ ID NO:105)

20 μl of 100 μM (SEQ ID NO:107)

100 μl of 10X phosphatase buffer (supplied by manufacturer, New England Biolabs)

800 μl water

80 μl of Antarctic phosphatase.

The reaction was incubated at 37°C for one hour and heat inactivated at 65°C for 5 minutes. The reaction mixture was then split into two tubes and precipitated by adding 3 μl Glyco-blue (Ambion Catalog #AM9516), 60 μl of 3 M NaOAc pH 5.2, and 1200 μl of ethanol per tube, mixed and centrifuged at 12K at 4°C for 20 minutes. The solvent was aspirated away from the pellet, which was resuspended in 100 μl of water. The recovery of the oligo linker was in the range of approximately 50%, with a final concentration of about 10 μM, which was determined by nanodrop.

Sonication Treatment

In some embodiments of the method of the invention, genomic DNA was fragmented by sonication prior to DNAse I treatment as follows.

Genomic DNA was diluted in water or in a Tris buffer (2 μg DNA with 500 μL 50 mM Tris) without EDTA and without Mn^{++} (Note: EDTA will chelate the Mn^{++} ions needed by the DNAse I in the next step). If EDTA was present in the sonication buffer, then a clean-up step (e.g., Qiagen Qiaquick® column) was used to remove the EDTA prior to DNAse I treatment.

Sonication was carried out in a 1.5 mL tube in an ice-water slurry, such that the sonication instrument tip was inserted into the DNA containing solution at the depth of the 100 ul mark. Sonication for each sample was carried out for 4 minutes with an amplitude of 45%, pulse on=20 seconds, pulse off=50 seconds.

The sonicated sample was then treated with DNAse I as described below.

B. DNase I treatment of genomic DNA

As described in Example 1, it was determined that bovine pancreatic deoxyribonuclease I (DNase I) induces random double-stranded breaks in DNA in the absence of Mg^{++} and in the presence of Mn^{++}.

IDT Exhibit 1020.2054

DNase I Digestion:

20 µl (2 µg) of total human genomic DNA (Clontech)

10 µl 10X reaction buffer (50 mM Tris pH 7.6 0.5 mg/ml acetylated BSA)

12.5 µl 40 mM $MnCl_2$

47.5 µl water

10 µl DNase I (N.E.B. Cat. No. M0303S, diluted 1:1500* in 1X buffer (100 µl 10X buffer, 125 µl $MnCl_2$, and 775 µl water)

100 µl total volume

The DNase I reaction was incubated at room temperature for 10 minutes and stopped by the addition of 0.2 volumes of 100 mM EDTA and immediately transferred to ice.

*The dilution of DNAse was chosen to generate fragments averaging in length from about 50 to about 500 bp, which was determined using a DNase I dilution series as described in Example 1.

The reaction mixture was then purified over a Qiaquick® spin column (Qiagen), with a recovery of about 40% of the input DNA in about 200 µl, with a size cut-off below about 40 bp. The column purified DNA was then concentrated by precipitation and resuspended in water to a final concentration of 80 ng/µl.

C. Blunt-End Polishing, Ligation and PCR of Target Ligation Products

(i) The DNase I treated genomic DNA was blunt-end treated as follows:

10 µl DNAse treated genomic DNA (80 ng/µl)

10 µl 2X blunt buffer (NEB Quick Blunt Reaction #E120S)

4 µl 10X blunt buffer (NEB Quick Blunt Reaction #E120S)

4 µl dNTP (10 mM in each dNTP)

10.4 µl H_2O

1.6 µl T4 polymerase plus T4 polynucleotide kinase enzyme

40 µl total

The reaction was incubated at room temperature for 30 minutes, then heated at 70°C for 10 minutes to yield approximately at least 40 ng/µl of DNase I treated and blunt end-polished genomic DNA that was ready for ligation to the stem-loop linkers.

IDT Exhibit 1020.2055

(ii) Ligations between phosphatase treated stem-loop linkers and DNase I treated genomic DNA were carried out as follows:

10 µl 2X buffer (NEB Quick Ligation kit (#M2200S)

2 µl DNase I treated and blunt end polished gDNA (40 ng/µl)

4 µl (SEQ ID NO:105) first stem-loop linker #1, phosphatase treated (10 µM)

(Note: Separate ligation reactions were carried out with each uniquely bar-coded stem-loop linker #1)

4 µl (SEQ ID NO:107) second stem-loop linker #2, phosphatase treated (10 µM)

1 µl Quick ligase® (NEB Quick Ligation Kit) (#M2200S)

20 µl total

The ligation reaction was incubated at room temperature for 10 minutes (not heat inactivated) then diluted with 180 µl of TEzero (10 mM Tris pH 7.6 and 0.1 mM EDTA) and stored at -20°C or used in the PCR amplification step described below.

(Note: The overall concentration of vector plus insert was preferably between 1 to 10 µg/ml for efficient ligation. For ligation products with single insertions, vector:insert ratios between 2:1 and 6:1 were preferable. It was observed that vector:insert ratios below 2:1 resulted in lower ligation efficiency, while vector: insert ratios above 6:1 promoted multiple inserts.)

(iii) PCR Amplification of Ligation Reaction (Suppression PCR)

PCR was used to produce >5 µg of product for the first round of solution-based target capture and enrichment. In order to generate this amount of product, 4X 100 µl PCR reactions were carried out for each library generated.

First PCR primer: 5'-AATGATACGGCGACCACCGA-3' (SEQ ID NO:109)

Second PCR primer: 5'-CAAGCAGAAGACGGCATACG-3' (SEQ ID NO:110)

PCR Reaction Mixture with 5% DMSO:

10 µl DNA template (the ligation mixture described above diluted 10-fold with TEzero)

20 µl 5X buffer (supplied by the manufacturer with the EXPAND^{plus}® kit, Roche)

IDT Exhibit 1020.2056

10 µl 25 mM $MgCl_2$

10 µl 10 µM First PCR primer (SEQ ID NO:109)

10 µl 10 µM Second PCR primer (SEQ ID NO:110)

5 µl DMSO (100%)

5 µl dNTPs (10 mM in each dNTP)

30 µl H_2O

1 µl Taq Polymerase (Invitrogen)

1 µl EXPAND$^{PLUS®}$ Polymerase (Roche)

100 µl total

PCR Cycling Conditions:

1 cycle:

72°C for 1 minute; 94°C for 2 minutes. (Note: This step copies the sequence from the stem-loop linker to the complementary strand.)

10 cycles:

94°C for 30 sec

60°C for 30 sec

72°C for 1 minute

10 cycles:

94°C for 30 sec

60°C for 30 sec

72°C for 1 minute plus 10 sec/cycle

1 cycle:

72°C for 7 minutes

4°C hold

D. Evaluation of Library Quality

Two standard modes were used to evaluate library quality. The first was to load 100 ng of purified library PCR product on a 2% agarose gel followed by visual inspection of size distribution of the library. The minimum size range of the library was expected to be ≥130 bp, which is the sum of the adaptor sequences left over after PCR (90 bp) and the minimum insert size of 40 bp. Smaller bands are indicative of ligated adaptor dimers and libraries with detectable quantities of this material were rejected. We expected the bulk of library material to be a smear that ranges in size from 140 bp to 800 bp. Libraries

IDT Exhibit 1020.2057

dominated by longer fragments than this size range show poor capture performance and the resulting sequences are spread over a large area, requiring excessive sequencing to obtain the desired sequencing depth.

While visual inspection provided information on the bulk library characteristics, it could not be used to assess the content of sequences present.

To assess specific gene content of libraries, real-time PCR with TaqMan® assays were carried out as follows.

As a typical example, a 100 µl PCR reaction mixture was purified over a Qiaquick® column (Qiagen) and the DNA was quantified. As shown in FIGURE 3, the purified DNA comprises a sequence-ready library (at step 630), which can be directly sequenced (at step 640) or enriched for target sequences (as shown in FIGURE 3, steps 650-670) prior to sequence analysis.

The gene content of the library was measured and compared to a reference genomic DNA sample using qPCR. A panel of four gene-specific quantitative PCR (qPCR) assays were used (AKT1, KRAS, PIK3CA, and PTEN). The libraries and reference genomic DNA were adjusted to uniform 10 ng/µl concentrations prior to measurement.

qPCR Reaction Mixture:

200 µl 2X TaqMan master mix (provided by the manufacturer, Applied Biosystems)

100 µl H_2O

20 µl primer/probe

220 µl total

8 µl of the qPCR reaction mixture was then aliquoted into a well of a 384 well qPCR plate. 2 µl of DNA template was added. The qPCR reactions were conducted on an ABI 7900 real-time instrument in a 384 well format using the manufacturer's recommended PCR conditions over 40 cycles.

qPCR Results:

The Counts (cts) were converted to raw quantities with the following formula:

Raw Quantity = $10^{((\log 10(1/2)*Ct)+10)}$

The % raw abundance of library samples relative to the reference genomic DNA was then calculated. It was observed that the abundances of gene content in the libraries

IDT Exhibit 1020.2058

fell short of the reference for gene content. While not wishing to be bound by theory, the reason for this was believed to be two-fold--first, enzymatic shearing created a high likelihood of digesting within a qPCR TaqMan primer binding site, therefore sheared DNA would be expected to have a lower gene-specific activity than an unsheared reference genomic DNA control; second, the stem-loop linkers represented a substantial mass in the library (e.g., in a library with 100 bp inserts, half the mass of the library is adaptor). Therefore, a substantial portion of the mass of library DNA is comprised of ligated linkers.

TABLE 6 shows the relationship between insert size, the % of library composed of adaptor, and the TaqMan signal detected. The key point for assessing library quality was that gene content was readily detectable (target genomic DNA is present in the initial library and the insert size is ≥50 bp) and that insert size is not excessive, as judged by gel appearance combined with qPCR.

TABLE 6: Decrease in PCR Signal as a Function of Decreasing Insert Sizes

Average Insert Size	% Stem-Loop Linker	% of gDNA Signal
10	91	0.00
20	83	0.28
25	80	0.8
30	77	1.5
40	71	4
50	67	7
60	63	10
70	59	13
80	56	16
100	50	22
150	40	35
200	33	45
400	20	65
1000	9	84

IDT Exhibit 1020.2059

These quality assessments were applied to the nine sample libraries described in this example. The agarose gels (not shown) produced the desired size distribution of fragments, ranging in size from ≥130 bp to ≤800 bp; the majority of fragments were in the 200-400 bp size range.

The results of qPCR for the libraries using four of the genes are shown below in TABLE 7.

TABLE 7: qPCR Results of the Gene Content of Representative Libraries

%'s	lib1	lib2	lib3	lib4	lib5	lib6	lib7	lib8	lib9	100+ control	200+ control
Gene											
AKT1	5	32	53	11	4	11	14	5	12	11	38
KRAS	24	48	31	25	17	14	18	16	23	21	45
PIK3CA	6	10	34	13	18	16	7	5	14	24	34
PTEN	7	34	35	15	7	15	19	13	19	16	39
average	11	31	38	16	12	14	14	10	17	18	39
insert size	60	150	150	80	70	70	70	60	80	80	150

The qPCR signal for genes, shown in rows in TABLE 7, is reported as a percent of the signal detected in unsheared genomic DNA. The combined values for four genes and the numerical averages for each library are shown in columns. The 100+ control and 200+ controls corresponded to well-characterized genomic libraries with known insert sizes and gene content. Here it can be seen that the nine libraries reported as an example all produced qPCR measurements consistent with the creation of useful libraries. Gel analysis showed a desirable distribution of fragment sizes and qPCR gave consistent results showing gene content metrics comparable to the two well-characterized control samples. These results show that the genomic libraries generated using these methods have the desired insert size and gene content representative of the starting genomic DNA.

EXAMPLE 3

This example describes the use of solution-based capture from a sequencing-ready library generated from genomic DNA using biotinylated capture oligos in order to enrich the 52 coding exons from 5 target genes for subsequent resequencing studies.

IDT Exhibit 1020.2060

Rationale

In contrast to generation of resequencing libraries with PCR-generated starting material (e.g., as described in Example 1), the generation of a genomic library followed by solution-based sequence capture eliminates the need for the initial step of individually PCR-amplifying the regions of interest. Therefore, as shown in FIGURE 3, the use of solution-based capture requires manipulation of a single sample throughout the resequencing library construction process, regardless of the size or complexity of the target region that is being addressed. An additional advantage is that the capture of target sequences can be applied in several rounds, with PCR amplification of the enriched library fractions between steps. This allows for the creation of resequencing samples that are largely composed of target sequences.

As shown in FIGURE 4, the central basis of solution-based direct capture is the annealing of the library comprising ligation products 50A, 50B with sense 100 and anti-sense 100' capture probes, thereby forming a plurality of bi-molecular DNA complexes (at step B) between a target strand (e.g., 50A) and a target insert sequence-specific capture probe 100 comprising a moiety 110 that binds a capture reagent 400. Following annealing, these bi-molecular DNA complexes are bound by the capture reagent 400, such as streptavidin-coated 410 paramagnetic beads, which are then purified away from the bulk solution by magnetic retention to a magnetic source 500.

For example, as shown in more detail in FIGURE 4 at step A, a representative nucleic acid molecule 50A is shown that is a member of a library comprising a population of double-stranded nucleic acid molecules 50A, 50B. Each double-stranded nucleic acid molecule 50A, 50B in the library comprises an insert 10 with a candidate nucleic acid sequence flanked by a first linker region 20 and a second linker region 30.

Although this example was carried out using a library made from fragmented genomic DNA, it will be understood by those of skill in the art that the population of inserts 10 with candidate nucleic acid sequences for solution-based capture may be generated from genomic DNA or cDNA (as described in Example 2) or from PCR products (as described in Example 1).

As shown in FIGURE 4 at step A, a population of sense target capture probes 100 and a population of anti-sense target capture probes 100' are mixed with the denatured library comprising sense 50A, 50B nucleic acid molecules and anti-sense 50A', 50B' nucleic acid molecules. Each sense target capture probe 100 comprises a target-specific

IDT Exhibit 1020.2061

binding region 102 having a nucleic acid sequence that is substantially complementary to the sense strand of a target insert 10 of interest, and a region 104 for attaching a moiety 110 for binding to a capture reagent 400 (e.g., streptavidin-coated magnetic beads).

Similarly, each anti-sense target capture probe 100' comprises a target-specific binding region 102' having a nucleic acid sequence that is substantially complementary to the anti-sense-strand of a target insert 10' of interest, and a region 104 for attaching a moiety 110 for binding to a capture reagent 400 (e.g., streptavidin-coated magnetic beads).

In operation, as shown in FIGURE 4, step B, the target-specific binding region 102 of sense 100 or antisense 100' target capture probes bind to a substantially complementary nucleic acid sequence contained in an insert region 10 or 10' of a nucleic acid molecule 50 in the library. The moiety 110 (e.g., biotin) attached to the capture probe 100, 100' is then contacted with a capture reagent 400 (e.g., a magnetic bead) having a binding region 410 (e.g., streptavidin coating) and pulled out of solution with a sorting device 500 that binds to the capture reagent 400, such as a magnet.

The solution-based capture method may be used to enrich a library for target sequences of interest. For example, as shown in FIGURE 3, a sequencing-ready library generated from total genomic DNA 630, generated using the methods described, *supra*, includes a population of double-stranded nucleic acid molecules 50, each double-stranded nucleic acid molecule 50 comprising an insert 10 having a candidate nucleic acid sequence flanked by a first linker region 20 and a second linker region 30. Within the population of double-stranded nucleic acid molecules 50 in the library, there exists a subpopulation of molecules 50 that contain inserts 10 with target nucleic acid sequences within a greater population of molecules 50 that contain inserts 10 with non-target nucleic acid sequences. The subpopulation of molecules 50 that contain inserts 10 with target nucleic acid sequences may be captured in solution from the starting non-enriched genomic library using capture probes, leaving behind the larger population of molecules 50 that contain inserts 10 with non-target sequences.

With continued reference to FIGURE 3, the non-enriched starting genomic DNA library 630, which is used in the first round of target capture, typically contains very few target sequences 10 in comparison to non-target sequences. In the solution-based capture methods at 650, in the first round of enrichment, the capture oligo probes are typically

IDT Exhibit 1020.2062

present in a molar excess in the first and second rounds of enrichment. An optional third round of enrichment may also be carried out which contains an excess amount of capture oligo probes that is reduced about 10-fold from the amount of capture oligo probes used in the second round of enrichment. Alternatively, a third round of enrichment may be carried out with a limiting amount of capture probe in order to normalize the content of the library (data not shown).

Methods

The libraries containing nucleic acid molecules with inserts containing target sequences of interest were generated as described above in Example 2, starting with genomic DNA, DNase I treating, blunt end polishing, and ligating on stem-loop linkers (SEQ ID NO:105 and SEQ ID NO:107), followed by 20 cycles of PCR and purification over a Qiaquick® column.

Solution-based Capture Using Biotinylated Sequence Specific Oligonucleotide Capture Probes:

Capture probes

A set of sense and antisense biotinylated capture oligos were generated that target the exons in the 5-gene set--AKT1, KRAS, PIK3CA, PTEN, and TP53--as shown below in TABLE 8. For exons less than 70 nucleotides in length, two sense oligos were synthesized. For exons of intermediate size (e.g., 70 nt to 200 nt), (referred to as "100+") alternating targeting oligos evenly spaced on opposite strands were chosen. For regions longer than 200 nt (referred to as "200+"), alternating targeting oligos spaced at intervals of about 45 nt to 65 nt were chosen. The potential collection of capture oligos sequences were screened for sequences that would anneal to multiple places in the human reference genome. Such oligos were removed from the synthesis list and replaced by oligos that would be expected to anneal to a nearby site with more unique sequence characteristics.

The oligos were synthesized by Operon and provided at a concentration of 100 μM. The biotinylated oligos were pooled for subsequent validation of the solution-based capture methodology.

IDT Exhibit 1020.2063

TABLE 8: Biotinylated Capture Oligos for Direct Capture (50-mers)

Oligo Reference ID	Sequence	SEQ ID No:
AKT_2_1_S	[BioTEG]TGGGCTCGGGGAGCGCCAGCCTGAGAGGAGCGCGTGAGCGTC GCGGGAGC	111
AKT_2_2_AS	[BioTEG]AACCCTCCTTCACAATAGCCACGTCGCTCATGGTGCCCGAGGC TCCCGCG	112
AKT_3_3_S	[BioTEG]ACCTGGCGGCCACGCTACTTCCTCCTCAAGAATGATGGCACCT TCATTGG	113
AKT_3_4_AS	[BioTEG]GTTGAGGGGAGCCTCACGTTGGTCCACATCCTGCGGCCGCTCC TTGTAGC	114
AKT_4_5_S	[BioTEG]CTGATGAAGACGGAGCGGCCCCGGCCCAACACCTTCATCATCC GCTGCCT	115
AKT_4_6_AS	[BioTEG]CTCCACATGGAAGGTGCGTTCGATGACAGTGGTCCACTGCAGG CAGCGGA	116
AKT_5_7_S	[BioTEG]TGTGGCTGACGGCCTCAAGAAGCAGGAGGAGGAGGAGATGGA CTTCCGGT	117
AKT_5_8_AS	[BioTEG]CCATCTCTTCAGCCCCTGAGTTGTCACTGGGTGAGCCCGACCG GAAGTCC	118
AKT_6_9_S	[BioTEG]CCTGAAGCTGCTGGGCAAGGGCACTTTCGGCAAGGTGATCCTG GTGAAGG	119
AKT_6_10_AS	[BioTEG]TCTTGAGGATCTTCATGGCGTAGTAGCGGCCTGTGGCCTTCTCC TTCACC	120
AKT_7_11_S	[BioTEG]ACGAGGTGGCCCACACACTCACCGAGAACCGCGTCCTGCAGAA CTCCAGG	121
AKT_7_12_S	[BioTEG]CACTCACCGAGAACCGCGTCCTGCAGAACTCCAGGCACCCCTT CCTCACA	122
AKT_8_13_S	[BioTEG]GCCCTGAAGTACTCTTTCCAGACCCACGACCGCCTCTGCTTTGT CATGGA	123
AKT_8_14_S	[BioTEG]AGACCCACGACCGCCTCTGCTTTGTCATGGAGTACGCCAACGG GGGCGAG	124
AKT_9_15_S	[BioTEG]CGGGAGCGTGTGTTCTCCGAGGACCGGGCCCGCTTCTATGGCG CTGAGAT	125
AKT_9_16_AS	[BioTEG]CACGTTCTTCTCCGAGTGCAGGTAGTCCAGGGCTGACACAATC TCAGCGC	126
AKT_10_17_S	[BioTEG]ATGCTGGACAAGGACGGGCACATTAAGATCACAGACTTCGGGC TGTGCAA	127
AKT_10_18_AS	[BioTEG]GGTGTGCCGCAAAAGGTCTTCATGGTGGCACCGTCCTTGATCC CCTCCTT	128
AKT_11_19_S	[BioTEG]AGTGGACTGGTGGGGGCTGGGCGTGGTCATGTACGAGATGATG TGCGGTC	129
AKT_11_20_AS	[BioTEG]CATGAGGATGAGCTCAAAAAGCTTCTCATGGTCCTGGTTGTAG AAGGGCA	130

IDT Exhibit 1020.2064

Oligo Reference ID	Sequence	SEQ ID No:
AKT_11_21_S	[BioTEG]AGATCCGCTTCCCGCGCACGCTTGGTCCCGAGGCCAAGTCCTTGCTTTCA	131
AKT_12_22_S	[BioTEG]GCTTGGCGGGGGCTCCGAGGACGCCAAGGAGATCATGCAGCATCGCTTCT	132
AKT_12_23_AS	[BioTEG]CTTCTTCTCGTACACGTGCTGCCACACGATACCGGCAAAGAAGCGATGCT	133
AKT_13_24_S	[BioTEG]CCCACCCTTCAAGCCCCAGGTCACGTCGGAGACTGACACCAGGTATTTTG	134
AKT_13_25_AS	[BioTEG]GGTCAGGTGGTGTGATGGTGATCATCTGGGCCGTGAACTCCTCATCAAAA	135
AKT_14_26_S	[BioTEG]ATGACAGCATGGAGTGTGTGGACAGCGAGCGCAGGCCCCACTTCCCCCAG	136
AKT_14_27_AS	[BioTEG]GCGCAGTCCACCGCCGCCTCAGGCCGTGCCGCTGGCCGAGTAGGAGAACT	137
KRAS_2_1_S	[BioTEG]TGAAAATGACTGAATATAAACTTGTGGTAGTTGGAGCTGGTGGCGTAGGC	138
KRAS_2_2_AS	[BioTEG]ATTCGTCCACAAAATGATTCTGAATTAGCTGTATCGTCAAGGCACTCTTG	139
KRAS_3_3_S	[BioTEG]GTCTCTTGGATATTCTCGACACAGCAGGTCAAGAGGAGTACAGTGCAATG	140
KRAS_3_4_AS	[BioTEG]TGGCAAATACACAAAGAAAGCCCTCCCCAGTCCTCATGTACTGGTCCCTC	141
KRAS_4_5_S	[BioTEG]TGTACCTATGGTCCTAGTAGGAAATAAATGTGATTTGCCTTCTAGAACAG	142
KRAS_4_6_AS	[BioTEG]AGGAATTCCATAACTTCTTGCTAAGTCCTGAGCCTGTTTTGTGTCTACTG	143
KRAS_5_7_S	[BioTEG]TGCCTTCTATACATTAGTTCGAGAAATTCGAAAACATAAAGAAAAGATGA	144
KRAS_5_8_AS	[BioTEG]TTACACACTTTGTCTTTGACTTCTTTTTCTTCTTTTTACCATCTTTGCTC	145
PIK3CA_2_1_S	[BioTEG]TTGATGCCCCCAAGAATCCTAGTAGAATGTTTACTACCAAATGGAATGAT	146
PIK3CA_2_2_AS	[BioTEG]CTTTATGGTTATTAATGTAGCCTCACGGAGGCATTCTAAAGTCACTATCA	147
PIK3CA_2_3_S	[BioTEG]GTGTTACTCAAGAAGCAGAAAGGGAAGAATTTTTTGATGAAACAAGACGA	148
PIK3CA_2_4_AS	[BioTEG]GTTCAATTACTTTTAAAAAGGGTTGAAAAAGCCGAAGGTCACAAAGTCGT	149
PIK3CA_3_5_S	[BioTEG]AGGACTTCCGAAGAAATATTCTGAACGTTTGTAAAGAAGCTGTGGATCTT	150
PIK3CA_3_6_A	[BioTEG]GAGGATAGACATACATTGCTCTACTATGAGGTGAATTGAGGTC	151

IDT Exhibit 1020.2065

Oligo Reference ID	Sequence	SEQ ID No:
S	CCTAAGA	
PIK3CA_4_7_S	[BioTEG]AAATAATGACAAGCAGAAGTATACTCTGAAAATCAACCATGAC TGTGTAC	152
PIK3CA_4_8_A S	[BioTEG]GTTCAGAGGATAGCAACATACTTCGAGTTTTTTTCCTGATTGCT TCAGCA	153
PIK3CA_4_9_S	[BioTEG]GTTTTAGAATATCAGGGCAAGTATATTTTAAAAGTGTGTGGAT GTGATGA	154
PIK3CA_5_10_S	[BioTEG]CAATTTGATGTTGATGGCTAAAGAAAGCCTTTATTCTCAACTGC CAATGG	155
PIK3CA_5_11_ AS	[BioTEG]CTCCATTCATATATGGTGTAGCTGTGGAAATGCGTCTGGAATA AGATGGC	156
PIK3CA_5_12_S	[BioTEG]ATCCCTTTGGGTTATAAATAGTGCACTCAGAATAAAAATTCTTT GTGCAA	157
PIK3CA_6_13_S	[BioTEG]ATCTATGTTCGAACAGGTATCTACCATGGAGGAGAACCCTTAT GTGACAA	158
PIK3CA_6_14_ AS	[BioTEG]CTGGGATTGGAACAAGGTACTCTTTGAGTGTTCACATTGTCAC ATAAGGG	159
PIK3CA_7_15_S	[BioTEG]ATGAATGGCTGAATTATGATATATACATTCCTGATCTTCCTCGT GCTGCT	160
PIK3CA_7_16_ AS	[BioTEG]TAGCACCCTTTCGGCCTTTAACAGAGCAAATGGAAAGGCAAAG TCGAGCA	161
PIK3CA_8_17_S	[BioTEG]TAAACTTGTTTGATTACACAGACACTCTAGTATCTGGAAAAAT GGCTTTG	162
PIK3CA_8_18_ AS	[BioTEG]TAGGGTTCAGCAAATCTTCTAATCCATGAGGTACTGGCCAAAG ATTCAAA	163
PIK3CA_9_19_S	[BioTEG]GGAGTTTGACTGGTTCAGCAGTGTGGTAAAGTTCCCAGATATG TCAGTGA	164
PIK3CA_9_20_ AS	[BioTEG]TAAATCCTGCTTCTCGGGATACAGACCAATTGGCATGCTCTTCA ATCACT	165
PIK3CA_10_21_ S	[BioTEG]AGAGACAATGAATTAAGGGAAAATGACAAAGAACAGCTCAAA GCAATTTC	166
PIK3CA_10_22_ AS	[BioTEG]AAAATCTTTCTCCTGCTCAGTGATTTCAGAGAGAGGATCTCGTG TAGAAA	167
PIK3CA_11_23_ S	[BioTEG]ACACTATTGTGTAACTATCCCCGAAATTCTACCCAAATTGCTTC TGTCTG	168
PIK3CA_11_24_ S	[BioTEG]CCAAATTGCTTCTGTCTGTTAAATGGAATTCTAGAGATGAAGT AGCCCAG	169
PIK3CA_12_25_ S	[BioTEG]AACCTGAACAGGCTATGGAACTTCTGGACTGTAATTACCCAGA TCCTATG	170
PIK3CA_12_26_ AS	[BioTEG]CATCTGTTAAATATTTTTCCAAGCACCGAACAGCAAAACCTCG AACCATA	171

IDT Exhibit 1020.2066

Oligo Reference ID	Sequence	SEQ ID No:
PIK3CA_13_27_S	[BioTEG]TAAAATATGAACAATATTTGGATAACTTGCTTGTGAGATTTTTACTGAAG	172
PIK3CA_13_28_AS	[BioTEG]AATGCCAAAAGAAAAAGTGCCCAATCCTTTGATTAGTCAATGCTTTCTTC	173
PIK3CA_14_29_S	[BioTEG]GCCTGCTTTTGGAGTCCTATTGTCGTGCATGTGGGATGTATTTGAAGCAC	174
PIK3CA_14_30_AS	[BioTEG]CAGTTAAGTTAATGAGCTTTTCCATTGCCTCGACTTGCCTATTCAGGTGC	175
PIK3CA_15_31_S	[BioTEG]GATGAAGTTTTTAGTTGAGCAAATGAGGCGACCAGATTTCATGGATGCTC	176
PIK3CA_15_32_AS	[BioTEG]TTTCCTAGTTGATGAGCAGGGTTTAGAGGAGACAGAAAGCCCTGTAGAGC	177
PIK3CA_16_33_S	[BioTEG]CGAATTATGTCCTCTGCAAAAAGGCCACTGTGGTTGAATTGGGAGAACCC	178
PIK3CA_16_34_AS	[BioTEG]AAAGATGATCTCATTGTTCTGAAACAGTAACTCTGACATGATGTCTGGGT	179
PIK3CA_17_35_S	[BioTEG]ATTTACGGCAAGATATGCTAACACTTCAAATTATTCGTATTATGGAAAAT	180
PIK3CA_17_36_S	[BioTEG]ATTATTCGTATTATGGAAAATATCTGGCAAAATCAAGGTCTTGATCTTCG	181
PIK3CA_18_37_S	[BioTEG]GTGGGACTTATTGAGGTGGTGCGAAATTCTCACACTATTATGCAAATTCA	182
PIK3CA_18_38_AS	[BioTEG]TGTGTGGCTGTTGAACTGCAGTGCACCTTTCAAGCCGCCTTTGCACTGAA	183
PIK3CA_19_39_S	[BioTEG]CCATTGACCTGTTTACACGTTCATGTGCTGGATACTGTGTAGCTACCTTC	184
PIK3CA_19_40_AS	[BioTEG]CTTTCACCATGATGTTACTATTGTGACGATCTCCAATTCCCAAAATGAAG	185
PIK3CA_20_41_S	[BioTEG]TGGATCACAAGAAGAAAAAATTTGGTTATAAACGAGAACGTGTGCCATTT	186
PIK3CA_20_42_AS	[BioTEG]CTTGGGCTCCTTTACTAATCACTATTAAGAAATCCTGTGTCAAAACAAAT	187
PIK3CA_21_43_S	[BioTEG]ATGCCAATCTCTTCATAAATCTTTTCTCAATGATGCTTGGCTCTGGAATG	188
PIK3CA_21_44_AS	[BioTEG]CTCAGTTTTATCTAAGGCTAGGGTCTTTCGAATGTATGCAATGTCATCAA	189
PIK3CA_21_45_S	[BioTEG]TTTCATGAAACAAATGAATGATGCACATCATGGTGGCTGGACAACAAAAA	190
PTEN_1_1_S	[BioTEG]TTCAGCCACAGGCTCCCAGACATGACAGCCATCATCAAAGAGATCGTTAG	191

IDT Exhibit 1020.2067

Oligo Reference ID	Sequence	SEQ ID No:
PTEN_1_2_AS	[BioTEG]GGTCAAGTCTAAGTCGAATCCATCCTCTTGATATCTCCTTTTGT TTCTGC	192
PTEN_2_3_S	[BioTEG]ATATTTATCCAAACATTATTGCTATGGGATTTCCTGCAGAAAGA CTTGAA	193
PTEN_2_4_AS	[BioTEG]CTTACTACATCATCAATATTGTTCCTGTATACGCCTTCAAGTCT TTCTGC	194
PTEN_3_5_S	[BioTEG]ttaagGTTTTTGGATTCAAAGCATAAAAACCATTACAAGATATAC AATCT	195
PTEN_3_6_S	[BioTEG]GTTTTTGGATTCAAAGCATAAAAACCATTACAAGATATACAAT CTgtaag	196
PTEN_4_7_S	[BioTEG]ttttagTTGTGCTGAAAGACATTATGACACCGCCAAATTTAATTGC AGAG	197
PTEN_4_8_S	[BioTEG]TTGTGCTGAAAGACATTATGACACCGCCAAATTTAATTGCAGA Ggtaggt	198
PTEN_5_9_S	[BioTEG]AGCTAGAACTTATCAAACCCTTTTGTGAAGATCTTGACCAATG GCTAAGT	199
PTEN_5_10_AS	[BioTEG]CACCAGTTCGTCCCTTTCCAGCTTTACAGTGAATTGCTGCAACA TGATTG	200
PTEN_5_11_S	[BioTEG]GTGCATATTTATTACATCGGGGCAAATTTTTAAAGGCACAAGA GGCCCTA	201
PTEN_6_12_S	[BioTEG]GCGCTATGTGTATTATTATAGCTACCTGTTAAAGAATCATCTGG ATTATA	202
PTEN_6_13_AS	[BioTEG]GGAATAGTTTCAAACATCATCTTGTGAAACAACAGTGCCACTG GTCTATA	203
PTEN_7_14_S	[BioTEG]ATATATTCCTCCAATTCAGGACCCACACGACGGGAAGACAAGT TCATGTA	204
PTEN_7_15_AS	[BioTEG]TACTTTGATATCACCACACACAGGTAACGGCTGAGGGAACTCA AAGTACA	205
PTEN_8_16_S	[BioTEG]TCTTCATACCAGGACCAGAGGAAACCTCAGAAAAAGTAGAAA ATGGAAGT	206
PTEN_8_17_AS	[BioTEG]CTTGTCATTATCTGCACGCTCTATACTGCAAATGCTATCGATTT CTTGAT	207
PTEN_8_18_S	[BioTEG]CTAGTACTTACTTTAACAAAAAATGATCTTGACAAAGCAAATA AAGACAA	208
PTEN_9_19_S	[BioTEG]GAGGCTAGCAGTTCAACTTCTGTAACACCAGATGTTAGTGACA ATGAACC	209
PTEN_9_20_AS	[BioTEG]ATTCTCTGGATCAGAGTCAGTGGTGTCAGAATATCTATAATGA TCAGGTT	210
TP53_2_1_S	[BioTEG]CCAGACTGCCTTCCGGGTCACTGCCATGGAGGAGCCGCAGTCA GATCCTA	211

IDT Exhibit 1020.2068

Oligo Reference ID	Sequence	SEQ ID No:
TP54_2_2_AS	[BioTEG]TTCCATAGGTCTGAAAATGTTTCCTGACTCAGAGGGGGCTCGACGCTAGG	212
TP55_3_3_S	[BioTEG]ggactgactttctgctcttgtctttcagACTTCCTGAAAACAACGTTCTG	213
TP56_3_4_S	[BioTEG]ACTTCCTGAAAACAACGTTCTGgtaaggacaagggttgggctggggacct	214
TP57_4_5_S	[BioTEG]GGACGATATTGAACAATGGTTCACTGAAGACCCAGGTCCAGATGAAGCTC	215
TP58_4_6_AS	[BioTEG]GGTGCAGGGGCCGCCGGTGTAGGAGCTGCTGGTGCAGGGGCCACGGGGGG	216
TP57_4_7_S	[BioTEG]ATCTTCTGTCCCTTCCCAGAAAACCTACCAGGGCAGCTACGGTTTCCGTC	217
TP59_5_8_S	[BioTEG]CTGCCCTGTGCAGCTGTGGGTTGATTCCACACCCCCGCCCGGCACCCGCG	218
TP60_5_9_AS	[BioTEG]ACCTCCGTCATGTGCTGTGACTGCTTGTAGATGGCCATGGCGCGGACGCG	219
TP61_6_10_S	[BioTEG]CCTCCTCAGCATCTTATCCGAGTGGAAGGAAATTTGCGTGTGGAGTATTT	220
TP62_6_11_AS	[BioTEG]CTCATAGGGCACCACCACACTATGTCGAAAAGTGTTTCTGTCATCCAAAT	221
TP63_7_12_S	[BioTEG]CTGACTGTACCACCATCCACTACAACTACATGTGTAACAGTTCCTGCATG	222
TP64_7_13_AS	[BioTEG]CTTCCAGTGTGATGATGGTGAGGATGGGCCTCCGGTTCATGCCGCCCATG	223
TP65_8_14_S	[BioTEG]ACAGCTTTGAGGTGCGTGTTTGTGCCTGTCCTGGGAGAGACCGGCGCACA	224
TP66_8_15_AS	[BioTEG]GCAGCTCGTGGTGAGGCTCCCCTTTCTTGCGGAGATTCTCTTCCTCTGTG	225
TP67_9_16_S	[BioTEG]CACTGCCCAACAACACCAGCTCCTCTCCCCAGCCAAAGAAGAAACCACTG	226
TP68_9_17_S	[BioTEG]CTCCCCAGCCAAAGAAGAAACCACTGGATGGAGAATATTTCACCCTTCAG	227
TP69_10_18_S	[BioTEG]TGGGCGTGAGCGCTTCGAGATGTTCCGAGAGCTGAATGAGGCCTTGGAAC	228
TP70_10_19_AS	[BioTEG]TGAGCCCTGCTCCCCCCTGGCTCCTTCCCAGCCTGGGCATCCTTGAGTTC	229
TP71_11_20_S	[BioTEG]CCACCTGAAGTCCAAAAAGGGTCAGTCTACCTCCCGCCATAAAAAACTCA	230
TP72_11_21_AS	[BioTEG]AGAAGTGGAGAATGTCAGTCTGAGTCAGGCCCTTCTGTCTTGAACATGAG	231

(BioTEG indicates that the above oligonucleotides were biotinylated at the 5' end).

IDT Exhibit 1020.2069

Preparation of Beads

InVitrogen's Dynabeads MyOne® Streptavidin C1 magnetic beads (InVitrogen #650-01) were used (which have a binding capacity of ~50 pmol of biotinylated dsDNA/50 µl beads). 120 µl beads were combined with 500 µl 2X binding buffer (20 mM Tris pH 7.6, 0.2 mM EDTA, 2M NaCl) and 380 µl water. The beads were pulled with a magnet and washed twice with 1 ml of 1X binding buffer and resuspended in 1200 µl 1X binding buffer.

A. First Round of Solution-Based Capture Using Biotinylated Sequence Specific Oligo Probes to Generate a Once-enriched Genomic DNA Library

A range of pooled biotinylated target-specific capture oligos (SEQ ID NO:111-231) was tested at the following concentrations (10 pmol, 1 pmol, 100 attomol, 10 attomol, 1 attomol, no oligo control). Two different wash buffers were also tested: (1X binding buffer; (high salt): 10 mM Tris pH 7.6, 0.1 mM EDTA, 1M NaCl) (osmolarity = 2000 molar) or TEzero (10 mM Tris pH 7.6 and 0.1 mM EDTA) = low salt (no NaCl) (osmolarity = 10 millimolar).

A dilution series was set up as follows. A first reaction mixture was prepared with 222 µl (10 µg) of PCR product (genomic library), 277.5 µl of 2X binding buffer (20 mM Tris pH 7.6, 0.2 mM EDTA, 2M NaCl), 22.2 µl of 1 µM pooled biotinylated oligos (20 pmol) and 33.3 µl water. Four tubes were prepared with 200 µl PCR product, 250 µl 2X binding buffer and 50 µl water. Serial 10-fold dilutions were then made of 55 µl of the first reaction mixture with the biotinylated oligo through the series of the 4 non-biotin containing tubes. A control was prepared with 200 µl PCR product, 250 µl 2X binding buffer and 50 µl water.

For the solution-based capture, 10 µl of the 1 µM pooled capture oligos were combined with 50 µl of 100 ng/µl genomic library (or a pool of genomic libraries containing 625 ng of each of eight genomic libraries ligated to a particular barcode), 125 µl of 2X binding buffer and 65 µl water, for a total volume of 250 µl.

The reaction mixture was annealed as follows:

94°C for 30 sec

90°C for 30 sec

85°C for 30 sec

80°C for 30 sec

IDT Exhibit 1020.2070

75°C for 30 sec

70°C for 30 sec

65°C for 30 sec

60°C for 30 sec

55°C for 30 sec

50°C for 30 sec

45°C for 30 sec

40°C for 30 sec

After the last annealing temperature, the cycler was allowed to come to room temperature. The 250 µl annealed mixture was combined with 100 µl washed beads and 150 µl 1X binding buffer and incubated at room temperature for 15 minutes. The beads were pulled out from the mixture with a magnet. The beads were then washed four times with either:

(1) 500 µl 1X binding buffer (10 mM Tris pH 7.6, 0.1 mM EDTA, 1M NaCl); or

(2) 500 µl TEzero (10 mM Tris pH 7.6 and 0.1 mM EDTA, no NaCl)

For washing, the beads were resuspended in either 1X binding buffer or TEzero and rocked for 5 minutes prior to pull down. This washing process was carried out four times. The washed beads were then eluted by resuspending them in 50 µl water, heating to 94°C for 30 seconds, then the beads were pulled with a magnet and the supernatant was removed. The elution process was repeated with another 50 µl of water, giving a total volume of 100 µl eluate that contained the enriched fragment library.

B. Amplification of the Eluted Once-enriched Library to Generate an Amplified Once-enriched Genomic DNA Library

The eluate containing the enriched fragment library was PCR amplified as follows:

PCR Reaction Mixture (5% DMSO):

28 µl H_2O

20 µl 5X buffer (supplied by the manufacturer with the EXPAND^{plus}® kit, Roche)

10 µl 25 mM $MgCl_2$

10 µl template (1/10th of total eluate of once-enriched library)

5 µl dNTPs (10 mM each dNTP)

IDT Exhibit 1020.2071

5 µl DMSO

10 µl 10 µM Forward PCR primer (SEQ ID NO:109)

10 µl 10 µM Reverse PCR primer (SEQ ID NO:110)

1 µl Taq polymerase

1 µl EXPAND$^{plus®}$ polymerase, Roche

100 µl total volume

PCR cycling Conditions

1 cycle:

94°C for 2 minutes

10 cycles:

94°C for 30 sec

60°C for 30 sec

72°C for 1 minute

15 cycles:

94°C for 30 sec

60°C for 30 sec

72°C for 1 minute plus 10 sec/cycle

1 cycle:

72°C for 7 minutes

4°C hold

The PCR reaction products were purified over a Qiaquick® column and quantified.

1 µl of PCR product was analyzed on a 2% agarose gel.

Analysis

The PCR products were analyzed by gene specific qPCR assays to determine the specific activity of target fragments in the enriched, amplified libraries.

IDT Exhibit 1020.2072

TABLE 9: Enhancement of Target-Specific Fragments in Library After Solution-Based Capture

	AKT1 (fold enhance-ment)	KRAS (fold enhance-ment)	PIK3CA (fold enhance-ment)	PTEN (fold enhance-ment)	TP53 (fold enhance-ment)
Capture conditions					
1X bind/10 pmol	33	24	18	24	22
1X bind/1 pmol	33	21	18	26	33
1X bind/100 amol	24	12	9	15	21
1Xbind/10 amol	4	4	2	4	3
1X bind/1 amol	0	1	1	1	1
1X bind/0 control	1.2	1.1	0.5	1.0	1.0
TE/10 pmol	755	675	703	762	488
TE/1 pmol	847	728	740	861	678
TE/100 amol	359	336	248	388	366
TE/10 amol	61	79	77	52	128
TE/1 amol	0	6	0	0	4
TE/0 control	2.6	0.1	0.1	0.1	0.3

Results

Two different wash buffers were also tested: 1X binding buffer (high salt): 10 mM Tris pH 7.6, 0.1 mM EDTA, 1M NaCl) (osmolarity = 2000 molar) or TEzero (10 mM Tris pH 7.6 and 0.1 mM EDTA) = low salt (no NaCl) (osmolarity = 10 millimolar).

As shown above in TABLE 9, modest enrichment was observed with high capture oligo concentrations (10 pmol to 10 amol) and a high salt wash (1X binding buffer (high salt): 10 mM Tris pH 7.6, 0.1 mM EDTA, 1M NaCl) (osmolarity = 2000 molar). It was also observed that low salt washes TEzero (10 mM Tris pH 7.6 and 0.1 mM EDTA) = low salt (no NaCl) (osmolarity = 10 millimolar) made a significant difference in enrichment specificity, which was oligo concentration dependent and very uniform across the five TaqMan assays. In this regard, it is noted that the Tris buffer in TE (low salt conditions) stabilizes the solution pH and the DNA duplex, but does not have the electrostatic effect of adding a monovalent cation anion such as NaCl. In

IDT Exhibit 1020.2073

contrast, the monovalent cation anion NaCl was observed to have a negative effect on stringency and enrichment.

This experimental data indicates that capture oligonucleotide concentrations in the range of 1.0 to 10 pmol are optimal for capture for 5 μg of input genomic DNA. Given that the capture was performed in 1 ml, this corresponds to a concentration of 500 ng/ml DNA target and 1 nM to 10 nM capture oligo. This data also indicates that low salt wash (TE (10 mM Tris pH 7.6, 0.1 mM EDTA) is a superior wash buffer over a high salt wash (10 mM Tris pH 7.6, 0.1 mM EDTA, 1 M NaCl).

The theoretical maximum of fold enrichment is 3,000,000 Kbp-human genome/20 Kb target region = 150,000-fold. As shown above in TABLE 9, the level of target fragment specific enrichment achieved after one round of capture was in the range of 500- to 900-fold using the low salt buffer wash conditions. This motivated the following experiment in which it was determined whether a second round of capture using the first round material as input could further enrich for target sequences. In the experiment that follows, low salt conditions TEzero (10 mM Tris pH 7.6 and 0.1 mM EDTA) = low salt (no NaCl) (osmolarity = 10 millimolar) were also applied during the annealing step. The results that follow indicated that annealing of the genomic DNA library to capture oligos in high salt buffer (1X binding buffer (high salt): 10 mM Tris pH 7.6, 0.1 mM EDTA, 1 M NaCl) (osmolarity = 2000 molar) followed by washing the bound material in a low salt wash buffer: TEzero (10 mM Tris pH 7.6 and 0.1 mM EDTA) = low salt (no NaCl) (osmolarity = 10 millimolar) worked best for enriching target sequences. Additionally, and importantly, it was determined that successive rounds of capture resulted in highly-enriched target sequences.

C. Second Round of Solution-Based Capture Using Biotinylated Sequence Specific Oligo Probes to Generate a Twice-enriched Genomic DNA Library

Preparation of Capture Beads

Two sets of Capture Beads were prepared as follows.

Set 1: low salt: 20 μl beads were combined with 480 μl TEzero (10 mM Tris pH 7.6, 0.1 mM EDTA). The beads were pulled out with a magnet, washed twice with 500 μl of TEzero, and resuspended in 500 μl TEzero low salt buffer (10 mM Tris pH 7.6 and 0.1 mM EDTA) = low salt (no NaCl) (osmolarity = 10 millimolar). 250 μl of washed beads/reaction were used.

IDT Exhibit 1020.2074

Set 2: high salt: 20 µl beads were combined with 250 2X binding buffer (1 M NaCl) and 230 µl water. The beads were pulled out with a magnet, washed twice with 500 µl of 1X binding buffer, and resuspended in 500 µl 1X binding buffer (high salt): 10 mM Tris pH 7.6, 0.1 mM EDTA, 1 M NaCl) (osmolarity = 2000 molar). 250 µl of washed beads/reaction were used.

Four samples were prepared for solution-based direct capture using the following libraries.

1. Once-enriched Library (TE/10 pmol - Table 9) annealed in high salt (1 M NaCl). 10 µl of 1 µM pooled biotinylated capture oligos (SEQ ID NOS:111-231) were combined with 67 µl of 75 ng/µl of once-enriched gDNA library (total of 5 µg DNA, produced in 20 cycles), 125 µl 2X binding buffer and 48 µl water for a total volume of 250 µl.
2. Starting Genomic DNA Library (not enriched), annealed in high salt (1 M NaCl). 10 µl of 1 µM pooled biotinylated capture oligos (SEQ ID NOS:111-231) were combined with 100 µl of 50 ng/µl starting genomic, non-enriched library (total of 5 µg DNA, produced in 20 cycles), 125 µl of 2X binding buffer and 15 µl of water for a total volume of 250 µl.
3. Once-enriched Library (TE/10 pmol - Table 9) annealed in low salt. 10 µl of 1 µM pooled biotinylated capture oligos (SEQ ID NOS:111-231) were combined with 67 µl of 75 ng/µl of once-enriched gDNA library (total of 5 µg DNA, produced in 20 cycles) and 173 µl of TEzero for a total volume of 250 µl.
4. Starting Genomic DNA Library (not enriched), annealed in low salt. 10 µl of 1 µM pooled biotinylated capture oligos (SEQ ID NOS:111-231) were combined with 100 µl of 50 ng/µl starting genomic, non-enriched library (total of 5 µg DNA, produced in 20 cycles) and 140 µl of TEzero for a total volume of 250 µl.

The reaction mixtures were each annealed as follows.

94°C for 30 sec

85°C for 30 sec

80°C for 30 sec

IDT Exhibit 1020.2075

75°C for 30 sec
70°C for 30 sec
65°C for 30 sec
60°C for 30 sec
55°C for 30 sec
50°C for 30 sec
45°C for 30 sec
40°C for 30 sec
35°C for 30 sec

Capture

250 µl of annealed mixture was combined with 250 µl of set 1 beads (washed in low salt) or with 250 µl of set 2 beads (washed in high salt). The mixture was incubated at room temperature with mixing for 15 minutes. The beads were pulled out with a magnet and washed four times with 500 µl of TEzero. For each wash step, the beads were resuspended and rocked for 5 minutes prior to pulling down with the magnet.

Elution

The washed beads were resuspended in 50 µl water, heated to 94°C for 30 seconds, pulled down with a magnet, and the supernatant with the bound DNA was collected. The process was repeated with an additional 50 µl, for a total eluate volume of 100 µl.

Amplification of Eluate:

PCR Reaction Mixture (5% DMSO):

28 µl H_2O
20 µl 5X buffer (supplied by the manufacturer with EXPAND$^{plus®}$ kit, Roche)
10 µl 25 mM $MgCl_2$
10 µl template (1/10th of the total eluted material from the twice-enriched library)
5 µl dNTPs (10 mM of each dNTP)
5 µl DMSO
10 µl 10 µM Forward PCR primer (SEQ ID NO:109)
10 µl 10 µM Reverse PCR primer (SEQ ID NO:110)
1 µl Taq polymerase
1 µl Expand polymerase

IDT Exhibit 1020.2076

100 µl total volume

PCR cycling Conditions

1 cycle:

94°C for 2 minutes

10 cycles:

94°C for 30 sec

60°C for 30 sec

72°C for 1 minute

10 or 15 cycles:

94°C for 30 sec

60°C for 30 sec

72°C for 1 minute plus 10 sec/cycle

1 cycle:

72°C for 7 minutes

4°C hold

The PCR reaction products were purified over a Qiaquick column and quantified. 1 µl of PCR product was analyzed on a 2% agarose gel.

D. Comparison of Sequence-Specific Capture From (1) A Genomic DNA Starting Library, (2) A Genomic DNA Library Once-enriched, and (3) A Genomic DNA Library Twice-enriched

The samples generated as described above were analyzed using qPCR to determine the level of enrichment achieved and the influence of salt concentration on the wash step.

20 ng of starting gDNA library (not enriched) and 20 pg of once- or twice-enriched samples were analyzed by qPCR as follows:

1. No template control
2. Starting material: gDNA library (amplified, not enriched)
3. Starting material: once-enriched gDNA library (TABLE 9: TE/10 pmol) diluted 1000-fold))
4. Low salt annealed, twice-enriched (diluted 1000-fold)
5. Low salt annealed, once-enriched (diluted 1000-fold)
6. High salt annealed, twice-enriched (diluted 1000-fold)

IDT Exhibit 1020.2077

7. High salt annealed, once-enriched (diluted 1000-fold)

The above samples were run on a 2% agarose gel and it was observed that all libraries had a reasonable size distribution, with fragments > about 130 nt in length (data not shown).

The TaqMan data from the 5-gene qPCR assay on the above samples was processed as follows. The raw counts (Cts) were converted to raw quantities calculated with the universal formula: $10^{(\log10(1/2)*Ct+10)}$.

The results shown below in TABLE 10 are adjusted by 1000-fold to be normalized with the 20 ng samples.

TABLE 10: qPCR Data (Normalized Counts)

	AKT1	KRAS	PIK3CA	PTEN	TP53
Sample Assayed					
gDNA pool (gDNA starting library, not target enriched)	467	362	339	461	3175
E-Pool (starting material, one round capture, Table 9 TE/10pmol)	705,442	245,007	176,163	485,587	1,813,181
Low salt annealed twice-enriched (enriched after two rounds capture under low salt conditions)	586,661	79,010	86,977	420,247	938,748
Low salt annealed once-enriched (enriched after one round of capture under low salt conditions)	20	287	21	135	2850
High salt annealed twice-enriched (enriched after two rounds capture under high salt conditions)	32,398,922	12,469,722	9,078,119	24,276,124	110,388,507
High salt annealed once-enriched (enriched after one round of capture under high salt conditions)	277,775	241,019	123,291	293,021	1,254,427
No Template Negative control	0	0	0	0	0

IDT Exhibit 1020.2078

The results shown in TABLE 11 are the ratios of the values shown in TABLE 10, as described in the first column of the table, in order to show the fold enrichment level.

TABLE 11: The Fold Enrichment of Various Libraries for Gene Target Content

		AKT1 (Ratio of fold enrichment observed)	KRAS	PIK3CA	PTEN	TP53
Row number	Ratio measured to determine fold enrichment					
1	High salt annealed once-enriched/gDNA starting library	595-fold	666-fold	364-fold	635-fold	395-fold
2	High salt annealed twice-enriched/gDNA starting library	69,448-fold	34,453-fold	26,797-fold	52,634-fold	34,763-fold
3	High salt twice-enriched pool/High salt once-enriched pool	46-fold	51-fold	52-fold	50-fold	61-fold

Discussion of Results

As shown above in TABLE 10, annealing in high salt (1X binding buffer: 10 mM Tris pH 7.6, 0.1 mM EDTA, 1 M NaCl) worked much better for library enrichment than annealing in low salt (10 mM Tris pH 7.6, 0.1 mM EDTA).

As shown above in TABLE 11, row number 1 is the ratio of high salt annealed once-enriched genomic pool/gDNA, which is a measure of a single round of enrichment from the starting genomic library (non enriched) to the enriched library, showing an average target enrichment level of approximately 500-fold for the 5 genes, which is very good.

Row number 2 of TABLE 11 shows an average of about 50,000-fold target enrichment in the high salt annealed, twice-enriched genomic library relative to the starting non-enriched library. This is a surprisingly successful achievement, given that theoretical perfection (3 billion bases human genome/20 kb target), would be an enrichment of 150,000-fold, which is only a factor of 3-6-fold difference. It is also noted that the approximately 50,000-fold enrichment is reasonably uniform across the five genes.

IDT Exhibit 1020.2079

Row number 3 of TABLE 11 shows that the second round of enrichment contributes substantially to the overall target enrichment process, contributing 50-fold more purification relative to a single round alone.

Another important feature demonstrated by the data in TABLES 10 and 11 is that all five monitored targets, chosen across the five genes in this study, are fairly uniformly enriched (within 2-fold or less). In contrast to the results observed using these methods, there are several reports that sequence based capture is severely hampered by unequal representation of target sequences. See, e.g., Albert, T.J., et al., *Nature Methods* 4(11):903-905 (2007); Okou, D.T., et al., *Nature Methods* 4(11):907-909 (2007); Porreca, G.J., et al., *Nature Methods* 4(11):931-6 (2007); and Hodges, E., et al., *Nature Genetics* 39:1522-2527 (2007).

The first and second rounds of enrichment described above were both carried out with a concentration of 500 ng/ml DNA target and 1 nM to 10 nM capture oligo.

Optional Third Round of Enrichment

A twice-enriched library can optionally be further enriched prior to sequence analysis by subjecting the twice-enriched library to one more round of solution-based capture. The use of another round of biotin capture on the amplified and enriched material serves to eliminate more of the off-target sequences that may have passed through the enrichment process, and may also be used to level or normalize the fragment representation in the library.

Methods

It was previously determined that 1 pmol of oligos (1 µl of a 1 µM solution) was sufficient to bind target sequences in a library generated using stem-loop linkers. 5 µg has $5x10^{-6}$ g/(160 bp average fragment size x 660 g/mol-bp) = 47 pmol of dsDNA library fragments. Therefore, hybridization of 5 µg (500 ng/ml) of the twice-enriched library with 1 pmol (1nM) of a pool of biotinylated capture oligos was used.

5 µg (39 µl) of the high salt annealed, twice-enriched library was combined with 1 µl of 1 µM biotinylated capture oligo pool (SEQ ID NOS:111-231), 125 µl of 2X binding buffer, and 85 µl water, for a total volume of 250 µl.

The reaction mixtures were annealed as follows:

94°C for 30 sec

90°C for 30 sec

IDT Exhibit 1020.2080

85°C for 30 sec
80°C for 30 sec
75°C for 30 sec
70°C for 30 sec
65°C for 30 sec
60°C for 30 sec
55°C for 30 sec
50°C for 30 sec
45°C for 30 sec
40°C for 30 sec

Capture

Washed beads were prepared by combining 10 µl beads, 125 µl 2X binding buffer, and 115 µl water. The beads were pulled over with a magnet, washed twice with 250 µl 1X binding buffer, and resuspended in 250 µl 1X binding buffer.

The annealed 250 µl mixture was combined with the 250 µl washed beads, mixed for 15 minutes, the beads were pulled over with a magnet, and the supernatant was decanted. The beads were then washed 4 times with TEzero (low salt).

Elution

The bound beads were eluted with two aliquots of 50 µl of water by incubation at 94°C for 30 seconds, pulling over the beads and removing the eluate, for a total eluate volume of 100 µl. Assuming 100% capture, the purified material should possess 1 pmol library/100 µl = 10 amol/µl. 2 µl was sequenced using a flow cell cluster sequencing platform, available from Illumina.

Results of third Round of Enrichment

A bioinformatic analysis of sequences derived from the twice-enriched pool versus the three-times enriched pool indicated that the third round of enrichment contributed an additional two-fold enrichment/purification of target sequences. In one experiment, it was observed that 25% of the sequencing reads from the twice-enriched pool aligned to the overall target region, whereas 50% of the sequencing reads from the three-times enriched pool aligned to the target region.

IDT Exhibit 1020.2081

EXAMPLE 4

This example describes solution-based capture using indirect capture via chimeric capture oligos with a gene-specific region and a region that hybridizes to a universal biotinylated adaptor oligo, with a set of indirect oligos that are specific for a set of 5 genes of interest.

Rationale

As demonstrated above in Example 3, the method of targeted sequence capture using biotinylated gene sequence specific oligonucleotides works well for its intended purpose of generating a sequencing library. However, the drawbacks to the use of biotinylated gene sequence specific oligonucleotides are that biotinylated oligos are expensive reagents to produce, they require a long time to synthesize, and the yields of oligonucleotides are generally low and unpredictably variable. An alternative approach is to use chimeric capture oligonucleotides where one portion of the capture oligonucleotides hybridizes to target sequences and one portion hybridizes to a common, biotinylated oligonucleotide, as shown in FIGURE 5. The chimeric capture oligonucleotides that are not biotinyated are straightforward to produce and the universal (i.e., common) biotinylated oligo is easily manufactured in a single large batch. In contrast to capture with directly biotinylated oligonucleotides, the advantage of the indirect capture approach is that only a single biotinylated oligonucleotide sequence needs to be synthesized, and the chimeric oligos are pure DNA oligos that are relatively inexpensive to synthesize.

As shown in FIGURE 5, an alternative approach for target gene enrichment of a genomic library is to use indirect capture by generating a chimeric capture probe 200, 200' with a first region 202 that hybridizes to a target nucleic acid sequence 10, 10' in the library and a second region 204 that hybridizes to a universal biotinylated oligo 300, mixing the chimeric oligo, the universal biotinylated oligo and the library containing a plurality of nucleic acid molecules 50 under hybridizing conditions to form a tri-molecular complex (i.e., 50/200/300), and using magnetic beads 400 coated with streptavidin 410 to bind to the biotinylated region 310 of the universal oligo 300 and pull out the target sequences 50 bound in the complex to the chimeric capture probe 200, using a magnet 400.

IDT Exhibit 1020.2082

This experiment compares library enrichment using biotinylated capture oligos 100 for direct capture versus chimeric capture oligos 200 that have a first region that hybridizes to a target sequence and a second region that hybridizes to a universal biotinylated oligo.

Methods

Oligonucleotides

A universal 5' biotinylated oligo was used:

5' [BioTEG] TAATTGCTCGAAGGGGTCCACATCCGCCACGCGT 3' (SEQ ID NO:232)

A set of chimeric capture oligos were generated that target AKT1, KRAS, PIK3CA, PTEN, and TP53, which were not biotinylated and that have a first 5' region with the identical sequence to the oligos shown above in TABLE 8, and a second 3' region consisting of the following additional sequence that hybridizes to universal oligo:

5' ACGCGTGGCGGATGTGGACCCCTTCGAGCAATTA 3' (SEQ ID NO:233)

A set of exemplary chimeric capture oligos is provided below in TABLE 12 that target AKT1, KRAS, PIK3CA, PTEN, and TP53 that contain a 5' first region (35 nt) that contains sequence that hybridizes to the target gene AKT1, and a 3' region (SEQ ID NO:232) (34 nt) that hybridizes to the universal biotinylated capture oligo (SEQ ID NO:233).

TABLE 12: Chimeric Capture Oligos Targeting 5 Genes of Interest

Reference ID	Sequence	SEQ ID NO:
AKT_2_1_S	TGGGCTCGGGGAGCGCCAGCCTGAGAGGAGCGCGTACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	234
AKT_2_2_AS	AACCCTCCTTCACAATAGCCACGTCGCTCATGGTGACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	235
AKT_3_3_S	ACCTGGCGGCCACGCTACTTCCTCCTCAAGAATGAACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	236
AKT_3_4_AS	GTTGAGGGGAGCCTCACGTTGGTCCACATCCTGCGACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	237
AKT_4_5_S	CTGATGAAGACGGAGCGGCCCCGGCCCAACACCTTACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	238

IDT Exhibit 1020.2083

Reference ID	Sequence	SEQ ID NO:
AKT_4_6_AS	CTCCACATGGAAGGTGCGTTCGATGACAGTGGTCCACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	239
AKT_5_7_S	TGTGGCTGACGGCCTCAAGAAGCAGGAGGAGGAGGACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	240
AKT_5_8_AS	CCATCTCTTCAGCCCCTGAGTTGTCACTGGGTGAGACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	241
AKT_6_9_S	CCTGAAGCTGCTGGGCAAGGGCACTTTCGGCAAGGACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	242
AKT_6_10_AS	TCTTGAGGATCTTCATGGCGTAGTAGCGGCCTGTGACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	243
AKT_7_11_S	ACGAGGTGGCCCACACACTCACCGAGAACCGCGTCACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	244
AKT_7_12_S	CACTCACCGAGAACCGCGTCCTGCAGAACTCCAGGACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	245
AKT_8_13_S	GCCCTGAAGTACTCTTTCCAGACCCACGACCGCCTACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	246
AKT_8_14_S	AGACCCACGACCGCCTCTGCTTTGTCATGGAGTACACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	247
AKT_9_15_S	CGGGAGCGTGTGTTCTCCGAGGACCGGGCCCGCTTACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	248
AKT_9_16_AS	CACGTTCTTCTCCGAGTGCAGGTAGTCCAGGGCTGACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	249
AKT_10_17_S	ATGCTGGACAAGGACGGGCACATTAAGATCACAGAACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	250
AKT_10_18_AS	GGTGTGCCGCAAAAGGTCTTCATGGTGGCACCGTCACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	251
AKT_11_19_S	AGTGGACTGGTGGGGGCTGGGCGTGGTCATGTACGACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	252
AKT_11_20_AS	CATGAGGATGAGCTCAAAAAGCTTCTCATGGTCCTACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	253
AKT_11_21_S	AGATCCGCTTCCCGCGCACGCTTGGTCCCGAGGCCACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	254
AKT_12_22_S	GCTTGGCGGGGGCTCCGAGGACGCCAAGGAGATCAACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	255
AKT_12_23_AS	CTTCTTCTCGTACACGTGCTGCCACACGATACCGGACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	256
AKT_13_24_S	CCCACCCTTCAAGCCCCAGGTCACGTCGGAGACTGACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	257
AKT_13_25_AS	GGTCAGGTGGTGTGATGGTGATCATCTGGGCCGTGACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	258

IDT Exhibit 1020.2084

Reference ID	Sequence	SEQ ID NO:
AKT_14_26_S	ATGACAGCATGGAGTGTGTGGACAGCGAGCGCAGGACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	259
AKT_14_27_AS	GCGCAGTCCACCGCCGCCTCAGGCCGTGCCGCTGGACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	260
KRAS_2_1_S	TGAAAATGACTGAATATAAACTTGTGGTAGTTGGAACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	261
KRAS_2_2_AS	ATTCGTCCACAAAATGATTCTGAATTAGCTGTATCACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	262
KRAS_3_3_S	GTCTCTTGGATATTCTCGACACAGCAGGTCAAGAGACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	263
KRAS_3_4_AS	TGGCAAATACACAAAGAAAGCCCTCCCCAGTCCTCACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	264
KRAS_4_5_S	TGTACCTATGGTCCTAGTAGGAAATAAATGTGATTACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	265
KRAS_4_6_AS	AGGAATTCCATAACTTCTTGCTAAGTCCTGAGCCTACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	266
KRAS_5_7_S	TGCCTTCTATACATTAGTTCGAGAAATTCGAAAACACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	267
KRAS_5_8_AS	TTACACACTTTGTCTTTGACTTCTTTTTCTTCTTTACGCGTGGCGGAT GTGGACCCCTTCGAGCAATTA	268
PIK3CA_2_1_S	TTGATGCCCCCAAGAATCCTAGTAGAATGTTTACTACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	269
PIK3CA_2_2_AS	CTTTATGGTTATTAATGTAGCCTCACGGAGGCATTACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	270
PIK3CA_2_3_S	GTGTTACTCAAGAAGCAGAAAGGGAAGAATTTTTTACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	271
PIK3CA_2_4_AS	GTTCAATTACTTTTAAAAAGGGTTGAAAAAGCCGAACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	272
PIK3CA_3_5_S	AGGACTTCCGAAGAAATATTCTGAACGTTTGTAAAACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	273
PIK3CA_3_6_AS	GAGGATAGACATACATTGCTCTACTATGAGGTGAAACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	274
PIK3CA_4_7_S	AAATAATGACAAGCAGAAGTATACTCTGAAAATCAACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	275
PIK3CA_4_8_AS	GTTCAGAGGATAGCAACATACTTCGAGTTTTTTTCACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	276
PIK3CA_4_9_S	GTTTTAGAATATCAGGGCAAGTATATTTTAAAAGTACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	277
PIK3CA_5_10_S	CAATTTGATGTTGATGGCTAAAGAAAGCCTTTATTACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	278

IDT Exhibit 1020.2085

Reference ID	Sequence	SEQ ID NO:
PIK3CA_5_11_AS	CTCCATTCATATATGGTGTAGCTGTGGAAATGCGTACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	279
PIK3CA_5_12_S	ATCCCTTTGGGTTATAAATAGTGCACTCAGAATAAACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	280
PIK3CA_6_13_S	ATCTATGTTCGAACAGGTATCTACCATGGAGGAGAACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	281
PIK3CA_6_14_AS	CTGGGATTGGAACAAGGTACTCTTTGAGTGTTCACACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	282
PIK3CA_7_15_S	ATGAATGGCTGAATTATGATATATACATTCCTGATACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	283
PIK3CA_7_16_AS	TAGCACCCTTTCGGCCTTTAACAGAGCAAATGGAAACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	284
PIK3CA_8_17_S	TAAACTTGTTTGATTACACAGACACTCTAGTATCTACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	285
PIK3CA_8_18_AS	TAGGGTTCAGCAAATCTTCTAATCCATGAGGTACTACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	286
PIK3CA_9_19_S	GGAGTTTGACTGGTTCAGCAGTGTGGTAAAGTTCCACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	287
PIK3CA_9_20_AS	TAAATCCTGCTTCTCGGGATACAGACCAATTGGCAACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	288
PIK3CA_10_21_S	AGAGACAATGAATTAAGGGAAAATGACAAAGAACAACGCGTGGC GGATGTGGACCCCTTCGAGCAATTA	289
PIK3CA_10_22_AS	AAAATCTTTCTCCTGCTCAGTGATTTCAGAGAGAGACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	290
PIK3CA_11_23_S	ACACTATTGTGTAACTATCCCCGAAATTCTACCCAACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	291
PIK3CA_11_24_S	CCAAATTGCTTCTGTCTGTTAAATGGAATTCTAGAACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	292
PIK3CA_12_25_S	AACCTGAACAGGCTATGGAACTTCTGGACTGTAATACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	293
PIK3CA_12_26_AS	CATCTGTTAAATATTTTTCCAAGCACCGAACAGCAACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	294
PIK3CA_13_27_S	TAAAATATGAACAATATTTGGATAACTTGCTTGTGACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	295
PIK3CA_13_28_AS	AATGCCAAAAGAAAAAGTGCCCAATCCTTTGATTAACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	296
PIK3CA_14_29_S	GCCTGCTTTTGGAGTCCTATTGTCGTGCATGTGGGACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	297
PIK3CA_14_30_AS	CAGTTAAGTTAATGAGCTTTTCCATTGCCTCGACTACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	298

IDT Exhibit 1020.2086

Reference ID	Sequence	SEQ ID NO:
PIK3CA_15_31_S	GATGAAGTTTTTAGTTGAGCAAATGAGGCGACCAGACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	299
PIK3CA_15_32_AS	TTTCCTAGTTGATGAGCAGGGTTTAGAGGAGACAGACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	300
PIK3CA_16_33_S	CGAATTATGTCCTCTGCAAAAAGGCCACTGTGGTTACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	301
PIK3CA_16_34_AS	AAAGATGATCTCATTGTTCTGAAACAGTAACTCTGACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	302
PIK3CA_17_35_S	ATTTACGGCAAGATATGCTAACACTTCAAATTATTACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	303
PIK3CA_17_36_S	ATTATTCGTATTATGGAAAATATCTGGCAAAATCAACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	304
PIK3CA_18_37_S	GTGGGACTTATTGAGGTGGTGCGAAATTCTCACACACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	305
PIK3CA_18_38_AS	TGTGTGGCTGTTGAACTGCAGTGCACCTTTCAAGCACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	306
PIK3CA_19_39_S	CCATTGACCTGTTTACACGTTCATGTGCTGGATACACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	307
PIK3CA_19_40_AS	CTTTCACCATGATGTTACTATTGTGACGATCTCCAACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	308
PIK3CA_20_41_S	TGGATCACAAGAAGAAAAAATTTGGTTATAAACGAACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	309
PIK3CA_20_42_AS	CTTGGGCTCCTTTACTAATCACTATTAAGAAATCCACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	310
PIK3CA_21_43_S	ATGCCAATCTCTTCATAAATCTTTTCTCAATGATGACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	311
PIK3CA_21_44_AS	CTCAGTTTTATCTAAGGCTAGGGTCTTTCGAATGTACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	312
PIK3CA_21_45_S	TTTCATGAAACAAATGAATGATGCACATCATGGTGACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	313
PTEN_1_1_S	TTCAGCCACAGGCTCCCAGACATGACAGCCATCATACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	314
PTEN_1_2_AS	GGTCAAGTCTAAGTCGAATCCATCCTCTTGATATCACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	315
PTEN_2_3_S	ATATTTATCCAAACATTATTGCTATGGGATTTCCTACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	316
PTEN_2_4_AS	CTTACTACATCATCAATATTGTTCCTGTATACGCCACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	317
PTEN_3_5_S	ttaagGTTTTTGGATTCAAAGCATAAAAACCATTAACGCGTGGCGGAT GTGGACCCCTTCGAGCAATTA	318

IDT Exhibit 1020.2087

Reference ID	Sequence	SEQ ID NO:
PTEN_3_6_S	GTTTTTGGATTCAAAGCATAAAAACCATTACAAGAACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	319
PTEN_4_7_S	ttttagTTGTGCTGAAAGACATTATGACACCGCCAACGCGTGGCGGATG TGGACCCCTTCGAGCAATTA	320
PTEN_4_8_S	TTGTGCTGAAAGACATTATGACACCGCCAAATTTAACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	321
PTEN_5_9_S	AGCTAGAACTTATCAAACCCTTTTGTGAAGATCTTACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	322
PTEN_5_10_AS	CACCAGTTCGTCCCTTTCCAGCTTTACAGTGAATTACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	323
PTEN_5_11_S	GTGCATATTTATTACATCGGGGCAAATTTTTAAAGACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	324
PTEN_6_12_S	GCGCTATGTGTATTATTATAGCTACCTGTTAAAGAACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	325
PTEN_6_13_AS	GGAATAGTTTCAAACATCATCTTGTGAAACAACAGACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	326
PTEN_7_14_S	ATATATTCCTCCAATTCAGGACCCACACGACGGGAACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	327
PTEN_7_15_AS	TACTTTGATATCACCACACACAGGTAACGGCTGAGACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	328
PTEN_8_16_S	TCTTCATACCAGGACCAGAGGAAACCTCAGAAAAAAACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	329
PTEN_8_17_AS	CTTGTCATTATCTGCACGCTCTATACTGCAAATGCACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	330
PTEN_8_18_S	CTAGTACTTACTTTAACAAAAAATGATCTTGACAAACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	331
PTEN_9_19_S	GAGGCTAGCAGTTCAACTTCTGTAACACCAGATGTACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	332
PTEN_9_20_AS	ATTCTCTGGATCAGAGTCAGTGGTGTCAGAATATCACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	333
TP53_2_1_S	CCAGACTGCCTTCCGGGTCACTGCCATGGAGGAGCACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	334
TP53_2_2_AS	TTCCATAGGTCTGAAAATGTTTCCTGACTCAGAGGACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	335
TP53_3_3_S	ggactgactttctgctcttgtctttcagACTTCCTACGCGTGGCGGATGTGGACCCCTT CGAGCAATTA	336
TP53_3_4_S	ACTTCCTGAAAACAACGTTCTGgtaaggacaagggACGCGTGGCGGATGT GGACCCCTTCGAGCAATTA	337
TP53_4_5_S	GGACGATATTGAACAATGGTTCACTGAAGACCCAGACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	338

IDT Exhibit 1020.2088

Reference ID	Sequence	SEQ ID NO:
TP53_4_6_AS	GGTGCAGGGGCCGCCGGTGTAGGAGCTGCTGGTGCACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	339
TP53_4_7_S	ATCTTCTGTCCCTTCCCAGAAAACCTACCAGGGCAACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	340
TP53_5_8_S	CTGCCCTGTGCAGCTGTGGGTTGATTCCACACCCCACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	341
TP53_5_9_AS	ACCTCCGTCATGTGCTGTGACTGCTTGTAGATGGCACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	342
TP53_6_10_S	CCTCCTCAGCATCTTATCCGAGTGGAAGGAAATTTACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	343
TP53_6_11_AS	CTCATAGGGCACCACCACACTATGTCGAAAAGTGTACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	344
TP53_7_12_S	CTGACTGTACCACCATCCACTACAACTACATGTGTACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	345
TP53_7_13_AS	CTTCCAGTGTGATGATGGTGAGGATGGGCCTCCGGACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	346
TP53_8_14_S	ACAGCTTTGAGGTGCGTGTTTGTGCCTGTCCTGGGACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	347
TP53_8_15_AS	GCAGCTCGTGGTGAGGCTCCCCTTTCTTGCGGAGAACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	348
TP53_9_16_S	CACTGCCCAACAACACCAGCTCCTCTCCCCAGCCAACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	349
TP53_9_17_S	CTCCCCAGCCAAAGAAGAAACCACTGGATGGAGAAACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	350
TP53_10_18_S	TGGGCGTGAGCGCTTCGAGATGTTCCGAGAGCTGAACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	351
TP53_10_19_AS	TGAGCCCTGCTCCCCCCTGGCTCCTTCCCAGCCTGACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	352
TP53_11_20_S	CCACCTGAAGTCCAAAAAGGGTCAGTCTACCTCCCACGCGTGGCGG ATGTGGACCCCTTCGAGCAATTA	353
TP53_11_21_AS	AGAAGTGGAGAATGTCAGTCTGAGTCAGGCCCTTCACGCGTGGCG GATGTGGACCCCTTCGAGCAATTA	354

Comparison of Direct versus Indirect Solution-based Capture Methods

Preparation of Oligo pools

A 100 μM pool was created of all the direct capture oligos (50-mers) (SEQ ID NOS:111-231), referred to as "D oligo pool."

IDT Exhibit 1020.2089

A 100 μM pool was created of all the indirect capture chimeric oligos (69-mers), (SEQ ID NOS:234-354) referred to as the "I oligo pool." 1 μM of the biotinylated adaptor capture oligo (SEQ ID NO:232) was added to the I oligo pool, referred to as "I oligo pool + capture adaptor oligo."

TABLE 13: Capture Probes Tested

Sample ID	Capture oligos used	Oligo Concentration
D-10	direct capture biotinylated oligos (SEQ ID NOS:111-231)	10 pmol
D-1	direct capture biotinylated oligos (SEQ ID NOS:111-231)	1.0 pmol
D-0.1	direct capture biotinylated oligos (SEQ ID NOS:111-231)	0.1 pmol
I-10	Indirect oligos (SEQ ID NOS: 234-354) plus biotinylated adaptor oligo (SEQ ID NO:232)	10 pmol of indirect oligos plus 10 pmol of biotinylated adaptor oligo
I-1	Indirect oligos (SEQ ID NOS: 234-354) plus biotinylated adaptor oligo (SEQ ID NO:232)	1.0 pmol of indirect oligos plus 1.0 pmol of biotinylated adaptor oligo
I-0.1	Indirect oligos (SEQ ID NOS: 234-354) plus biotinylated adaptor oligo (SEQ ID NO:232)	0.1 pmol of indirect oligos plus 0.1 pmol of biotinylated adaptor oligo

Capture Mixture

The above dilution series was prepared as follows:

A 1800 μl master mix was prepared by combining 36 μg (545 μl of 66 ng/μl pool) of a gDNA library (non-enriched), prepared with heterologous stem loop adaptors as described in Example 2, 900 μl 2X binding buffer, and 355 μl water. Aliquots were taken from the master mix with two tubes of 300 μl and four tubes of 270 μl. 12.5 μl of either the direct biotinylated oligo pool (1 μM D oligo pool), or the indirect chimeric oligo pool (1 μM I oligo pool + universal adaptor) to the tubes with 300 μl and 30 μl was serially transferred to the remaining tubes to create a dilution series as shown in TABLE 13. 250 μl of each sample was used in the capture method as follows:

The reaction mixtures were annealed as follows:

94°C for 1 minute

90°C for 1 minute

85°C for 1 minute

80°C for 1 minute

75°C for 1 minute

70°C for 1 minute

IDT Exhibit 1020.2090

65°C for 1 minute
60°C for 1 minute
55°C for 1 minute
50°C for 1 minute
45°C for 1 minute
40°C for 1 minute

Capture Reagents

Washed beads were prepared by combining 66 µl beads, 500 µl 2X binding buffer and 440 µl water. The beads were pulled over with a magnet and washed twice with 1 ml 1X binding buffer, and resuspended in 600 µl 1X binding buffer. 100 µl washed beads were transferred to individual tubes and 150 µl 1X binding buffer (10 mM Tris pH 7.6, 0.1 mM EDTA, 1 M NaCl) was added for a total volume of 250 µl.

First Round of Capture

The annealed 250 µl mixture was combined with the 250 µl washed beads. mixed for 15 minutes, the beads were pulled over with a magnet, and the supernatant was decanted. The beads were then washed 4 times with 500 µl TEzero (low salt = 10 mM Tris pH 7.6, 0.1 mM EDTA).

Elution

The DNA bound to the beads was eluted with two aliquots of 50 µl of water by incubation at 94°C for 30 seconds, pulling over the beads and removing the eluate, for a total eluate volume of 100 µl.

Amplification of Eluate (Once-Enriched Library)

PCR Reaction Mixture (5% DMSO)

29 µl H_2O
20 µl 5X buffer (supplied by manufacturer with the EXPAND$^{plus®}$ kit, Roche)
10 µl 25 mM $MgCl_2$
10 µl template (1/10th eluate from once-enriched fragment library)
5 µl dNTPs (10 nM each dNTP)
5 µl DMSO
10 µl 10 µM Forward PCR primer (SEQ ID NO:109)
10 µl 10 µM Reverse PCR primer (SEQ ID NO:110
1 µl Expand$^{PLUS®}$ polymerase (Roche)

IDT Exhibit 1020.2091

100 µl total volume

PCR cycling Conditions:

1 cycle:

94°C for 2 minutes

10 cycles:

94°C for 30 sec

60°C for 30 sec

72°C for 1 minute

10 or 15 cycles:

94°C for 30 sec

60°C for 30 sec

72°C for 1 minute plus 10 sec/cycle

1 cycle:

72°C for 7 minutes

4°C hold

The PCR reaction products were purified over a Qiaquick column and quantified. 1 µl of PCR product was analyzed on a 2% agarose gel.

Second Round Capture

5 µg of first round PCR product was mixed with the capture oligos (D-10:10 pmol D oligo); (I-10:10 pmol I oligo pool + adaptor oligo) in 1X binding buffer (high salt = 10 mM Tris pH 7.6, 0.1 mM EDTA, 1 M NaCl) to a total final volume of 250 µl and annealed under the same temperatures as the first round capture shown above.

The annealed mixture was then mixed with 10 µl of washed beads, as described *supra*. The beads were washed 4 times in TEzero (low salt= 10 mM Tris pH 7.6, 0.1 mM EDTA). The captured DNA was eluted by resuspending the beads in water, as described *supra*, to give a total volume of eluate of 100 µl (twice-enriched). 10 µl of the eluate was amplified in a 100 µl PCR reaction under the same conditions shown above, and purified over a Qiaquick® column.

Third Round of Capture and Enrichment

5 µg of the PCR amplified second round capture material (50 pmol of fragments) was combined with 1 pmol capture oligos in 500 µl 1X binding buffer. The incubation,

IDT Exhibit 1020.2092

wash, and elution steps were carried out as described above for the first and second capture.

TABLE 14: qPCR Analysis of Libraries Enriched by Direct or Indirect Solution-based Capture

Sample	AKT1	KRAS	PIK3CA	PTEN	TP53
No template control	0	0	0	0	0
gDNA	201	323	122	172	895
D10	55,618	99,723	42,283	75,788	331,250
D1	248,767	295,648	103,156	163,626	1,015,336
D0.1	151,637	179,229	47,130	106,709	804,054
I10	112,928	141,126	88,792	143,495	734,659
I1	115,325	127,103	26,589	81,358	592,030
I0.1	30,071	35,305	10,799	30,060	176,812
D10 - 2 rounds	15,539,150	22,955,591	9,332,159	7,222,279	31,904,829
I10 - 2 rounds	29,133,566	27,469,246	11,444,918	13,423,041	83,868,123

TABLE 15: Ratios of Enriched Signal to Starting gDNA Material Showing Fold Enrichment for Each Gene During the Solution-Based Capture Process for Direct Versus Indirect Capture Methodologies

Row	Ratios	AKT1	KRAS	PIK3CA	PTEN	TP53
1	D10/gDNA	277	309	345	440	370
2	D1/gDNA	1240	916	842	951	1134
3	D0.1/gDNA	756	555	385	620	898
4	I10/gINA	563	437	725	834	821
5	I1/glNA	575	394	217	473	661
6	I0.1/gINA	150	109	88	175	197
7	D10 2nd enrich/gDNA	77,450	71,087	76,212	41,977	35,636
8	I10 2nd enrich/gDNA	145,207	85,064	93,466	78,017	93,675
9	D10 2nd enrich/D10	279	230	221	95	96
10	I10 2nd enrich/I10	258	195	129	94	114

IDT Exhibit 1020.2093

As shown above in TABLES 14 and 15, direct capture and indirect capture worked equally well. It is also important to note that, using both methods, the fold enrichment observed for all 5 gene targets are similar, suggesting there was no preferential enrichment of certain sequences.

Sequence Verification of Enriched Library

Up to this point in this example, 5 qPCR assays, each positioned within one of several exons per gene, were used to assess gene enrichment. The collection of capture oligonucleotides was designed to enrich 56 total exon sequences. To establish that enrichment had taken place across all of the exons of targeted genes, the sample that had been twice-enriched with 10 pmol indirect capture oligos was applied to an Illumina sequencing flow cell and 3,272,895 alignable sequencing reads of 36 nucleotides each were obtained. Of these, 35% mapped uniquely to the 5 target gene regions. The majority of these sequencing reads occurred either within the coding regions of coding exons or within nearby flanking intron segments, as expected if sequence based capture was performing to enrich for target segments.

A representative alignment to the PIK3CA gene is shown in FIGURE 7. The upper part of FIGURE 7 is a graph showing number of sequencing reads (y-axis) that map to each base of the PIK3CA gene (displayed along the X-axis). The lower portion of FIGURE 7 shows the exon structure for PIK3CA, with solid boxes representing each coding exon that is spliced into the PIK3CA mRNA. As shown in FIGURE 7, all of the targeted exons in the PIK3CA gene (as well as the other targeted exons in the other 4 genes, not shown) showed a read density of >1000 reads at each targeted exonic base position. These data conclusively demonstrate that targeted capture for gene resequencing using indirect capture strategies is effective.

EXAMPLE 5

This example describes solution-based indirect capture using a population of 3,229 chimeric capture oligos having a first region that is substantially complementary to the sequence of an exon region of one of 77 target genes and a second region for binding to a universal biotinylated oligo, which in turn binds to a capture reagent.

Rationale

This example describes a scale-up from 5 gene targets, 56 exons and 13,267 bp of target sequences that were targeted with 121 oligonucleotides (as described in

IDT Exhibit 1020.2094

Example 4), to 77 genes, 1,221 exons, and 304,161 bp of target sequences targeted with 3,229 capture probes. As further described in this example, during the scale-up of this technique it was discovered that the magnitude of target enrichment was substantially enhanced by more stringent washing of the trimolecular capture complex.

Preparation of Capture Probes

A set of 77 genes was identified that is important in the PI3K kinase pathway, shown below in TABLE 16. All the exons of this set of 77 genes were identified, for a total of 1,221 exons, including alternatively spliced exons, for a total target region of 182,061 bases. An algorithm was then applied for picking alternating sense and antisense strand chimeric oligos with a 5' target-specific region (35 nt) with a sequence that hybridizes to either the sense or antisense strand of each of these exons, and a 3' region (SEQ ID NO:233) that hybridizes to the biotinylated adaptor capture oligo (SEQ ID NO:232), resulting in a total of 3,229 oligos.

These capture oligonucleotides were chosen as follows. For exons less than 69 nucleotides in length, 2 oligonucleotides, both targeting the same strand and oriented in the same direction and not overlapping one another in sequence by more than 10 nucleotides, were chosen. In some cases where exons were very short (i.e., <60 nucleotides), these capture oligonucleotides included flanking exon sequences.

For exons between 70 and 115 nucleotides in length, 2 oligonucleotides targeting opposite Watson and Crick strands and oriented in the opposite orientations were selected. The first oligonucleotide covered exon base positions 1-35 and the second oligonucleotide was positioned from base positions 80-115, which often included flanking intron sequences so that the oligos were each about 35 nt in length and spaced about 45 nt apart.

For exonic sequences greater than 115 nucleotides in length, the first capture oligonucleotide was placed at exon positions 1-35 and successive oligos were placed in alternating orientations with a spacing of 45 nucleotides between oligonucleotides.

Not to be bound by this example, it is envisioned that capture oligonucleotides could be spaced at many different intervals, have many different lengths, and the placement process could take into account genomic features such as genetic variation, G:C content, predicted oligo Tm, and the like.

IDT Exhibit 1020.2095

The oligos designed as described above were synthesized by Operon and provided in a plate at 100 μM and pooled into a single 50 ml sample using a Biomek robot. The pooled 3,229 capture oligos were then diluted to 10 μM and 1 μM.

TABLE 16: Overview of the Set of 3,299 Capture Oligos

Gene	Oligo Numbers in Candidate Set	Total # Oligos/ Per Gene Target	TaqMan Assay
AKT3	1-85	86	Y
AXIN1	86-162	77	
BAD	163-170	8	
BCL2	171-188	18	
BCL2L1	189-203	15	
BMP2	204-219	16	
BRAF	220-261	42	Y
BRCA1	262-397	136	
BRIP1	398-456	59	
CASP9	457-479	23	
CCND1	480-494	15	
CDC42	495-507	13	
CDKN1B	508-516	9	
CDKN2A	517-539	23	
CENTG1	540-585	46	
CHUK	586-630	45	
CTNNB1	631-671	41	Y
EGFR	672-746	75	Y
EP300	747-861	115	
ERBB2	862-930	69	
ESR1	931-957	27	
FBXW7	958-997	40	
FKBP1A	998-1005	8	
FRAP1	1006-1142	137	
GAB1	1143-1177	35	
GSK3B	1178-1203	26	
HGF	1204-1250	47	
HRAS	1251-1262	12	

IDT Exhibit 1020.2096

Gene	Oligo Numbers in Candidate Set	Total # Oligos/ Per Gene Target	TaqMan Assay
IGF1R	1263-1328	66	
IKBKB	1329-1373	45	
IL6	1374-1386	13	
ILK	1387-1412	26	
IRS1	1413-1459	47	
KIAA1303	1460-1534	75	
KRAS	1535-1546	12	Y
LRRFIP1	1547-1585	39	
MAPK1	1586-1603	18	
MAPK3	1604-1625	22	
MAPK8	1626-1654	29	
MET	1655-1721	67	
MYC	1722-1741	20	
NFKB1	1742-1796	55	
NFKBIA	1797-1813	17	
NRAS	1814-1823	10	
PAK1	1824-1855	32	
PALB2	1856-1911	56	
PDPK1	1912-1944	33	
PIK3CA	1945-1999	55	Y
PIK3R1	2000-2041	42	
PLCG1	2042-2119	78	
PPARG	2120-2147	28	
PPP2CA	2148-2164	17	
PRKCA	2165-2202	38	
PRKCZ	2203-2241	39	
PRKDC	2242-2458	217	
PTEN	2459-2481	23	Y
PTK2	2482-2552	71	
RAF1	2553-2588	36	
RB1	2589-2646	58	
RBBP8	2647-2695	49	
RET	2696-2754	59	Y
RICTOR	2755-2851	97	

IDT Exhibit 1020.2097

Gene	Oligo Numbers in Candidate Set	Total # Oligos/ Per Gene Target	TaqMan Assay
RPS6KB1	2852-2884	33	
SMAD3	2885-2905	21	
SRC	2906-2934	29	
STAT1	2935-2986	52	
STAT3	2987-3040	54	
STK11	3041-3062	22	
TGFBR2	3063-3090	28	
TNF	3091-3103	13	
TP53	3104-3126	23	Y
TSC1	3127-3189	63	
TSC2	3190-3288	99	
YWHAH	3289-3299	11	Y

TaqMan assays were developed for the 10 genes (AKT1, BRAF, CTNNB1, EGFR, KRAS, PIK3CA, PTEN, RET, TP53, and YWHAH), as shown in TABLE 16. TaqMan assays were also developed for off-target genes ANKHD and MKRN1 for use as negative controls. These genes were not targeted by capture oligonucleotides and it was shown that their representation diminished during the course of target library enrichment.

Library Generation

Genomic DNA libraries were generated as described above with a 1/100 DNase I treated library (smaller size distribution) and a 1/200 DNase I treated library (larger insert size distribution), forward stem loop adaptors (SEQ ID NO:105) and reverse stem loop adaptors (SEQ ID NO:7) were ligated onto the inserts, followed by PCR amplification for 20 cycles with PCR forward primer (SEQ ID NO:109) and PCR reverse primer (SEQ ID NO:110), then the PCR product was purified over a Qiaquick column.

Solution-Based Capture and Enrichment of Libraries for Target Sequences

In a preliminary experiment, it was determined that although the 5-gene capture worked well under the conditions described above, the 77-gene capture showed a higher level of non-specific binding. While not wishing to be bound by theory, it is likely that the increased oligo diversity going from 121 oligos up to 3,229 oligos creates a more diverse landscape of sequences, which may cause non-specific binding effects.

IDT Exhibit 1020.2098

It was also observed that certain types of plastic microfuge tubes were not optimal for use with the magnetic beads MyOne™ Streptavidin C1 (InVitrogen #650-01). It was determined that microfuge Axygen M-175C tubes work well for magnetic capture using these beads.

As described in this example, it was determined that the addition of 25% formamide in a low salt wash buffer (10 mM Tris pH 7.6, 0.1 mM EDTA) was effective to increase binding specificity in the 77-gene capture environment.

In a related experiment, it was further determined that the addition of 0.1% triton X100 (or tween or NP40) nonionic detergent during the annealing phase enhanced the binding specificity by an order of magnitude (data not shown).

Capture reagents

10 μM of 3,229 capture oligos for the 77 candidate genes described in TABLE 16 were mixed with 10 μM of the biotinylated adaptor oligo (SEQ ID NO:232). The genomic library DNA was prepared as described in Example 2

Capture Mixture

125 μl of 2X binding buffer (2 M NaCl, 20 mM Tris pH 7.6, 0.2 mM EDTA), 60 μl (4.3 μg) of gDNA library, 5 μl capture oligo pool (50 pM of 10 μM of 3229 oligo pool + adaptor oligo), and 60 μl water, for a total volume of 250 μl.

The reaction mixture was annealed as follows:

94°C for 1 minute
90°C for 1 minute
85°C for 1 minute
80°C for 1 minute
75°C for 1 minute
70°C for 1 minute
65°C for 1 minute
60°C for 1 minute
55°C for 1 minute
50°C for 1 minute
45°C for 1 minute
40°C for 1 minute
25°C - hold

IDT Exhibit 1020.2099

Note: It was determined in another experiment that longer annealing times between capture oligos and target DNA (15 minutes per 5°C step) further improved the quality of the capture experiment (data not shown).

Capture Reagents

Washed beads were prepared by combining 6 aliquots of 50 µl beads (in principle, each 50 µl of beads is capable of binding 50 pmol of dsDNA complex), 500 µl 2X binding buffer and 440 µl water. The beads were pulled over with a magnet and washed twice with 1 ml 1X binding buffer.

First Round of Capture/Enrichment

The aliquots of washed oligos were combined with the annealed oligos into a total volume of 1 ml of 1X binding buffer and mixed gently for 15 minutes.

Wash Solutions

A series of wash buffers with increasing formamide were tested, each with 100 mM Tris pH 7.6, 1 mM EDTA, and a range of formamide from 15%, 20%, 25%, 30%, and 50%.

It was previously determined that the presence of 20 mM NaCl in the 10 mM Tris pH 7.6, 1 mM EDTA buffer enhanced non-specific binding (data not shown); therefore, the NaCl was eliminated in the wash buffer in this experiment.

The capture oligos/library/bead complexes were washed 4 times with the above-described wash buffers including formamide, 1 ml each wash for 5 minutes.

Elution

The DNA bound to the beads was eluted with 2 aliquots of 50 µl of water by incubation at 94°C for 1 minute each, pulling over the beads and removing the eluate, for a total eluate volume of 100 µl.

Amplification of Eluate

The eluted material was amplified through 20 cycles of PCR as described in Example 5.

Analysis

qPCR analysis was carried out for the 10 genes shown in TABLE 16. The fold enrichment observed under different wash conditions are shown in TABLE 17.

IDT Exhibit 1020.2100

TABLE 17: Fold Enrichment of Targets in gDNA Library With Wash Buffers Containing Increasing Amounts of Formamide

Wash buffer (core buffer: 100 mM Tris pH 7.6, 1 mM EDTA)	Average fold-enrichment	Stdev	Stdev as % of average
0% formamide	26	7	25
15% formamide	128	36	28
20% formamide	186	51	27
25% formamide	269	77	29
30% formamide	266	107	40
50% formamide	168	278	166

As shown above in TABLE 17, the addition of formamide to the wash buffer had a significantly positive impact on the fold enrichment observed after one round of capture. Because the goal is uniform enrichment of all targets in the library, the 25% formamide washed beads were used in the methods described herein, due to the smaller standard deviation observed.

Second Round of Capture/Enrichment

The eluent obtained from the beads washed in 25% formamide was amplified and 5 µg of purified PCR product was subjected to a second round of oligo capture as follows:

125 µl 2X binding buffer (2 M NaCl, 20 mM Tris pH 7.6, 0.2 mM EDTA), 5 µg PCR product (once-enriched), 5 µl (50 pmol of 50 pM of 10 µM of 3229 chimeric indirect capture oligo pool + adaptor capture oligo), water to a final volume of 250 µl. Annealing was carried out as described for the first capture round. The bound complex was washed 4 times in 1 ml of wash buffer (12.5 mls formamide, 500 µl 1 M Tris pH 7.6, 10 µl of 0.5 M EDTA and 37 mls water for a total of 50 mls wash solution at 25% formamide).

Elution

The DNA bound to the beads was eluted with two aliquots of 50 µl of water by incubation at 94°C for 1 minute each, pulling over the beads and removing the eluate, for a total eluate volume of 100 µl. The eluted twice-enriched material was then PCR amplified through 20 cycles using the PCR conditions described above for the once-enriched eluate.

IDT Exhibit 1020.2101

Third Round of Capture/Enrichment

10 μg of purified twice-enriched PCR amplified product (about 100 nmols of fragments) was annealed to 5 μl of 1 μM (5 pmols) capture oligo pool as follows:

125 μl 2X binding buffer, 10 μg PCR product, 2 μl (2 pmol) of 1 μM indirect candidate oligo pool + adaptor oligo, water was added to a total volume of 250 μl. The captured complexes were annealed to 5 μl of washed beads as described above. The wash in 25% formamide buffer was carried out as described above for the first and second capture. Captured nucleic acids were eluted from the beads with two 25 μl aliquots of water, 1 minute at 94°C each. In the third round of capture/enrichment, the amount of capture oligo was reduced to 5 pmol (instead of 50 pM in the first and second rounds). After 2 rounds of enrichment, the solution of PCR products contains an excess of targeted fragments. When a limiting amount of capture oligos are added to the excess of fragments, the capture oligos will become saturated. Therefore, in the third round of enrichment, capture oligos were added in the same molar abundance so that the composition of the sequencing material would be represented in nearly equal amounts.

The amount of beads was reduced to 5 μl (instead of 50 μl in the first and second round of capture) in order to provide just enough beads to bind all complexes present, in order to minimize any non-specific binding effects that may occur with the use of excess beads in the third round of capture.

Analysis

The starting gDNA library, once-enriched and twice-enriched libraries were analyzed by qPCR and submitted for sequence analysis. The qPCR results, which monitored 1 exon, each within 10 target genes (out of 1,221 total targeted exons: AKT1, BRAF, CTNNB1, EGFR, KRAS, PIK3CA, PRET, PTEN, TP53, and YWHAH) and 1 exon, each within 2 non-targeted genes (ANKHD and MKRN1) are shown in TABLE 18. As described above, shorter insert (100+) and longer insert (200+) libraries of control reference human genomic DNA were captured. TABLE 19 shows the fold enrichment for each individual gene and the averages for all 10 target genes. The shorter insert (100+) library showed 4,650-fold enrichment averaged over all 10 genes and significant depletion of the non-targeted genes. The longer insert library (200+) showed slightly less enrichment, 3,756-fold, as expected for longer targets that are more dispersed across the genome.

IDT Exhibit 1020.2102

TABLE 18: Raw qPCR Values for 10 Targeted and 2 Non-Targeted Genes Across 2 Libraries (100+ and 200+) and 2 Enrichment Steps Per Library

Raw values	100+ gDNA library	100+ 1st enrichment	100+ 2nd enrichment	200+ gDNA library	200+ 1st enrichment	200+ 2nd enrichment
AKT1	113	43,343	1,020,097	367	118,523	2,733,551
ANKHD	28	51	27	125	48	26
BRAF	39	9,954	214,621	161	24,796	337,500
CTNNB1	92	25,257	743,927	361	109,018	2,953,848
EGFR	184	49,839	1,126,117	531	157,430	4,098,106
KRAS	261	42,021	409,344	470	31,328	196,314
MKRN1	60	35	18	216	203	143
PIK3CA	50	4,672	38,174	210	14,221	103,576
PRET	167	33,649	625,767	371	67,372	891,252
PTEN	129	29,636	686,516	228	37,929	593,306
TP53	726	126,317	3,309,405	1,261	305,070	5,131,470
YWHAH	183	28,616	328,326	575	86,582	1,220,647

TABLE 19: Fold-Enrichment Values Over the Starting Raw Genomic DNA Libraries After the 1st and 2nd Rounds of Enrichment for the Shorter-Insert (100+) and Longer Insert (200+) Libraries

Fold-enrichment	100+ library 1st enrichment	100+ library 2nd enrichment	200+ library 1st enrichment	200+ library 2nd enrichment
AKT1	383	9,017	323	7,457
ANKHD	2	1	0	0
BRAF	256	5,520	154	2,098
CTNNB1	275	8,096	302	8,173
EGFR	271	6,122	297	7,724
KRAS	161	1,571	67	418
MKRN1	1	0	1	1
PIK3CA	94	764	68	493
PRET	201	3,741	182	2,405
PTEN	229	5,311	166	2,600
TP53	174	4,556	242	4,068
YWHAH	157	1,798	151	2,123
Average	220	4,650	195	3,756

IDT Exhibit 1020.2103

Sequence Analysis

A sequencing flow cell was created as shown in TABLE 20 in order to determine the specific coverage of target genes as a function of library enrichment and normalization.

TABLE 20: Sequencing Analysis by Flow Cell

Sample ID	Library	Amount Added	# of Clusters Per Sequencing Tile
1	100+ insert/gDNA library (non-enriched)	4μl of 1 ng/μl to 2 μl	34,642
2	100+ insert/once-enriched	4μl of 1 ng/μl to 2 μl	33,088
3	100+insert/twice-enriched	4μl of 1 ng/μl to 2 μl	27,829
4	100+insert/twice-enriched and normalized (third round of enrichment)	20 μl to 4 μl	30,047
5	200+insert/gDNA library (non-enriched)	4 μl of 1 ng/μl to 2 μl	24,813
6	200+insert/once-enriched	4 μl of 1 ng/μl to 2 μl	23,472
7	200+insert/twice-enriched	4 μl of 1 ng/μl to 2 μl	20,455
8	200+insert/twice-enriched and normalized (third round of enrichment)	20 μl to 4 μl	22,375

TABLE 21: Alignment of Sequencing Reads to Target Region

Sample ID	Library	total number of 27nt sequencing reads	# of reads that align to the gene loci	% reads that align to the gene loci	# of reads that align to coding exons	% reads that align to coding exons	fold enrichment of reads aligning to coding exons
1	100+ insert/gDNA library (non-enriched)	3,798,404	135,139	3.55%	262	.0069%	1
2	100+ insert/once-enriched	4,220,177	276,322	6.54%	31,865	0.755%	122
3	100+insert/twice-enriched	4,274,896	1,099,077	25.71%	653,057	15.27%	2493
4	100+insert/twice-enriched and normalized (third round of enrichment)	4,431,923	2,144,103	48.38%	1,428,594	32.23%	5453

IDT Exhibit 1020.2104

Sample ID	Library	total number of 27nt sequencing reads	# of reads that align to the gene loci	% reads that align to the gene loci	# of reads that align to coding exons	% reads that align to coding exons	fold enrichment of reads aligning to coding exons
5	200+insert/gDNA library (non-enriched)	4,335,770	156,622	3.61%	324	.0074%	1
6	200+insert/once-enriched	4,243,550	260,528	6.13%	29,577	0.696%	91
7	200+insert/twice-enriched	3,855,917	970,067	25.15%	464,853	12.05%	1435
8	200+insert/twice-enriched and normalized (third round of enrichment)	3,913,182	1,697,263	43.37%	933,503	23.85%	2881

To assess the overall characteristics of the 77-gene targeted libraries, the 100+ twice-enriched processed library was applied to one lane of an Illumina sequencing flow cell. As shown above in TABLE 21, for the 100+ short insert library, approximately 50% of the overall sequencing reads in the twice-enriched and normalized (third round of enrichment) sample (#4) aligned to the gene target region (all the sequence space around a particular gene), while 32% of these reads aligned specifically to the target exonic regions. This demonstrates that 67% (1,428,594/2,144,103 x 100%) of the sequencing reads that map to the gene region are within the target exons. Further analysis indicated that the remaining 33% of the gene region mapped sequencing reads aligned to the intronic regions immediately adjacent to the exons (data not shown).

For the 200+ longer insert library, approximately 43% of the overall sequencing reads in the twice-enriched and normalized sample (third round of enrichment) (#8) aligned to the gene target region, while 24% of these reads aligned specifically to the target exonic regions. This demonstrates that 55% (933,503/1,697,263 x 100%) of the sequencing reads that map to the gene region are within the target exons and 45% of the reads mapped to adjacent introns, as would be expected for longer inserts that extend into intronic regions.

With regard to the sequence coverage at the level of an individual gene, as shown in FIGURE 8, a base-by-base read depth from the 100+ short insert twice-enriched and normalized (third round of enrichment) library was determined for the exemplary gene AKT1. FIGURE 8 shows the exon structure for AKT1, with solid boxes representing

IDT Exhibit 1020.2105

exons and dotted lines representing intron regions. The base-by-base sequencing read depth is plotted on a scale from 0 to 20 reads. As shown in FIGURE 8, each exon region was covered by a sequencing read depth of at least 20 reads, while the intronic regions that were sequenced all clustered around the exonic targets of interest.

To address the issue of uniformity of sequencing coverage of a target region, the performance of the 100+ short insert twice-enriched library (sample #4 in TABLE 21) was analyzed in more detail. FIGURE 9 shows the overall characteristics of this data, with the X-axis showing the sequencing coverage depth, defined as the number of times each individual base was found in an aligned sequence. The y-axis shows the percentage of bases, defined as the percentage of bases that have $\geq$ the coverage depth shown on the x-axis. The percent of target bases was plotted as a function of sequence coverage depth (i.e., number of sequencing reads). The line plotted in FIGURE 9 shows that 99% of the target bases were covered by at least one sequencing read and the arrow shows that 90% of the target bases were covered by 16 or more sequencing reads. This result is important because sequencing read depths $\geq$16 are necessary to reveal single nucleotide polymorphisms (SNPs) with confidence. Therefore, this overall coverage analysis indicated that the data obtained from one flow cell lane on a given sample (~4,000,000 reads), there would be adequate sequence coverage depth to determine the presence of a small nucleotide polymorphism (SNP) with confidence across >90% of the target capture region.

In view of these results, an additional criterion for the selection of capture probe sequences will be to scan the candidate capture probe sequences for the presence of any known duplicated regions and eliminate these from use. Another approach will be to design the capture probes to selectively align to a particular genomic region of interest, such as a region less than one megabase of the human genome.

In summary, it has been demonstrated that the methods of generating an enriched library for a target region of interest are very useful for high throughput resequencing. In particular, it has been demonstrated that a genomic library that has been twice-enriched and normalized (third round of enrichment) library using the target capture methods described herein provides a highly enriched fraction of sequencing templates containing the target regions we endeavor to sequence.

IDT Exhibit 1020.2106

EXAMPLE 6

This example describes the use of the solution-based capture method for sequence analysis of copy number variation from a sequencing-ready library prepared from total genomic library.

Rationale

In contrast to the methods described in Examples 3-5, in which it was advantageous to enrich the library for target sequences by capturing the target sequences in several rounds of solution-based capture, in this example, the concept was to generate low coverage shotgun sequencing of total genomic libraries that contained the number of target regions that were representative of the starting sample and sequencing the library. Read density maps were then generated by mapping the sequencing reads back to large, 500 Kb intervals of a reference genome corresponding to the type of sample. This example describes the application of this method to chromosome 14 of a human subject.

Methods

A sequencing-ready library of total genomic DNA inserts was generated as described in Example 2, starting with genomic DNA isolated from a healthy human subject, DNase I treating, blunt end polishing, and ligating on stem-loop linkers (SEQ ID NO:105 and SEQ ID NO:107), followed by 20 cycles of PCR and purification over a Qiaquick® column.

Analysis

Read density maps were generated by mapping the sequencing reads of the once-enriched library back to large, 500 Kb intervals of the sequenced 87.3 Mb portion of human chromosome 14.

FIGURE 10A illustrates the measurement of copy number variation using low-coverage genomic sequencing and molecular karyotyping, with the density of aligning sequencing reads per 100 Kb plotted along the x-axis as the apparent copy number. FIGURE 10A, shows a sample containing a normal diploid chromosomal region (shown on the left), which exhibits a uniform 2 n density of sequencing reads across the entire region. In contrast, as further shown in FIGURE 10A, a sample containing 1 normal chromosome and 1 chromosome with a deletion and a tandem duplication (shown on the right) will generate an abnormally low read density "dip" to 1 n across the

IDT Exhibit 1020.2107

deleted region and an abnormally high read density "spike" to 3 n across the duplicated region.

FIGURE 10B shows the actual molecular karyotype across the 87.3 Mb sequenced portion of chromosome 14 showing uniform 2 n coverage from the normal human subject using the methods described in this Example. The density of aligning reads per 100 Kb region is plotted on the line shown.

The results in this example demonstrate the use of the described methods to analyze copy number variation of a target region of genomic DNA of interest from a subject.

EXAMPLE 7

This example describes a combination of whole transcriptome amplification, sequencing-ready library generation of the amplified whole transcriptome, enrichment of the library for target sequences of interest, and targeted resequencing of the library.

Rationale

Recent genome-wide association studies, which attempt to correlate a plurality of subjects sharing a particular phenotype with a particular genetic variation, have generated a puzzling result; more than half of the statistically defensible associations reported to date (>170 reports as of mid 2007) map to chromosomal regions devoid of any known genes.

This example describes the generation of a sequencing-ready library generated from whole transcriptome amplified nucleic acids that is enriched for a target locus of interest in order to provide sufficient sequencing coverage of a specific chromosomal location in a cost-effective manner.

In order to demonstrate proof of concept for the combined approach, we focused on the cardiovascular disease risk locus on chromosome 9p21 (Helgadottir et al., *Science 316*:1491 (2007), shown in FIGURE 11A which contains two SNPs (shown as SNPA and SNPB) associated with cardiovascular risk.

Methods

cDNA Synthesis for Whole-Transcriptome Library

A whole-transcriptome library was first generated as described in U.S. Patent Application Publication No. 2008/0187969, herein incorporated by reference. Briefly summarized, the method involves using a population of oligonucleotides to prime the

IDT Exhibit 1020.2108

amplification of a target population of nucleic acid molecules within a larger population of nucleic acid molecules, wherein each oligonucleotide comprises a hybridizing portion, wherein the hybridizing portion consists of one of 6, 7, or 8 nucleotides; and the population of oligonucleotides is selected to hybridize under defined conditions to a first subpopulation of the target nucleic acid population (i.e., mRNA molecules obtained from a human subject), but not hybridize under the defined conditions to a second subpopulation of the target nucleic acid population (i.e., ribosomal RNA). A population of not-so-random oligonucleotides that is capable of amplifying all transcripts except 18S and 28S transcripts was used to prime the amplification of mRNA and was generated as described in U.S. Patent Application Publication No. 2008/0187969.

Total RNA was extracted from a human subject and reverse transcriptase was used for first strand cDNA synthesis from the template RNA with the set of non-so-random primers. Second strand cDNA synthesis was then carried out and the double-stranded cDNA was used as the starting material for preparation of a sequencing-ready library, as described in Example 2.

Capture Oligonucleotides

A universal 5' biotinylated oligo was used:

5' [BioTEG] TAATTGCTCGAAGGGGTCCACATCCGCCACGCGT 3' (SEQ ID NO:232)

A series of closely-spaced chimeric capture oligos were generated that span a 200 Kb region across a segment of chromosome 9p21 encompassing SNPA and SNPB, as shown in FIGURE 11B. The chimeric oligos were not biotinylated and each has a first 5' region that hybridizes to the target region of chromosome 9p21, and a second 3' region consisting of the following additional sequence that hybridizes to universal oligo:

5' ACGCGTGGCGGATGTGGACCCCTTCGAGCAATTA 3' (SEQ ID NO:233)

Solution-based Capture

Three rounds of solution-based capture were carried out as described in Example 3. Each fraction of library (raw library, first round enriched material, second round enriched material, and third round enriched material) was then sequenced and analyzed as described below.

IDT Exhibit 1020.2109

Analysis

While very few ESTs map to the cardiovascular disease risk locus on chromosome 9p21, sequencing data obtained from sequencing-ready libraries generated from whole transcriptome amplification using the not-so-random primer amplification method, as described in this example, provided evidence that a large segment surrounding the disease-associated SNPA and SNPB is actively transcribed. As shown in FIGURE 11A, an expanded view of this region of interest shows transcriptional activity covering >800 Kb that can be transcriptionally assigned to one or two transcriptional units.

As shown in FIGURE 11B, by increasing the intensity of transcript-derived sequences from this region, the transcript structure can be mapped with confidence. Based on this information, assays can be developed that interrogate expression patterns of this region in tissues (e.g., a body atlas) to determine if expression of such loci correlates with patient phenotypes.

EXAMPLE 8

This example describes the use of the solution-based capture method for sequence analysis of genomic DNA isolated from clinical patient samples in order to identify genetic markers prognostic for treatment outcomes.

Rationale

While the application of genome-wide association studies to samples obtained from patients during clinical trials has the potential to point to genetic markers that may be prognostic for treatment outcomes, such associations are often weak. One explanation for this weak association is that while a genotyped SNP may be linked to an important genetic variation, it may not itself be causal for the observed phenotype. Moreover, although the genotyped SNP may be a common variant, it may also be linked to a rare, and as yet undiscovered, variation that is much more strongly associated with the phenotypic trait. Therefore, targeted resequencing can be used to uncover rare genetic variants, for example, in the genomic region surrounding a previously identified SNP.

Methods

Nucleic acids are isolated (DNA or RNA) from clinical samples obtained from subjects undergoing a particular treatment, or from a group of subjects exhibiting a particular phenotype of interest. Sequencing-ready libraries are made from the isolated

IDT Exhibit 1020.2110

nucleic acids, and the libraries are then enriched for a particular target region of interest. For example, a target region of interest may encompass the region surrounding a known SNP, such as common SNP "A" that is weakly associated with a rare and unfavorable adverse event. Targeted resequencing of a ~40 Kb region surrounding this SNP uncovers a rare C →T SNP that is more strongly associated with the adverse event. Genotyping for the rare T variant in treatment populations would enable clinicians to eliminate subjects vulnerable to unfavorable outcomes.

The methods described in this example may be carried out on a plurality of nucleic acid-containing samples obtained over a period of time from the human subject in order to monitor the subject for genetic mutations in a target region of interest or to monitor the effect of a particular treatment regimen on a subject.

While the preferred embodiment of the invention has been illustrated and described, it will be appreciated that various changes can be made therein without departing from the spirit and scope of the invention.

IDT Exhibit 1020.2111

CLAIMS

The embodiments of the invention in which an exclusive property or privilege is claimed are defined as follows:

1. A method of generating a population of DNA molecules, each DNA molecule comprising a nucleic acid insert region flanked by a first primer binding region and a second primer binding region, the method comprising:

(a) fragmenting a starting population of DNA molecules into a population of fragmented insert DNA molecules;

(b) combining in a ligation reaction, the population of fragmented insert DNA molecules of step (a) with:

(i) a plurality of first stem-loop linker oligonucleotides comprising a sequence that is complementary to a first primer binding region and

(ii) a plurality of second stem-loop linker oligonucleotides comprising a sequence that is complementary to a second primer binding region;

(c) contacting the ligation reaction of step (b) with a polymerase under conditions suitable to synthesize the complementary strands corresponding to the first and second stem-loop linkers, thereby generating a plurality of double-stranded DNA molecules, each DNA molecule comprising an insert region flanked by a first primer binding region and a second primer binding region; and

(d) performing a polymerase chain reaction on the double-stranded molecules of step (c) with a plurality of first PCR primers that bind to the first primer binding region and a plurality of second PCR primers that bind to the second primer binding region to selectively amplify the population of DNA molecules comprising an insert fragment flanked by a first stem-loop linker oligonucleotide and a second stem-loop linker oligonucleotide.

2. The method of Claim 1, wherein the starting population of DNA molecules comprises PCR products amplified from a plurality of target-specific amplicons from a nucleic acid containing sample.

3. The method of Claim 2, wherein the PCR products comprise a plurality of exon regions of at least one target gene of interest.

IDT Exhibit 1020.2112

4. The method of Claim 1, wherein the starting population of DNA molecules comprises total genomic DNA.

5. The method of Claim 1, wherein the starting population of DNA molecules comprises cDNA.

6. The method of Claim 5, wherein the cDNA is generated using a population of oligonucleotides selected to hybridize under defined conditions to a population of mRNA molecules but not hybridize under the defined conditions to ribosomal RNA.

7. The method of Claim 1, wherein the population of fragmented insert DNA molecules has an average length of from 50 nucleotides to 10,000 nucleotides.

8. The method of Claim 1, wherein the population of fragmented insert DNA molecules has an average length of from 50 nucleotides to 500 nucleotides.

9. The method of Claim 1, wherein the DNA molecules are fragmented with DNase I in the presence of manganese under conditions suitable to generate random double-stranded nucleic acid breaks.

10. The method of Claim 1, wherein the first or second stem-loop linker oligonucleotides further comprise a sequence for binding to a high-throughput sequencing platform.

11. The method of Claim 1, wherein the first or second stem-loop linker oligonucleotides further comprise a molecular bar code.

12. The method of Claim 1, wherein the first and second stem-loop linker oligonucleotides have a length of at least 40 nucleotides.

13. The method of Claim 4 or Claim 5, wherein the method further comprises

(e) annealing the nucleic acid molecules amplified according to step (d) with a plurality of target capture probes, each target capture probe comprising:

IDT Exhibit 1020.2113

(i) a target-specific binding region consisting of a nucleic acid sequence that is at least 95% identical to at least a portion of the sense or antisense stand of a target insert of interest; and

(ii) a region for binding to a capture reagent,

(f) adding a capture reagent that binds to the target capture probes to generate a plurality of nucleic acid molecule/target capture probe/capture reagent complexes,

(g) removing the complexes from the mixture with a sorting device that binds to the capture reagent; and

(h) eluting the nucleic acid sequences from the captured complexes, wherein the plurality of eluted nucleic acid sequences are enriched for the target regions of interest.

14. The method of Claim 13, wherein the method further comprises sequencing at least a portion of the population of DNA molecules from step (h).

15. A population of double-stranded DNA molecules comprising a nucleic acid insert region flanked by a first primer binding region and a second primer binding region generated according to the method of Claim 1.

16. A method of enriching a library for target nucleic acid regions of interest, the method comprising:

(a) contacting a library of DNA molecules comprising a subpopulation of nucleic acid target insert sequences of interest flanked by a first primer binding region and a second primer binding region within a larger population of nucleic acid insert sequences flanked by the first primer binding region and the second primer binding region with a set of capture probes, the set of capture probes comprising a plurality of capture oligonucleotides, each comprising a first target sequence-specific binding region and a second capture reagent binding region, under conditions that allow binding between the capture oligonucleotides and the nucleic acid target regions of interest, to form a mixture comprising a plurality of complexes between target regions of interest and capture probes;

(b) contacting the mixture of step (a) with a capture reagent and separating the capture reagent bound complex from the mixture; and

IDT Exhibit 1020.2114

(c) eluting the target regions of interest flanked by the first primer binding region and the second primer binding region from the capture reagent bound complex.

17. The method of Claim 16, wherein the second capture reagent binding region directly binds to the capture reagent.

18. The method of Claim 16, wherein the second capture reagent binding region binds to an adaptor capture oligonucleotide comprising a region that binds to the capture reagent; wherein the method further comprises contacting the mixture of step (a) with a plurality of adaptor capture oligonucleotides.

19. The method of Claim 16, wherein step (a) is carried out in a solution comprising from 100 mM to 2 M NaCl.

20. The method of Claim 16, further comprising washing the separated capture reagent bound complex with a wash solution comprising less than 10 mM NaCl prior to step (c).

21. The method of Claim 17, wherein the wash solution further comprises from 15% to 30% formamide.

22. The method of Claim 13, wherein the set of capture probes comprises a plurality of capture oligonucleotides, each capture probe comprising a first target-specific binding region that is at least 95% identical to at least a portion of the sense or antisense strand of the exons in at least 5 different genes.

23. The method of Claim 16, wherein the set of capture probes comprises a plurality of capture oligonucleotides, each comprising a first target-specific binding region that is at least 95% identical to at least a portion of the sense or antisense strand of the exons in at least 70 different genes.

24. The method of Claim 16, further comprising amplifying the eluted target regions of interest flanked by the first primer binding region and the second primer binding region with a forward PCR primer and a reverse PCR primer that bind to the first

IDT Exhibit 1020.2115

and second primer binding regions to generate a library that is once-enriched for target regions of interest.

25. The method of Claim 24, further comprising:

(d) contacting the library that is once-enriched for target regions of interest with the capture probe pool under conditions that allow binding between the capture oligonucleotides and the nucleic acid target regions of interest, to form a plurality of complexes between target regions of interest and capture probes;

(e) contacting the mixture of step (d) with a capture reagent and separating the capture reagent bound complex from the mixture; and

(f) eluting the target regions of interest flanked by the first primer binding region and the second primer binding region from the capture reagent bound complex.

26. The method of Claim 25 further comprising amplifying the eluted target regions of interest flanked by the first primer binding region and the second primer binding region with a forward PCR primer and a reverse PCR primer that bind to the first and second primer binding regions to generate a library that is twice-enriched for target regions of interest.

27. A method of generating a target enriched, sequencing ready library for resequencing at least one target region of interest from a nucleic acid containing sample, the method comprising:

(a) providing a library comprising fragmented nucleic acid molecules flanked by a first primer binding region and a second primer binding region; and

(b) enriching the library for target sequences with a set of capture probes comprising a plurality of capture oligonucleotides, each comprising a first target-specific binding region and a second capture reagent binding region, thereby generating an enriched sequencing ready library for resequencing at least one target region of interest.

28. The method of Claim 27, further comprising PCR amplifying the enriched library with PCR primers that bind to the first primer binding region and the second primer binding region to generate an amplified product.

IDT Exhibit 1020.2116

29. The method of Claim 27, wherein at least one of the first stem-loop linker or the second stem-loop linker oligonucleotides comprises a molecular bar code.

30. The method of Claim 27 or Claim 28, further comprising sequencing at least a portion of the enriched library to determine the sequence of the regions of interest.

31. The method of Claim 27, wherein the library is generated from nucleic acids obtained from a human subject.

IDT Exhibit 1020.2117

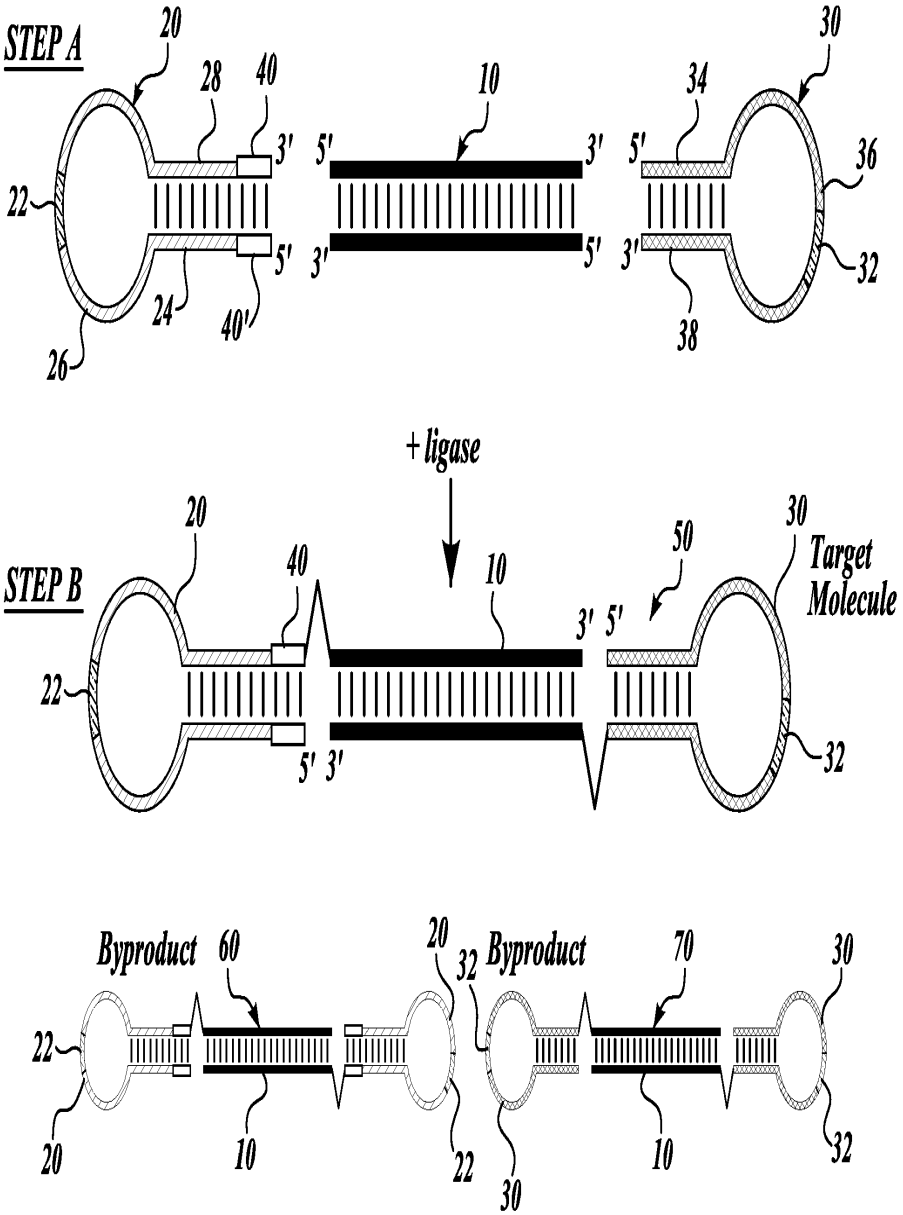

STEP A
20
28
40
10
34
30
3'
5'
3'
5'
36
22
5'
3'
5'
3'
32
24
40'
38
26
+ ligase
STEP B
20
40
10
50
30
Target
Molecule
3'
5'
22
32
5'
3'
Byproduct
60
20
32
Byproduct
70
30
22
10
10
32
20
22
30

Fig.1.

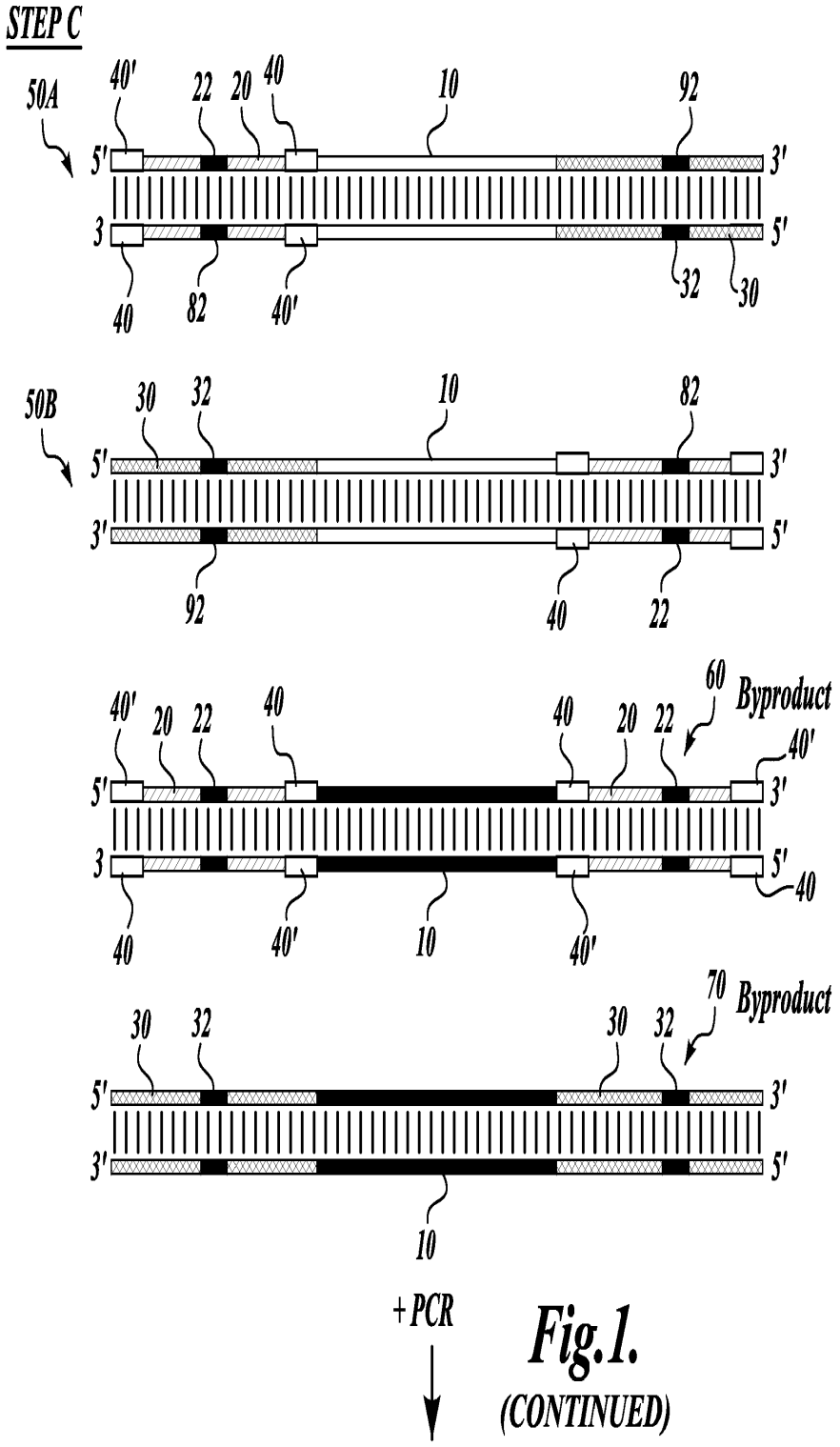


Fig.1.
(CONTINUED)

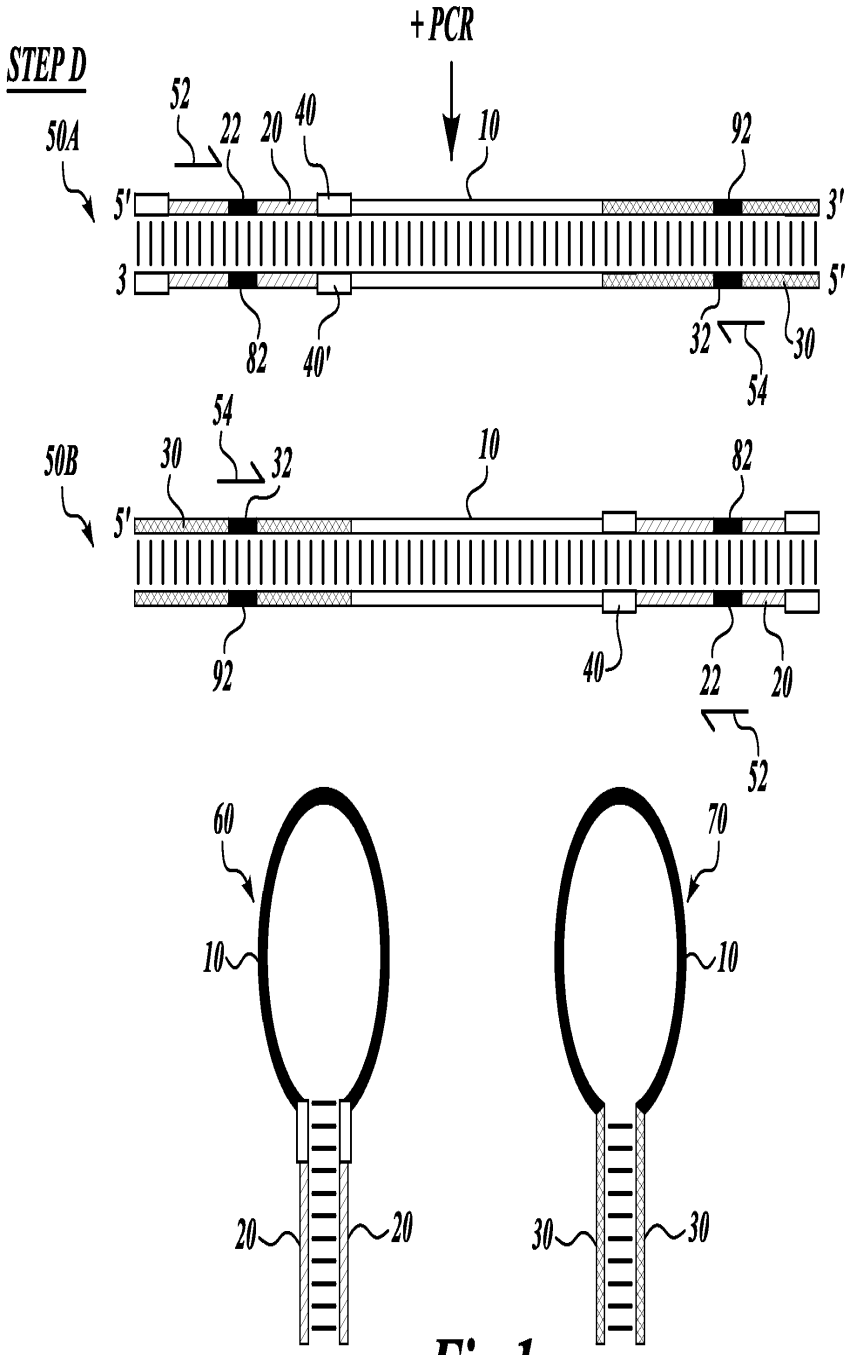


Fig.1.
(CONTINUED)

PCR PRODUCTS:

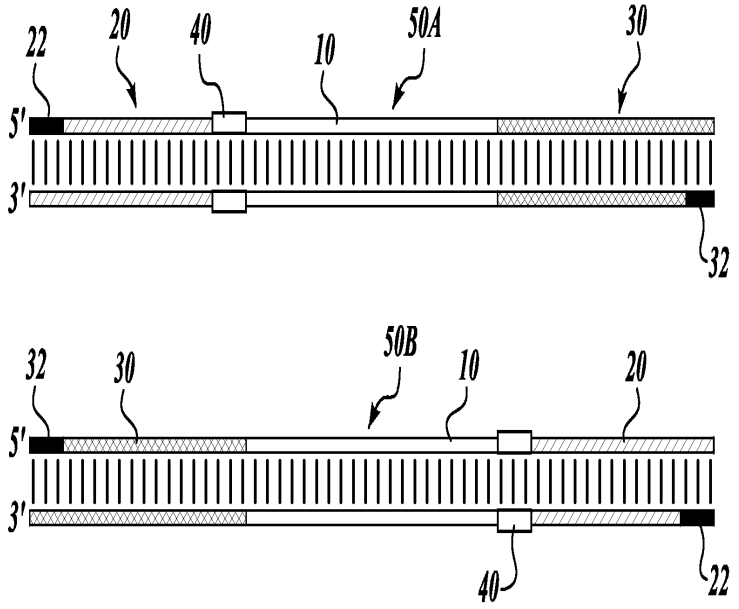


Fig.1.
(CONTINUED)

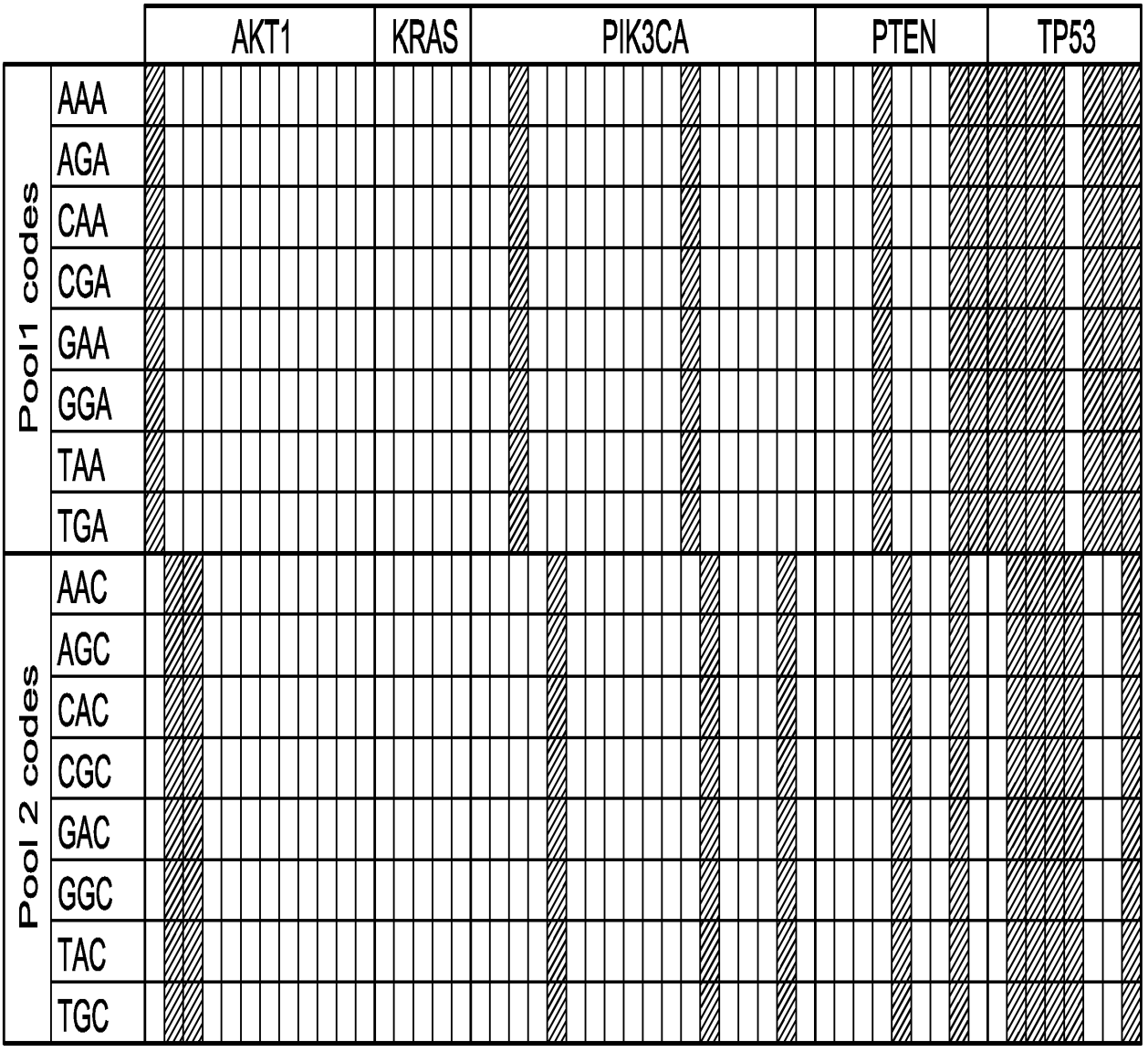

AKT1
KRAS
PIK3CA
PTEN
TP53
Pool1 codes
AAA
AGA
CAA
CGA
GAA
GGA
TAA
TGA
Pool 2 codes
AAC
AGC
CAC
CGC
GAC
GGC
TAC
TGC

Fig.2A.

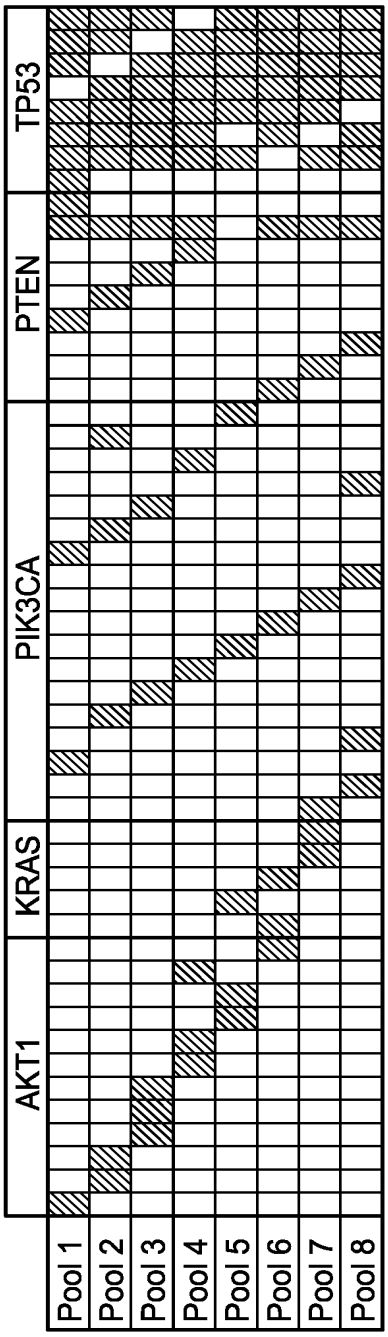

AKT1
KRAS
PIK3CA
PTEN
TP53
Pool 1
Pool 2
Pool 3
Pool 4
Pool 5
Pool 6
Pool 7
Pool 8

Fig.2B.

METHOD OF GENERATING A SEQUENCE-READY LIBRARY

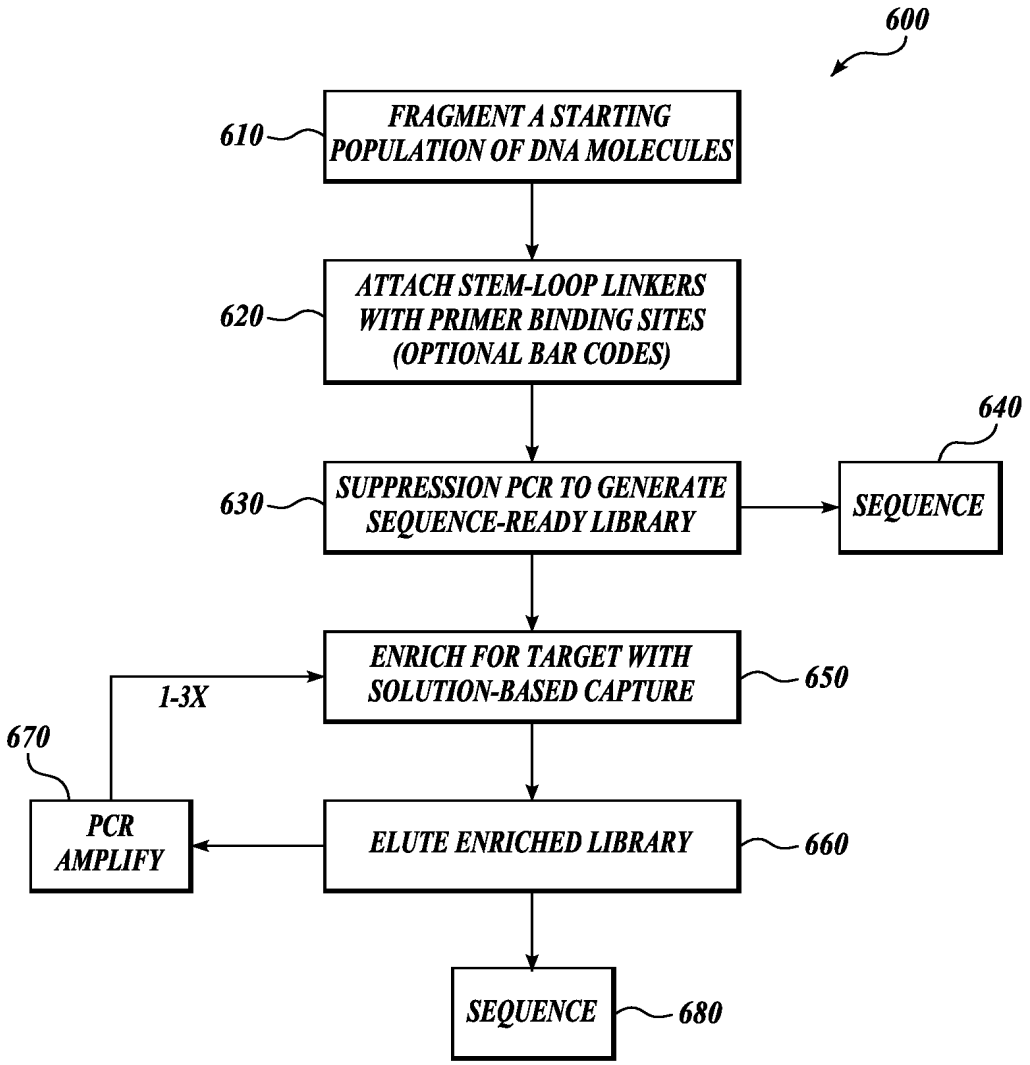


Fig.3.

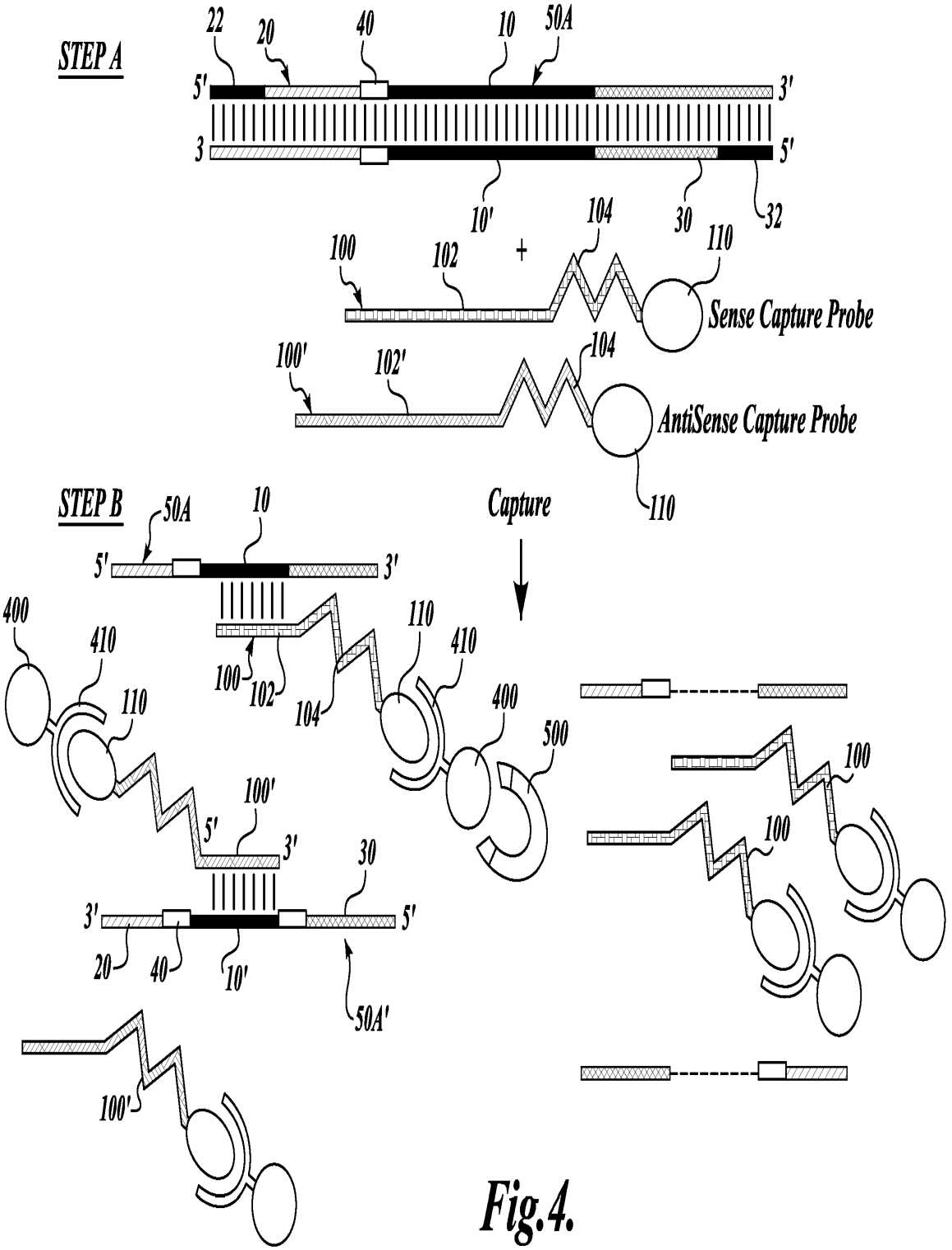

STEP A
22
20
40
10
50A
5'
3'
3
5'
10'
30
32
104
100
102
+
110
Sense Capture Probe
100'
102'
104
AntiSense Capture Probe
110
STEP B
50A
10
Capture
5'
3'
400
410
110
100
102
104
110
410
400
500
100'
5'
3'
30
3'
5'
20
40
10'
50A'
100'
100
100

Fig.4.

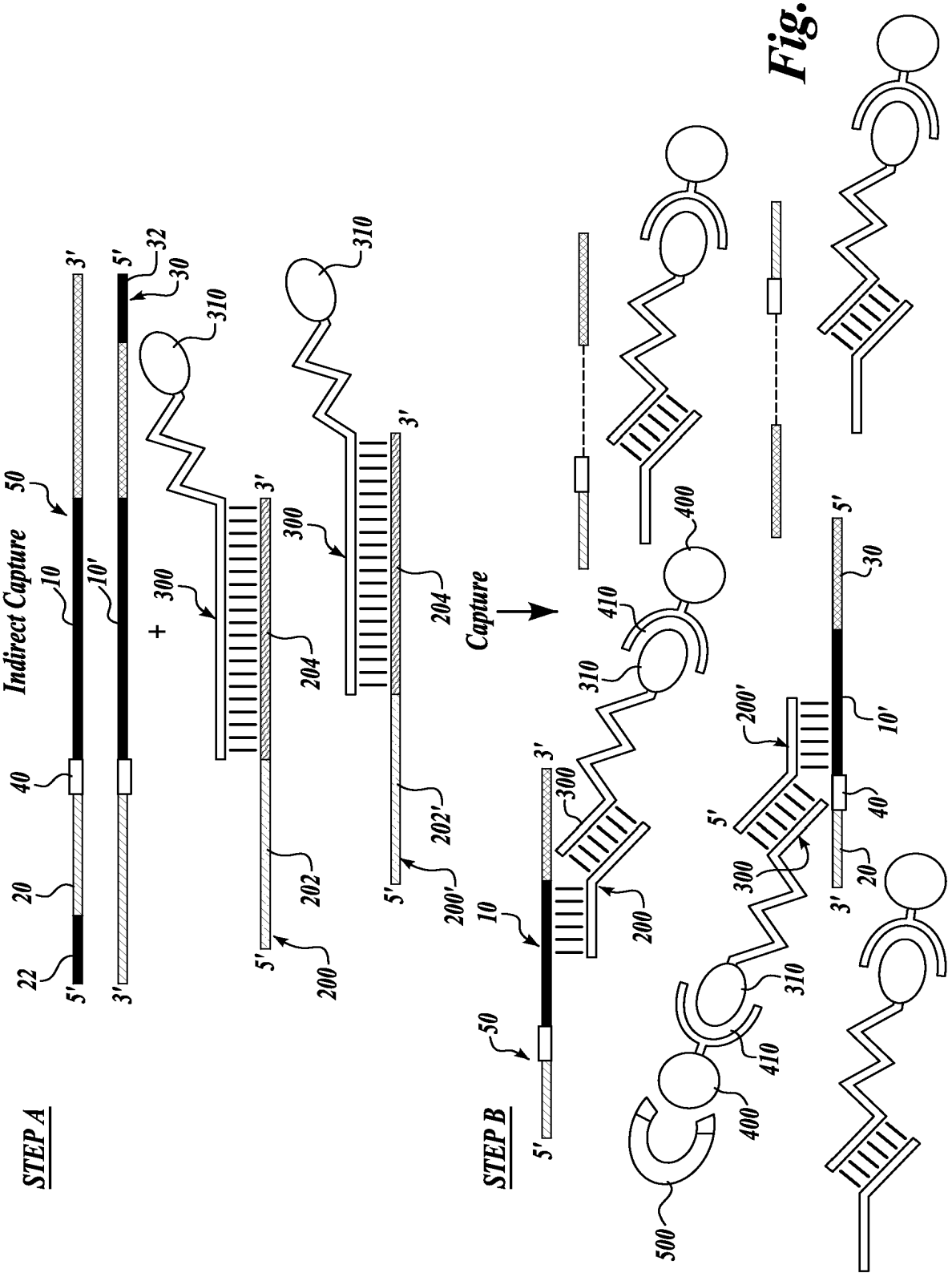

STEP A
Indirect Capture
Capture
STEP B
5'
3'
10
10'
20
22
30
32
40
50
200
200'
202
202'
204
300
310
400
410
500

Fig.5.

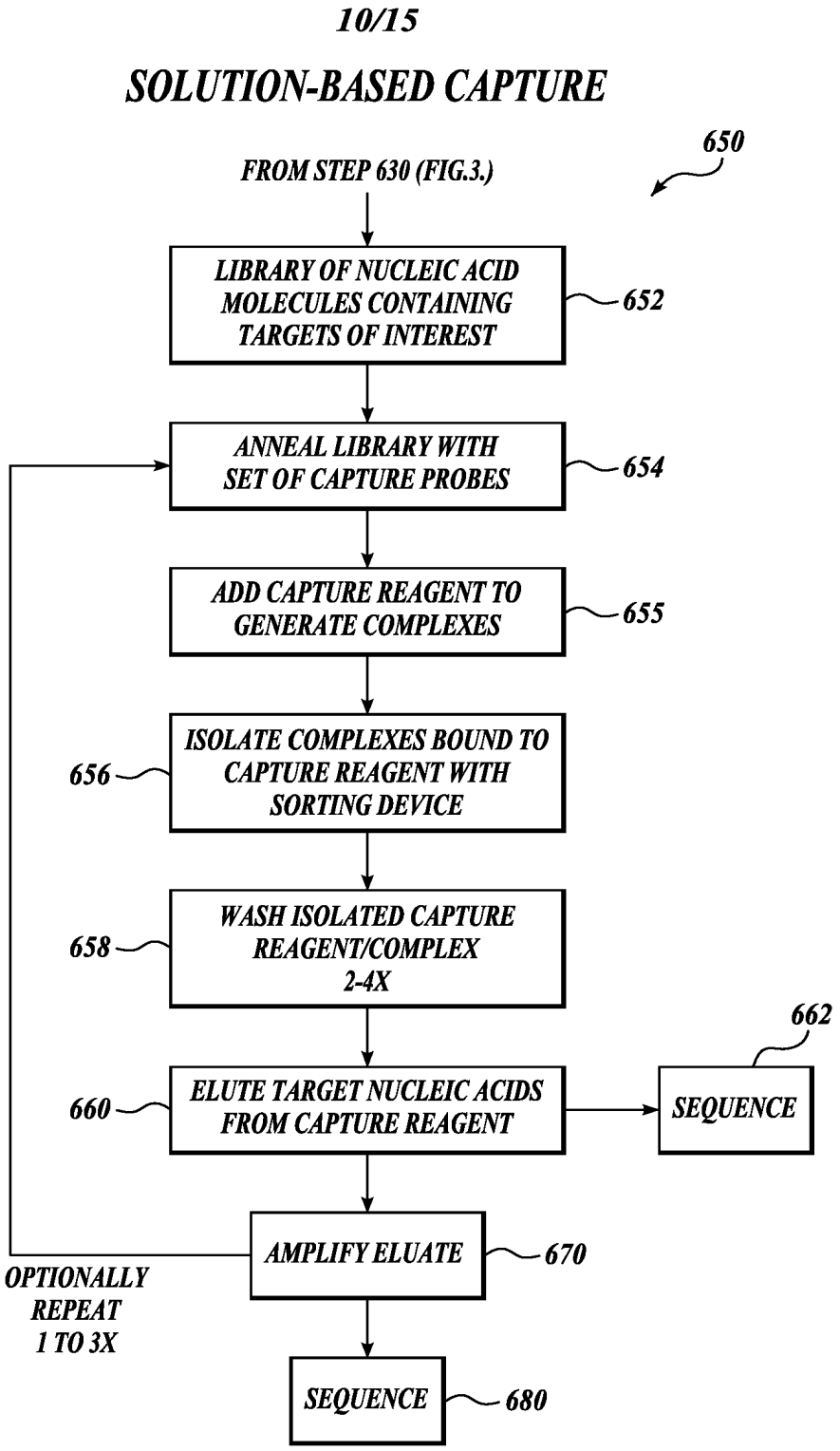

SOLUTION-BASED CAPTURE
650
FROM STEP 630 (FIG.3.)
LIBRARY OF NUCLEIC ACID MOLECULES CONTAINING TARGETS OF INTEREST
652
ANNEAL LIBRARY WITH SET OF CAPTURE PROBES
654
ADD CAPTURE REAGENT TO GENERATE COMPLEXES
655
ISOLATE COMPLEXES BOUND TO CAPTURE REAGENT WITH SORTING DEVICE
656
WASH ISOLATED CAPTURE REAGENT/COMPLEX 2-4X
658
662
ELUTE TARGET NUCLEIC ACIDS FROM CAPTURE REAGENT
660
SEQUENCE
AMPLIFY ELUATE
670
OPTIONALLY REPEAT 1 TO 3X
SEQUENCE
680

Fig. 6.

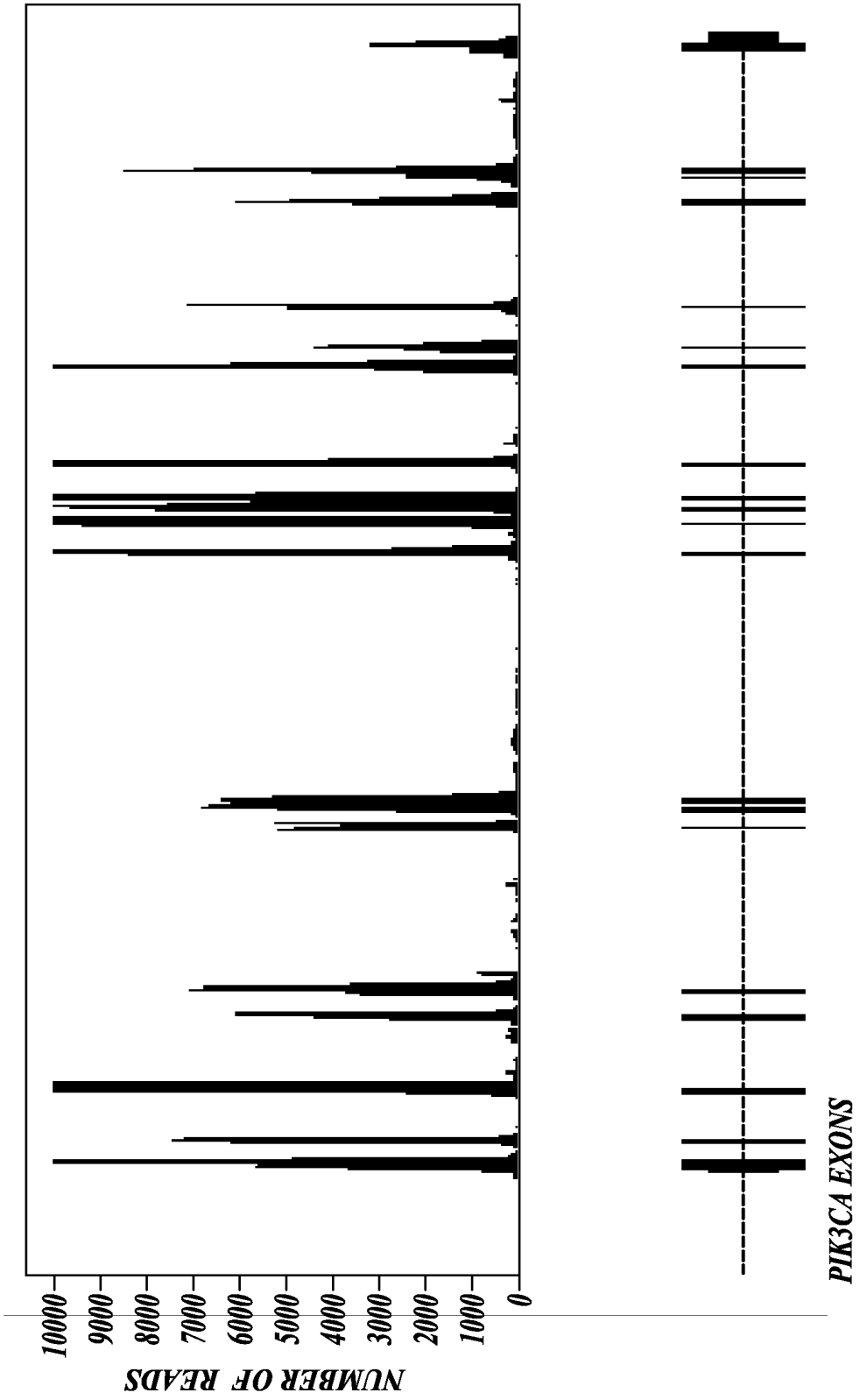

NUMBER OF READS
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
PIK3CA EXONS

Fig. 7.

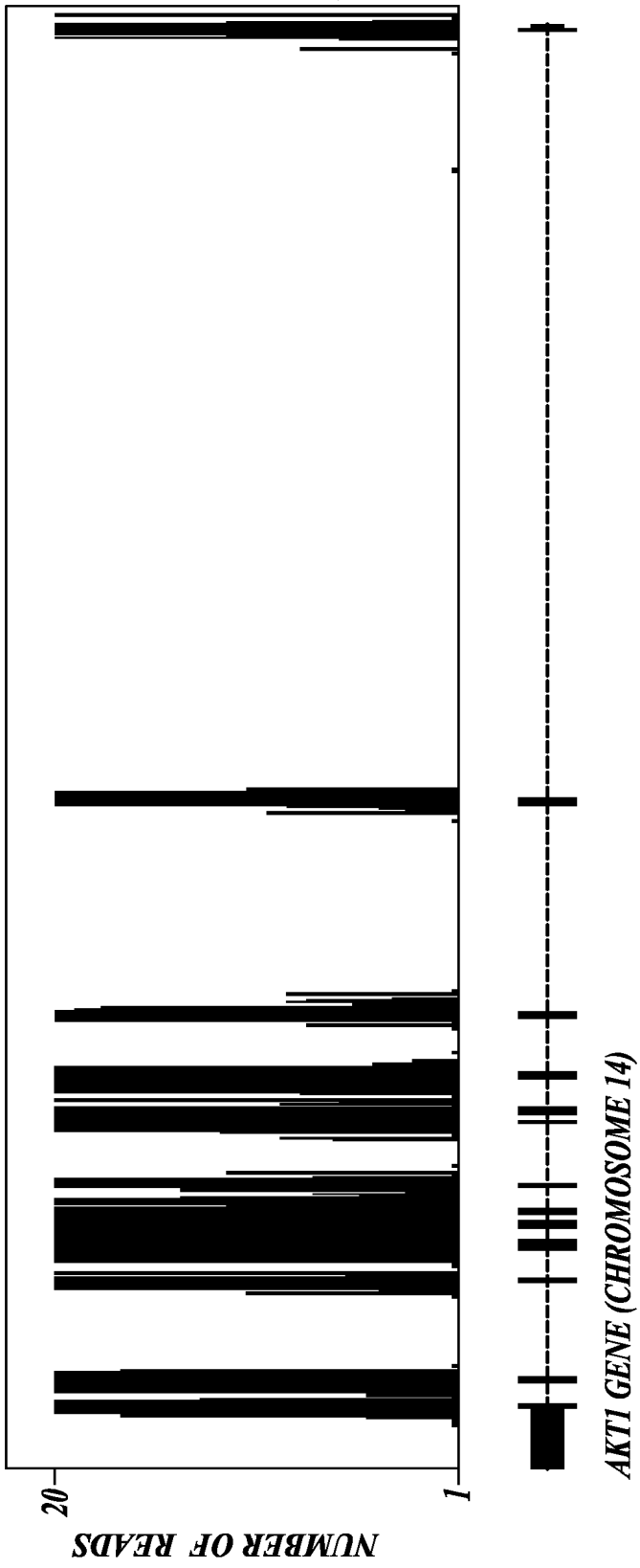


Fig.8.

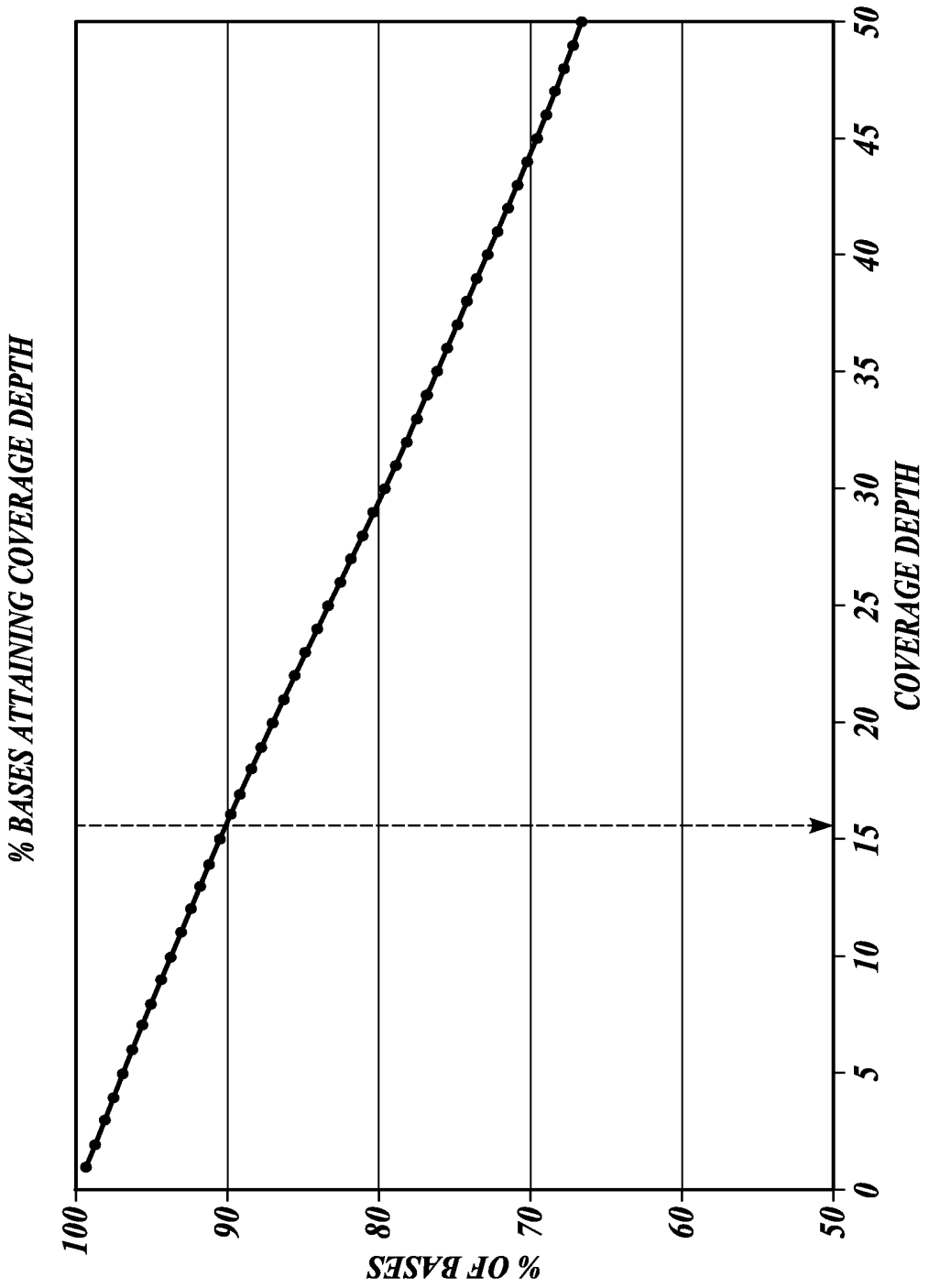

% BASES ATTAINING COVERAGE DEPTH
% OF BASES
100
90
80
70
60
50
0
5
10
15
20
25
30
35
40
45
50
COVERAGE DEPTH

Fig.9.

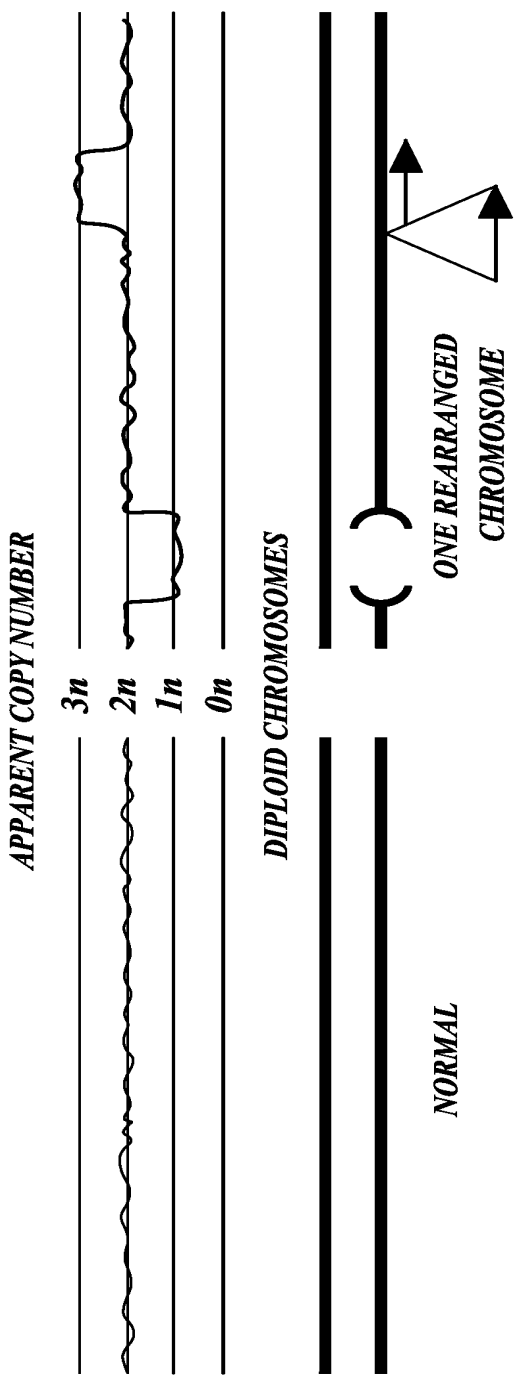


Fig.10A.

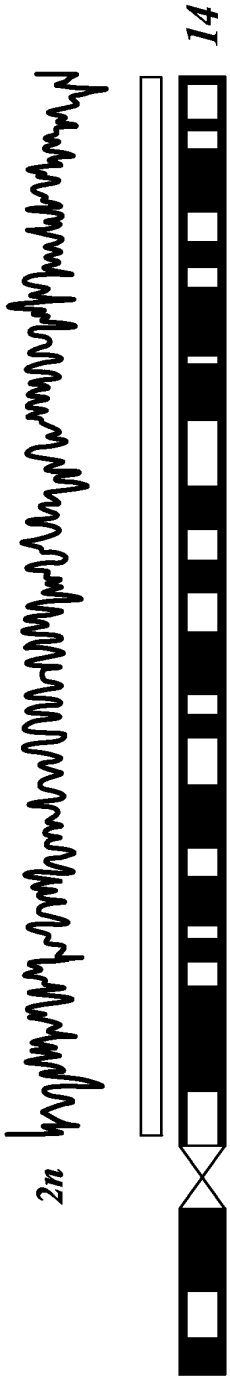


Fig.10B.

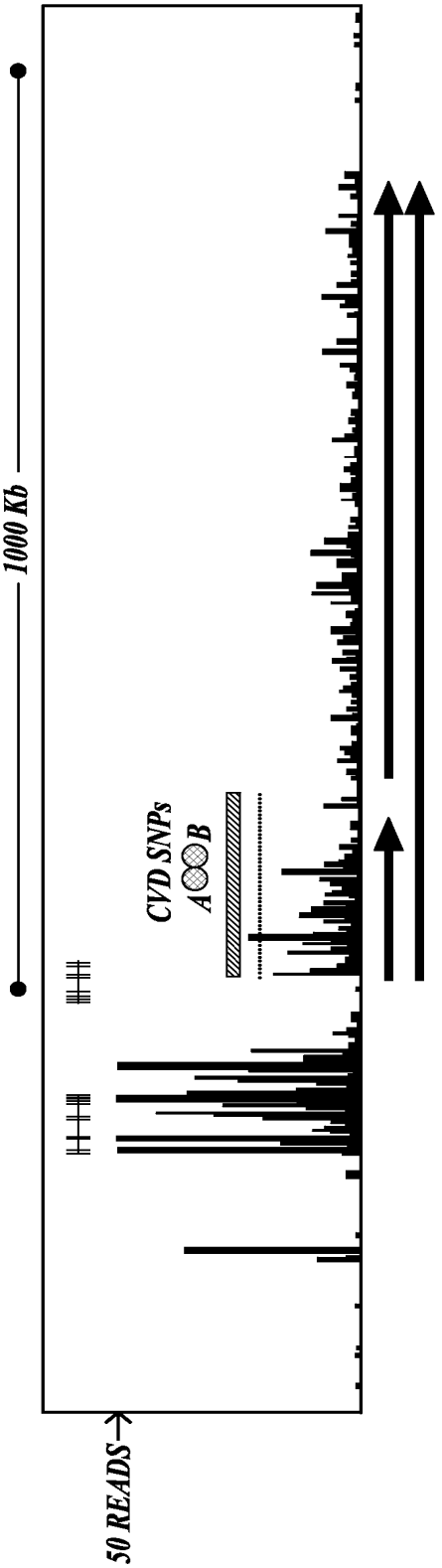


Fig.11A.

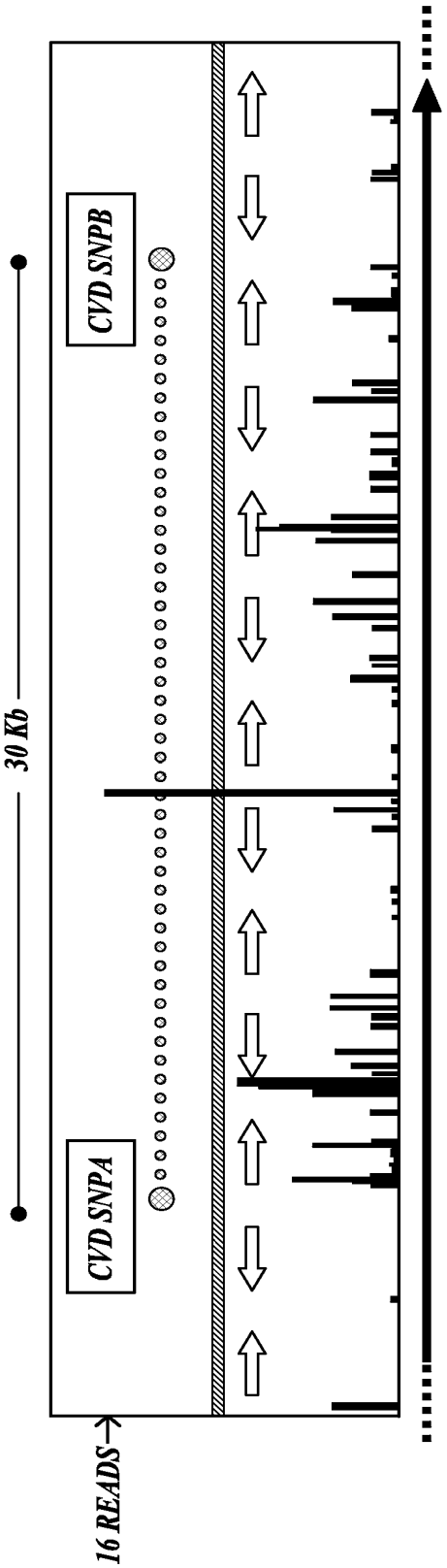


Fig.11B.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 09/56380

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12P 19/34 (2009.01)
USPC - 435/91.2
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
USPC: 435/91.2
IPC(8): C12P 19/34 (2009.01)

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)
WEST, Google Scholar: dna and linker and stem adj loop and insert and pcr and oligonucleotide and ligation and exon and cdna and ribosomal and dnase and manganese and bar adj code and antisense and probe and capture and complex and formamide and nacl and primer and population and library

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X ----- Y	US 2006/0051789 A1 (KAZAKOV et al.) 9 March 2006 (09.03.2006) para [0016], [0027], [0031], [0050], [0061]-[0065], [0078], [0142], [0161]-[0179]	1-6, 9, and 15 ---------- 7, 8, 10-14 and 16-31
Y	US 5,681,726 A (HUSE et al.) 28 October 1997 (28.10.1977) col 4, ln 33-35; col 14, ln 9-11	7, 8
Y	US 2007/0031857 A1 (MAKAROV et al.) 8 February 2007 (08.02.2007) para [0031, [0227], [0231], [0241]	10
Y	US 2004/0248153 A1 (DEAR et al.) 9 December 2004 (09.12.2004) para [0010] and [0068]	11 and 29
Y	US 2004/0115815 A1 (LI et al.) 17 June 2004 (17.06.2004) para [0006]	12
Y	US 2002/0150919 A1 (WEISMANN et al.) 17 October 2002 (17.10.2002) para [0022]-[0026], [0058], [0065]-[0068], [0102]-[0105], [0114]-[0118]	13, 14 and 16-31

☐ 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
12 October 2009 (12.10.2009)	22 OCT 2009
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

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

Espacenet

Bibliographic data: WO2007030759 (A2) — 2007-03-15

IMPROVED NUCLEIC ACID AMPLIFICATION PROCEDURE

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:** ***C12Q1/68***
- cooperative: **C12Q1/6865**

Application number: WO2006US35154 20060907 Global Dossier

Priority number (s): US20050714966P 20050907

Also published as: WO2007030759 (A3) CA2621267 (A1) EP1929046 (A2) EP1929046 (B1) US2009130721 (A1) more

Abstract of WO2007030759 (A2)

The invention provides methods for amplification of polynucleotide sequences using primers containing single-stranded RNA. The methods employ use of an enzyme capable of cleaving single-stranded RNA, such as RNase I, to degrade a first RNA-containing primer prior to addition of a second RNA-containing primer. The invention also provides compositions and kits for practicing the amplification methods, as well as methods which use the amplification products.

IDT Exhibit 1020.2134

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 March 2007 (15.03.2007)

PCT

(10) International Publication Number
WO 2007/030759 A2

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

(21) International Application Number:
PCT/US2006/035154

(22) International Filing Date:
7 September 2006 (07.09.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/714,966 7 September 2005 (07.09.2005) US

(71) Applicant *(for all designated States except US)*: **NUGEN TECHNOLOGIES, INC.** [US/US]; 821 Industrial Road, Unit A, San Carlos, CA (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: **JACOBSON, Jill, A.** et al.; Morrison & Foerster LLP, 755 Page Mill Road, Palo Alto, CA 94304-1018 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, 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, 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*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2007/030759 A2

(54) Title: IMPROVED NUCLEIC ACID AMPLIFICATION PROCEDURE

(57) Abstract: The invention provides methods for amplification of polynucleotide sequences using primers containing single-stranded RNA. The methods employ use of an enzyme capable of cleaving single-stranded RNA, such as RNase I, to degrade a first RNA- containing primer prior to addition of a second RNA-containing primer. The invention also provides compositions and kits for practicing the amplification methods, as well as methods which use the amplification products.

IDT Exhibit 1020.2135

IMPROVED NUCLEIC ACID AMPLIFICATION PROCEDURE

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 60/714,966, filed on September 7, 2005, which is hereby incorporated by reference in its entirety.

FIELD OF THE INVENTION

[0002] The invention relates generally to methods and compositions for improved efficiency of nucleic acid amplification procedures, and more specifically to use of a ribonuclease that is specific for single-stranded RNA to remove or reduce the concentration of single-stranded RNA of primers used in amplification procedures.

BACKGROUND

[0003] Composite primers comprising a 3'-DNA portion and an RNA portion are employed in previously-described DNA and RNA amplification methods, for example, as described in U.S. Patent Nos.6,251,639, 6,692,918, 6,815,164, 6,858,413, and 6,686,156, and in U.S. Application Nos. 2003-0087251, 2003-0003441, 2005-0014192, 2002-0164628, 2003-0215926, 2004-0023271, 2004-0005614, and 2005-0019793. Amplification of an RNA target, for example, as described in U.S. Application Nos. 2003-0087251, 2005-0003441, and 2005-0014192, and in PCT Application No. WO 02/072772, is initiated by a procedure for the generation of cDNA which utilizes composite primers (comprising an RNA and a 3'-DNA portion) for first strand cDNA synthesis. The RNA portion of the composite primer may comprise a sequence which does not hybridize to the target RNA sequence. The 3'-DNA portion comprises a sequence that hybridizes to the target RNA sequence. The 3'-DNA portion of the first strand cDNA primer may comprise a sequence that is complementary to the poly-A tail of mRNA, or a random sequence that is hybridizable to sequences across the RNA target sequence. Alternatively the 3'-DNA portion may comprise a sequence that is complementary to specific sequence(s) of the RNA target. The first strand cDNA synthesis is carried out by a reverse transcriptase, which extends the hybridized primer along the target RNA to form a cDNA/RNA heteroduplex. Any combination of first strand composite cDNA primers is possible. Thus,

IDT Exhibit 1020.2136

first strand cDNA synthesis may be carried out using a single composite primer, a mixture of composite primers with a random 3'-DNA sequence, such as a random hexamer, a combination of composite primers comprising random and sequence-specific 3'-DNA portions, etc. Second strand cDNA synthesis along the first strand cDNA, and reverse transcription of the RNA portion of the first strand primer extension product, results in the formation of unique double stranded cDNA molecules with a DNA/RNA heteroduplex at one end. The heteroduplex at the end of the double-stranded cDNA is a substrate for RNase H, which can degrade RNA of this heteroduplex to generate a unique partial duplex cDNA with a single-stranded DNA portion at the 3'-end of the second strand cDNA. This single-stranded sequence comprises a sequence that is complementary to the RNA portion of the first strand cDNA composite primer utilized, and serves as a priming site for subsequent amplification using a composite DNA/RNA amplification primer. Amplification is carried out using a composite amplification primer comprising a 3'-DNA portion and an RNA portion, a DNA polymerase with strand displacement activity, and an enzyme capable of degrading RNA in an RNA/DNA heteroduplex, such as RNase H.

[0004] In one procedure for amplifying an RNA target as described above, the first strand cDNA chimeric primer comprises a RNA portion that is not hybridizable to the target RNA sequence and comprises a sequence of the chimeric amplification primer. The double stranded cDNA generated at the completion of second strand cDNA synthesis, comprises a unique DNA/RNA heteroduplex at one end. The appended sequence at one end of the double stranded cDNA comprises the RNA portion of the first strand cDNA chimeric primer and its DNA complement. Incubation of this product with an enzyme that degrades RNA in an RNA/DNA heteroduplex, such as RNase H, results in the degradation of the RNA portion of the heteroduplex, releasing a site for primer hybridization to permit amplification with the chimeric amplification primer. Insofar as the amplification priming site contains sequences that are complementary to sequences in both the first strand cDNA chimeric primer and the chimeric amplification primer, any remaining first strand cDNA chimeric primer, which was not engaged in the synthesis of first strand cDNA, is capable of competing with the amplification chimeric primer for binding at the amplification primer binding site. This competition has the potential of impacting amplification efficiency. This competition is dependent on the concentrations of the two primers.

[0005] Whereas the chimeric amplification primer is typically added to the amplification reaction mixture at high concentration sufficient for effective productive

IDT Exhibit 1020.2137

hybridization and subsequent amplification to generate multiple copies of the single-stranded amplification products, the first strand primer is carried over into the amplification reaction mixture with the cDNA reaction mixture. The amount of first strand cDNA chimeric primer carried over into the amplification reaction mixture will be dependent on the amount added to the first strand synthesis reaction mixture, and the complexity of the primer composition. For example, the total amount of first strand chimeric primer may be particularly high when the primer employed is designed for random priming throughout the length of RNA transcripts, in contrast to a primer employed for cDNA synthesis which is initiated at defined sequences of RNA transcripts in the sample. The initiation of cDNA synthesis at specific sequences of RNA transcripts may entail hybridization and initiation of synthesis at the poly-A tail of eukaryotic mRNAs, or at sequences specific for defined internal mRNA species, such as a sequence common to a family of transcripts.

[0006] The amount (concentration) of the first strand chimeric primer added to a reaction mixture is often determined based on efficient priming and may therefore be in excess to the amount of RNA transcripts in the samples. Chimeric DNA/RNA primers designed to randomly prime cDNA synthesis throughout the length of an RNA transcript comprise a large population of primer sequences to accommodate the representation of random sequence at the 3' end, and thus require a large concentration of the total population of primers, for effective representation of each of the priming sequences. Similarly, the total concentration of primers added for priming at multiple transcript sequences will also be higher than that required for any single primer. Effective linear amplification of the entire population of transcripts in the mixture can be achieved with the use of a single amplification primer, when all the chimeric primers employed for first strand cDNA synthesis comprise an RNA portion of the same sequence. Thus, the effective combined concentration of the RNA portion of the chimeric primer is particularly high whereas the concentration of any of the 3' DNA portions is relatively low. The first strand cDNA chimeric primer remaining in the reaction mixture following the second strand synthesis reaction, and carried over into the amplification reaction mixture, is high when employing a random priming strategy. Insofar as all the RNA portions of the first strand chimeric primers include the RNA sequence of the chimeric amplification primer, the amplification efficiency may be impacted by competition of both sets of primers for binding to the priming site on the amplification target.

IDT Exhibit 1020.2138

[0007] There is a need for an improved amplification procedure in which single-stranded RNA of a first single-stranded-RNA-containing primer is degraded prior to a second reaction employing a second single-stranded-RNA-containing primer, to prevent or reduce competition between the two primers for binding to the target in an amplification reaction employing the second primer.

BRIEF SUMMARY OF THE INVENTION

[0008] The invention provides methods, compositions, and kits for amplification using primers containing single-stranded RNA, *e.g.*, RNA/DNA composite primers, as well as applications of the amplification methods. The methods described herein employ an enzyme capable of cleaving single-stranded RNA (*i.e.*, specific for single-stranded RNA) to cleave the RNA portion of excess unhybridized first composite primer prior to amplification with a composite amplification primer, and/or single-stranded RNA portions of amplification products

[0009] Accordingly, in one aspect, the invention provides methods for amplification of a template polynucleotide. In one embodiment, the method comprises: (a) extending a first primer hybridized to the polynucleotide template with at least one enzyme comprising a DNA polymerase activity, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, whereby a complex comprising a first primer extension product and the template polynucleotide is produced (b) extending a second primer hybridized to the first primer extension product to produce a complex comprising a first primer extension product and a second primer extension product; (c) incubating the reaction mixture comprising the products of step (b) with an enzyme that is capable of cleaving single-stranded RNA, whereby unhybridized single-stranded RNA of the RNA portion of the first primer is cleaved; (d) inactivating the enzyme that is capable of cleaving single-stranded RNA; (e) cleaving RNA from the first primer in the complex of first primer extension product and second primer extension product with at least one enzyme that cleaves RNA from an RNA/DNA hybrid such that a composite amplification primer hybridizes to the second primer extension product, wherein the composite amplification primer comprises an RNA portion and a 3' DNA portion; and (f) extending the composite amplification primer hybridized to the second primer extension product with at least one enzyme comprising DNA-dependent DNA polymerase activity, whereby the first primer extension product is displaced, RNA is cleaved from the composite amplification primer

IDT Exhibit 1020.2139

and another composite amplification hybridizes such that primer extension and strand displacement are repeated, and whereby multiple copies of a polynucleotide amplification product are generated.

[0010] In another embodiment, the method comprises:(a) incubating a reaction mixture, said reaction mixture comprising: (i) a template polynucleotide; (ii) a first primer, wherein the first primer is a composite primer that is hybridizable to at least one (*i.e.*, one or a multiplicity) template polynucleotide site(s), wherein the composite primer comprises an RNA portion and a 3' DNA portion; and (iii) a DNA-dependent DNA polymerase and/or an RNA-dependent DNA polymerase (which may be present as a separate enzyme or as an enzyme comprising both DNA-dependent DNA polymerase and RNA-dependent DNA polymerase activities); wherein the incubation is under conditions that permit composite primer hybridization and primer extension, whereby a complex comprising a first primer extension product and the template polynucleotide is generated; (b) incubating a reaction mixture, said reaction mixture comprising: (i) the first primer extension product; (ii) a second primer; (iii) at least one enzyme comprising DNA-dependent DNA polymerase activity; (iv) at least one enzyme comprising RNA-dependent DNA polymerase activity; and (v) optionally, at least one enzyme capable of cleaving RNA from an RNA/DNA hybrid; wherein the incubation is under conditions permitting formation of a complex comprising the first primer extension product and a second primer extension product; (c) incubating the reaction mixture of (b) (or an aliquot thereof) with an enzyme capable of cleaving single-stranded RNA (*e.g.*, RNase I); (d) inactivating the enzyme capable of cleaving single-stranded RNA (*e.g.*, by elevating the temperature of the reaction mixture) or removing the enzyme capable of cleaving single-stranded RNA from the reaction mixture; and (e) incubating a reaction mixture, said reaction mixture comprising (i) the reaction products generated according to steps (a), (b), (c), and (d) (or an aliquot thereof); (ii) a composite amplification primer, wherein the composite amplification primer comprises an RNA portion and a 3' DNA portion; (iii) a DNA-dependent DNA polymerase; and (iv) an enzyme that cleaves RNA from an RNA/DNA hybrid, *e.g.*, RNase H; wherein the incubation is under conditions that permit RNA cleavage from an RNA/DNA heteroduplex, primer hybridization, primer extension, and displacement of the first primer extension product from the complex comprising the first primer extension product and the second primer extension product, whereby another composite amplification

IDT Exhibit 1020.2140

primer hybridizes and primer extension and strand displacement are repeated; whereby multiple copies of a polynucleotide (generally, DNA) amplification product are generated.

[0011] In another embodiment, the method comprises: (i) contacting a reaction mixture comprising a first primer and a complex of first primer extension product and second primer extension product with an enzyme capable of cleaving single-stranded RNA, wherein the complex of first primer extension product and second primer extension product is produced by extension of a first primer hybridized to the template polynucleotide with a DNA polymerase, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, wherein the second primer extension product is generated by extension of a second primer hybridized to the first primer extension product; (ii) inactivating the enzyme capable of cleaving single-stranded RNA; and (iii) contacting the complex of first primer extension product and second primer extension product with at least one enzyme that cleaves RNA from an RNA/DNA hybrid and a composite amplification primer, said composite amplification primer comprising an RNA portion and a 3' DNA portion, wherein the composite amplification primer hybridizes to the second primer extension product; whereby said first primer extension product is displaced, RNA is cleaved from the composite amplification primer, and another composite amplification primer hybridizes such that primer extension and strand displacement are repeated; and whereby multiple copies of a polynucleotide amplification product are generated.

[0012] In some embodiments, the first composite primer comprises a random sequence or a partially randomized sequence In embodiments utilizing a composite primer with random or partially random sequence, the composite primer may be a population or pool of different primers comprising at least 2, at least 3, at least 4, at least 5, at least 10, at least 15, at least 20, at least 30, at least 40, at least 50, or at least 100 different sequences. In other embodiments, the composite primer contains one or more "degenerate" nucleotides that are able to hybridize to multiple different nucleotide bases (*e.g.*, inosine, which is able to hybridize to all four canonical bases).

[0013] The methods are applicable to amplifying any target polynucleotide, including, for example, DNA (such as genomic DNA, including human and other mammalian, bacterial, or viral genomic DNA, or any other species or combination of species) and RNA (including synthetic or natural RNA, such as total RNA, mRNA, noncoding RNA, ribosomal RNA, or viral RNA). One or more steps may be combined and/or performed sequentially (often in any order, as long as the requisite product(s) are able to be formed),

IDT Exhibit 1020.2141

and, as is evident, the invention includes various combinations of the steps described herein. It is also evident, and is described herein, that the invention encompasses methods in which the initial, or first, step is any of the steps described herein. For example, the methods of the invention do not require that the first step be hybridization of first composite primer. Methods of the invention encompass embodiments in which later, "downstream" steps are an initial step.

[0014] In another aspect, the invention provides methods for amplification of an RNA sequence of interest. In one embodiment, the invention provides a method for generating multiple copies of a sequence complementary to an RNA sequence of interest, comprising: (a) extending a first primer hybridized to a target RNA with at least one enzyme comprising RNA-dependent DNA polymerase activity, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, whereby a complex comprising a first primer extension product and the target RNA is produced; (b) cleaving the target RNA in the complex of step (a); (c) extending a second primer hybridized to the first primer extension product with at least one enzyme comprising DNA-dependent DNA polymerase activity and at least one enzyme comprising RNA-dependent DNA polymerase activity, whereby a second primer extension product is produced to form a complex of first and second primer extension products; (d) incubating the reaction mixture comprising the products of step (c) and excess first primer, if any, with an enzyme that is capable of cleaving single-stranded RNA, whereby unhybridized single-stranded RNA of the RNA portion of the first primer is cleaved; (e) inactivating the enzyme that is capable of cleaving single-stranded RNA; (f) cleaving RNA from the first primer in the complex of first and second primer extension products with at least one enzyme that cleaves RNA from an RNA/DNA hybrid such that a composite amplification primer hybridizes to the second primer extension product, wherein the composite amplification primer comprises an RNA portion and a 3' DNA portion; and (g) extending the composite amplification primer hybridized to the second primer extension product with at least one enzyme comprising DNA-dependent DNA polymerase activity; whereby said first primer extension product is displaced, RNA is cleaved from the composite amplification primer and another composite amplification primer hybridizes such that primer extension and strand displacement are repeated, and whereby multiple copies of a polynucleotide sequence complementary to the RNA sequence of interest are generated.

IDT Exhibit 1020.2142

[0015] In another embodiment, the invention provides a method of generating multiple copies of an RNA sequence of interest, comprising: (a) extending a first primer hybridized to a target RNA with at least one enzyme comprising RNA-dependent DNA polymerase activity, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, whereby a complex comprising a first primer extension product and the target RNA is produced; (b) cleaving the target RNA in the complex of step (a); (c) extending a second primer hybridized to the first primer extension product with at least one enzyme comprising DNA-dependent DNA polymerase activity and at least one enzyme comprising RNA-dependent DNA polymerase activity, whereby a second primer extension product is produced to form a complex of first and second primer extension products; (d) incubating the reaction mixture comprising the products of step (c) and excess first primer, if any, with an enzyme that is capable of cleaving single-stranded RNA, whereby unhybridized single-stranded RNA of the RNA portion of the first primer is cleaved; (e) inactivating the enzyme that is capable of cleaving single-stranded RNA; (f) cleaving RNA from the first primer in the complex of first and second primer extension products with at least one enzyme that cleaves RNA from an RNA/DNA hybrid such that a composite amplification primer hybridizes to the second primer extension product, wherein the composite amplification primer comprises an RNA portion and a 3' DNA portion; (g) extending said composite amplification primer hybridized to the second primer extension product with at least one enzyme comprising DNA-dependent DNA polymerase activity, whereby said first primer extension product is displaced, RNA is cleaved from the composite amplification primer and another composite amplification primer hybridizes such that primer extension and strand displacement are repeated; and (h) hybridizing the displaced first primer extension product with a polynucleotide comprising a propromoter and a region which is hybridizable to the displaced first primer extension product under conditions which allow transcription to occur by RNA polymerase, such that RNA transcripts are produced comprising sequences complementary to the displaced first primer extension product, whereby multiple copies of the RNA sequence of interest are generated.

[0016] In another embodiment, the invention provides a method of generating multiple copies of a polynucleotide sequence complementary to an RNA sequence of interest comprising: (a) incubating a reaction mixture, said reaction mixture comprising: (i) a target RNA; (ii) a first primer that is hybridizable to a target RNA, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion; and (iii) at least one

IDT Exhibit 1020.2143

enzyme comprising RNA-dependent DNA polymerase activity, wherein the incubation is under conditions that permit primer hybridization and primer extension, whereby a complex comprising a first primer extension product and the target RNA is formed; (b) incubating a reaction mixture, said reaction mixture comprising: (i) the first primer extension product; (ii) a second primer; (iii) at least one enzyme comprising DNA-dependent DNA polymerase activity;(iv) at least one enzyme comprising RNA-dependent DNA polymerase activity; and (v) optionally, at least one enzyme capable of cleaving RNA from an RNA/DNA hybrid; wherein the incubation is under conditions permitting formation of a complex comprising the first primer extension product and a second primer extension product; (c) incubating at least a portion of the reaction mixture of (b) with an enzyme capable of cleaving single-stranded RNA; (d) inactivating the enzyme capable of cleaving single-stranded RNA; and (e) incubating a reaction mixture, said reaction mixture comprising: (i) the reaction products generated according to steps (a), (b), (c), and (d); (ii) at least one enzyme capable of cleaving RNA from an RNA/DNA hybrid; (iii) a composite amplification primer, wherein the composite amplification primer comprises a RNA portion and a 3' DNA portion; (iv) at least one enzyme comprising DNA-dependent DNA polymerase activity, wherein the incubation is under conditions that permit cleavage of RNA, composite primer hybridization, primer extension, and displacement of the first primer extension product from the complex comprising the first primer extension product and the second primer extension product, whereby another composite amplification primer hybridizes and primer extension and strand displacement are repeated; whereby multiple copies of a polynucleotide sequence complementary to the RNA sequence of interest are generated.

[0017] In another embodiment, the invention comprises a method of generating multiple copies of a polynucleotide sequence complementary to an RNA sequence of interest, said method comprising: (i) contacting a reaction mixture comprising a complex of first and second primer extension products and a first primer with an enzyme capable of cleaving single-stranded RNA, wherein the first primer extension product is produced by extension of the first primer hybridized to a target RNA with at least one enzyme comprising RNA-dependent DNA polymerase activity, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, wherein the second primer extension product is generated by extension of a second primer hybridized to the first primer extension product; (ii) inactivating the enzyme capable of cleaving single-

IDT Exhibit 1020.2144

stranded RNA; and (iii) contacting the complex of first and second primer extension products with a composite amplification primer and at least one enzyme that cleaves RNA from an RNA/DNA hybrid, said composite amplification primer comprising an RNA portion and a 3' DNA portion, wherein the composite amplification primer hybridizes to the second primer extension product; whereby said first primer extension product is displaced, RNA is cleaved from the composite amplification primer and another composite amplification primer hybridizes such that primer extension and strand displacement are repeated; and whereby multiple copies of a polynucleotide sequence complementary to the RNA sequence of interest are generated.

[0018] In some embodiments of the methods described herein, the RNA portion of the first primer and/or amplification primer is 5' with respect to the 3' DNA portion. In some embodiments, the 5' RNA portion of the first primer and/or amplification primer is adjacent to the 3' DNA portion. In some embodiments, the first primer comprises a random sequence. In some embodiments, the first primer is hybridizable to a multiplicity of template polynucleotide sites. In some embodiments, the first primer is a tailed primer that comprises a 5' portion that is not hybridizable to the target RNA under conditions in which the first primer hybridizes to the target RNA. In some embodiments for amplification of mRNA, the 3' DNA portion of the first primer comprises a poly-dT sequence, and optionally comprises at least 1 random nucleotide at the 3' end.

[0019] In some embodiments, the second primer comprises DNA. In some embodiments, the second primer is a random primer. In some embodiments in which the target polynucleotide is RNA, the second primer comprises a fragment of the target RNA hybridized to the first primer extension product.

[0020] The enzymes which may be used in the methods and compositions are described herein. For example, the enzyme that cleaves RNA from an RNA/DNA heteroduplex may be an RNase H, and the RNA-dependent DNA polymerase may be reverse transcriptase. The RNA-dependent DNA polymerase may comprise an RNase H enzyme activity, or separate enzymes may be used. Similarly, a DNA polymerase may comprise both RNA-dependent and DNA-dependent DNA polymerase enzyme activities, or separate enzymes may be used. A DNA-dependent DNA polymerase, an RNA-dependent DNA polymerase, and the enzyme that cleaves RNA can also be the same enzyme, or separate enzymes comprising each of these activities may be used. The enzyme that is capable of cleaving single-stranded RNA may be, for example, RNase I, RNase T, RNase A, or combination

IDT Exhibit 1020.2145

thereof. Generally, an enzyme with little or no sequence specificity, and an enzyme that does not cleave the RNA in an RNA/DNA heteroduplex, is preferred. In some embodiments, the enzyme capable of cleaving single-stranded RNA is RNase I. In one embodiment, the RNase I is inactivated by heat.

[0021] In some embodiments, methods of the invention are used to generate labeled polynucleotide products (generally DNA products). In some embodiments of methods for generating labeled amplification, *e.g.*, DNA, products, at least one type of nucleotide, *e.g.*, dNTP, used is a labeled nucleotide, *e.g.*, dNTP. In other embodiments, a labeled nucleotide terminator is incorporated, for use in sequencing applications. In other embodiments of methods for generating labeled polynucleotide, *e.g.*, DNA, products, a labeled composite primer is used.

[0022] In some embodiments, a non-canonical nucleotide is incorporated into the amplification products by extension of the composite amplification primer in the presence of a non-canonical nucleotide. In one embodiment, the non-canonical nucleotide is dUTP. In some embodiments, polynucleotide products comprising a non-canonical nucleotide are fragmented and/or labeled, for example, according to methods described in U.S. Patent Application No. 2004/0005614.

[0023] In any of the methods described herein, the amplification products may be contacted with an enzyme that is capable of cleaving single-stranded RNA, *e.g.*, RNase I.

[0024] In another aspect, the invention provides methods for amplifying a polynucleotide sequence complementary to a target polynucleotide sequence, comprising: (a) extending a composite primer in a complex comprising (i) a polynucleotide template comprising the target sequence; and (ii) the composite primer, said composite primer comprising an RNA portion and a 3' DNA portion, wherein the composite primer is hybridized to the polynucleotide template; (b) cleaving the RNA portion of the annealed composite primer with an enzyme that cleaves RNA from an RNA/DNA hybrid such that another composite primer hybridizes to the template and repeats primer extension and strand displacement, whereby multiple copies of the complementary sequence of the target sequence are produced; and (c) incubating said multiple copies of the complementary sequence of the target sequence with an enzyme that is capable of cleaving single-stranded RNA, wherein RNA from the composite primer is cleaved.

[0025] In another aspect, the invention provides a method of generating multiple copies of a polynucleotide sequence complementary to an RNA sequence of interest, said method

IDT Exhibit 1020.2146

comprising: (a) extending a composite amplification primer in a complex comprising: (i) a complex of first and second primer extension products, wherein the first primer extension product is produced by extension of a first primer hybridized to a target RNA with at least one enzyme comprising RNA-dependent DNA polymerase activity, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, wherein the second primer extension product is generated by extension of a second primer hybridized to the first primer extension product, and wherein RNA from the complex of first and second primer extension products is cleaved with at least one enzyme that cleaves RNA from an RNA/DNA hybrid; and (ii) a composite amplification primer, said composite amplification primer comprising an RNA portion and a 3' DNA portion, wherein the composite amplification primer is hybridized to the second primer extension product; whereby said first primer extension product is displaced, RNA is cleaved from the composite amplification primer and another composite amplification primer hybridizes such that primer extension and strand displacement are repeated; and whereby multiple copies of a polynucleotide sequence complementary to the RNA sequence of interest are generated; and (b) contacting said multiple copies of the complementary sequence of the RNA sequence of interest with an enzyme that is capable of cleaving single-stranded RNA, wherein RNA from the composite primer is cleaved.

[0026] The invention also provides methods which employ (usually, analyze) the products of the amplification methods of the invention, such as detection of sequence alteration(s) (*e.g.*, genotyping, nucleic acid mutation detection, analysis of splice variants, and the like); determining presence or absence of a sequence of interest; quantifying a sequence of interest; gene expression profiling; subtractive hybridization; preparation of subtractive hybridization probe; differential amplification; preparation of libraries (including genomic, cDNA and differential expression libraries); preparation of an immobilized nucleic acid (which can be a nucleic acid immobilized on a microarray, preparing labeled probes for analysis on arrays (including high density arrays) for the detection and quantification of sequences of interest, including, for example, sequence determination, detecting sequence variation and genotyping; comparative genome hybridization; detection and/or identification of novel RNAs; and characterizing nucleic acids using the amplification nucleic acid products generated by the methods of the invention. Amplification products generated as described herein may also be used in gene expression analysis, including gene expression profiling or analysis of differential gene

IDT Exhibit 1020.2147

expression, for example, by quantitative PCR or analysis on microarrays for example, low, medium, or high density oligonucleotide or cDNA arrays.

[0027] Any of the methods of the invention can be used to generate polynucleotide products that are suitable for characterization of a polynucleotide sequence of interest in a sample. In one embodiment, the invention provides methods for characterizing (for example, detecting (presence or absence) and/or quantifying) a polynucleotide sequence of interest comprising: (a) amplifying a target polynucleotide by any of the methods described herein; and (b) analyzing the amplification products. Step (b) of analyzing the amplification products can be performed by any method known in the art or described herein, for example by detecting and/or quantifying amplification products that are hybridized to a probe. These amplification products may or may not be labeled. Any of the methods of the invention can be used to generate polynucleotide (such as DNA) products that are labeled by incorporating labeled nucleotides and/or labeled composite primers into appropriate step(s) of the methods. These labeled products are particularly suitable for quantification and/or identification by methods known in the art, which include the use of arrays such as cDNA microarrays and oligonucleotide arrays. In one aspect, the invention provides a method of characterizing a polynucleotide sequence of interest, comprising (a) amplifying a target polynucleotide by a method described herein to generate labeled polynucleotide products; and (b) analyzing the labeled polynucleotide products. In some embodiments, the step of analyzing polynucleotide products comprises determining amount of said products, whereby the amount of the polynucleotide sequence of interest present in a sample is quantified.

[0028] The amplification products can also serve as templates for further analysis such as sequence analysis, polymorphism detection (including multiplex SNP detection) using, *e.g.*, oligonucleotide ligation-based assays, analysis using Invader, Cleavase or limited primer extension, and other methods known in the art. For methods that generally require larger volumes of input material, the methods of the invention may be used to "pre" amplify a pool of polynucleotides to generate sufficient input material for subsequent analysis.

[0029] In another embodiment, the polynucleotide products can be analyzed by, for example, contacting them with at least one probe. In some embodiments, the at least one probe is provided as a microarray. The microarray can comprise at least one probe immobilized on a solid or semi-solid substrate fabricated from a material selected from the

IDT Exhibit 1020.2148

group consisting of paper, glass, ceramics, plastic, polypropylene, polystyrene, nylon, polyacrylamide, nitrocellulose, silicon, other metals, and optical fiber. A probe can be immobilized on the solid or semi-solid substrate in a two-dimensional configuration or a three-dimensional configuration comprising pins, rods, fibers, tapes, threads, beads, particles, microtiter wells, capillaries, and cylinders.

[0030] In another aspect, the invention provides methods of determining a gene expression profile in a sample, the methods comprising (a) amplifying RNA templates in a sample using any of the methods described herein; and (b) determining an amount of amplification products of each RNA sequence of interest in the sample, whereby the gene expression profile of the sample is determined. The invention further provides methods of determining a gene expression profile by determining an amount of amplification products of each RNA sequence of interest in a sample, the sample comprising multiple copies of RNA template amplified by any of the methods described herein, whereby the gene expression profile of the sample is determined.

[0031] Additionally, the invention also provides methods for archiving polynucleotide templates. In some embodiments, the amplification methods of the invention provide representative amplification of the sequences of the template polynucleotide, and amplified products produced by the these methods may be used as an archival source for the original template polynucleotide. Accordingly, the invention provides methods for archiving a polynucleotide template by storing the amplification products produced by the methods of the invention. The archived amplification products may be analyzed as described herein, or may be subjected to further amplification in accordance with the methods of the invention.

[0032] In one aspect, the invention provides methods for degrading excess RNA-containing primer in a nucleic acid amplification mixture that contains the RNA-containing primer and amplification products, by incubating the amplification mixture with an enzyme that specifically cleaves single-stranded RNA, such as RNase I, prior to use of the amplification products in subsequent applications.

[0033] In another aspect, the invention provides products (*e.g.*, multiple copies of a template polynucleotide or complement thereof) produced by the methods disclosed herein.

[0034] The invention also provides compositions, kits, complexes, reaction mixtures and systems comprising various components (and various combinations of the components) used in the amplification methods described herein.

IDT Exhibit 1020.2149

[0035] In another aspect, the invention provides compositions comprising any of the complexes (which are generally considered as intermediates with respect to the final amplification products), for example a complex of first and second primer extension product in a reaction mixture which has been incubated with an enzyme capable of cleaving single-stranded RNA, as described herein.

[0036] In another aspect, the invention includes any one or more products (including intermediates) and compositions comprising the products (including intermediates) produced by any aspect of the methods of the invention.

[0037] In another aspect, the invention provides reaction mixtures (or compositions comprising reaction mixtures) which contain various combinations of components described herein.

[0038] In another aspect, the invention provides kits for conducting the methods described herein. These kits, in suitable packaging and generally (but not necessarily) containing suitable instructions, contain one or more components used in the amplification methods.

[0039] In another aspect, the invention provides systems for effecting the amplification methods described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[0040] **Figure 1.** Schematic description of a procedure for mRNA amplification. **Figure 1A** depicts first and second strand synthesis. A first composite primer comprising a 3' poly-dT DNA sequence and a random nucleotide at the 3' end hybridizes to the poly-A tail of mRNA. The primer also includes a 5' RNA sequence which is not complementary to the mRNA sequences (unique sequence not complementary). After extension of the first primer, a second primer extension product is generated, forming a complex of first and second primer extension products with an RNA/DNA heteroduplex at one end. **Figure 1B** depicts cleavage of RNA from the first primer in the RNA/DNA heteroduplex with RNase H, which permits binding of an amplification primer and amplification of the target sequence using a DNA polymerase with strand displacement activity.

[0041] **Figure 2.** Schematic description of an RNA amplification method utilizing 3' DNA-5' RNA composite primers comprising a 3' portion comprising a random DNA sequence. **Figure 2A:** The amplification is initiated by hybridization of first composite

IDT Exhibit 1020.2150

primers to complementary sequences of the one or more (plurality of) mRNA species in the sample, and synthesis of first strand cDNAs by reverse transcriptase. Multiple first primer extension products are generated from the extension of the various random primers, each of the products comprising a 5' RNA sequence which is unique and non-complementary to the mRNA. **Figure 2B:** Second strand cDNA synthesis is carried out to generate multiple double stranded products comprising a common 5' end RNA/DNA heteroduplex. **Figure 2C:** RNase H activity cleaves the RNA portion of the heteroduplex to generate multiple partial duplex products comprising double stranded portions having sequence identity to the mRNAs in the sample, and a single-stranded DNA tail at one end comprising a unique sequence complementary to the unique RNA sequence of the composite primer. **Figure 2D:** These partial duplex products are substrates for amplification with a composite amplification primer. The composite amplification primer may be complementary to the cDNA complement of the unique sequence from the first primer, which may result in competition between the first composite primer and the composite amplification primer for binding to the single-stranded DNA sequence.

[0042] **Figure 3** schematically depicts RNA amplification using chimeric primers, incorporating treatment with RNase I following second strand cDNA synthesis. **Figure 3A:** Hybridization of a plurality of first chimeric primers comprising a 3' DNA portion with a random sequence to a mRNA target. **Figure 3B:** Formation of a double-stranded cDNA product with an RNA/DNA heteroduplex. **Figure 3C:** Degradation of RNA in the RNA/DNA heteroduplex to form product with 3' single-stranded DNA at one end. **Figure 3D:** Amplification without prior incubation with enzyme capable of degrading single-stranded RNA. **Figure 3E:** Amplification following treatment with RNase I to degrade excess first composite primers.

DETAILED DESCRIPTION

[0043] The invention provides methods, compositions, and kits for an improved nucleic amplification procedure. In a method such as an amplification method in which two primers containing single-stranded RNA are used, an enzyme capable of cleaving single-stranded RNA is employed to degrade single-stranded RNA of the first primer, prior to addition of a second single-stranded RNA containing primer. It is often desirable to

IDT Exhibit 1020.2151

remove a first single-stranded RNA containing primer (*i.e.*, one or a multiplicity of first primers) prior to addition of a second single-stranded RNA containing primer to reduce the concentration of excess first primer as a reaction mixture component, for example, in a procedure in which the first primer contains RNA sequences that could compete with the second primer for binding at a site to which the second primer is designed to hybridize.

[0044] An excess of a first primer could be removed from a reaction mixture by purification or separation of the reaction product from the primer (or primers) prior to incubation with the second primer. However, such a step is cumbersome and may increase variability in an amplification procedure. An "add and incubate" protocol is more advantageous and highly desirable for automation of the process. The methods of the invention do not require purification of a primer extension product from a single-stranded RNA-containing primer. According to the methods of the invention, an enzyme capable of degrading single-stranded RNA may be added to the primer extension reaction mixture for degradation of the primer prior to addition of a second primer. In the event that it is desirable to remove other competing or inhibitory components of the previous reaction prior to amplification, the removal of such other components may be achieved with various purification methods, which are well known in the art.

[0045] A major advantage afforded by the use of primers that contain an RNA portion, in addition to their utility in the isothermal amplification schemes described herein, is their unique susceptibility to RNA degrading enzymes which specifically recognize RNA sequences when free or hybridized to RNA or DNA complementary sequences. RNase H, which specifically degrades RNA sequences in an RNA/DNA heteroduplex, is employed in isothermal amplification methods described herein, *i.e.*, DNA and RNA isothermal amplification methods employing chimeric DNA/RNA primers. The specific recognition of RNA when hybridized to a complementary DNA sequence ensures the degradation of the RNA in the heteroduplex, for example, with RNase H, without affecting the RNA portion of unhybridized primer. The specific degradation of single-stranded RNA by an enzyme capable of specifically cleaving single-stranded RNA, *e.g.*, RNase I, RNase A, or RNase T, when not hybridized to form a DNA/RNA heteroduplex or an RNA/RNA hybrid, is advantageously used in the methods of the invention to remove or reduce the amount of excess, unhybridized single-stranded RNA containing primer. Often, RNase I is used in the methods of the invention, because it advantageously recognizes all four ribonucleotides, and thus its activity is not limited by sequence considerations. This enzyme is also heat

IDT Exhibit 1020.2152

labile and is easily inactivated by heat treatment, *i.e.*, incubation of the reaction mixture at an elevated temperature for a duration known to completely inactivate the enzyme. An added advantage of the use of RNase I for the degradation of the excess single-stranded RNA containing primer is its lack of requirement for specific divalent ions. Thus, RNase I is active within a reaction buffer mixture which is typically employed for the generation of double-stranded cDNA that can be subsequently amplified using a chimeric DNA/RNA amplification primer, DNA polymerase with strand-displacement activity, and RNase H activity.

[0046] In an amplification procedure comprising a first composite primer and a second composite primer, wherein the first composite primer comprises an RNA portion and a 3' DNA portion, wherein the second composite primer comprises an RNA portion and a 3' DNA portion, and wherein the second composite primer comprises a sequence that is hybridizable to a polynucleotide comprising a complement of the first composite primer, excess first composite primer can compete with the second composite primer when carried over into an amplification reaction using the second composite primer. This can be relieved by degradation of the RNA portion of the first chimeric primer prior to addition of the second composite primer to the reaction mixture. In an amplification procedure as described herein in which a first composite primer is used to generate a double-stranded cDNA with a RNA/DNA heteroduplex at one end, the RNA incorporated into the heteroduplex is not susceptible to degradation with an enzyme that specifically cleaves single-stranded RNA, such as RNase I. The remaining chimeric primer that was not consumed in the primer extension reaction is free in solution and its single-stranded 5'-RNA portion is susceptible degradation by RNase I, or another enzyme with single-stranded RNA specificity. Thus, the RNA portion of the remaining free first chimeric primer (or primers) can be degraded by incubation of the reaction mixture with RNase I.

[0047] In another embodiment of the invention, primer and RNA may be cleaned up following preparation of double stranded cDNA with a single stranded DNA at the 3'-end of the second strand cDNA. Preparation and utility of these partial duplexes have been previously described (U.S. Patent Application No. 2003/0215926). Treatment of the reaction mixture with one or more RNase enzyme specific for single-stranded RNA, *e.g.*, RNase I, achieves both removal of remaining primer(s) and the RNA target. Both free primer(s) and RNA may interfere with subsequent applications. The potential interference

IDT Exhibit 1020.2153

by the remaining free primer(s) is based on competition for binding to the 3'-end single-stranded DNA sequence of the partial duplex.

[0048] In a further embodiment of the invention, an enzyme that specifically cleaves single-stranded RNA, such as RNase I, may also be used for cleavage of the remaining ribose residues at the 5' end of RNA in an RNA/DNA heteroduplex after cleavage with an enzyme that cleaves RNA in such a heteroduplex, such as RNase H, in a single primer isothermal amplification procedure, such as the procedure described in U.S. Patent Nos. 6,251,639 and 6,692,918. It is well known that not all ribose residues of an RNA sequence are degraded by RNase H, when hybridized to a complementary DNA sequence. Thus, amplification products generated by the method utilizing a chimeric DNA/RNA primer, DNA polymerase with strand-displacement activity and RNase H, may include one or more ribose residue(s) at the 5'-end, which were incorporated with the partially hydrolyzed primer. Treatment of the amplification products by incubation with an enzyme capable of cleaving single-stranded RNA, such as RNase I, will advantageously remove residual ribose residues from the 5'-end to generate amplification products with a defined 5'-end sequence. This is particularly important when using the amplification methods described herein to produce amplification products for sequencing, where the sequencing ladders are better resolved when the 5' ends of the products are well defined.

General Techniques

[0049] The practice of the invention will 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).

[0050] Primers, oligonucleotides and polynucleotides employed in the invention can be generated using standard techniques known in the art.

IDT Exhibit 1020.2154

Definitions

[0051] A "template," "template strand," "template polynucleotide," "template DNA," "target sequence," "target nucleic acid," or "target DNA," "target polynucleotide," "template RNA," or "target RNA," as used herein, is a polynucleotide for which amplification is desired. The template polynucleotide can comprise DNA or RNA. The template sequence may be known or not known, in terms of its actual sequence. Generally, the terms "target," "template ," and variations thereof, are used interchangeably.

[0052] "Amplification," or "amplifying," as used herein, generally refers to the process of producing multiple copies of a desired sequence or its complement. "Multiple copies" means at least 2 copies. A "copy" does not necessarily mean perfect sequence complementarity or identity to the template sequence. For example, copies can include nucleotide analogs such as deoxyinosine, intentional sequence alterations (such as sequence alterations introduced through a primer comprising a sequence that is hybridizable, but not complementary, to the template, or a non-target sequence introduced through a primer), and/or sequence errors that occur during amplification. "Amplifying" a sequence may generally refer to making copies of a sequence or its complement, with the understanding that, as stated above, copying does not necessarily mean perfect sequence complementarity or identity with respect to the template sequence.

[0053] "Polynucleotide," or "nucleic acid," as used interchangeably herein, refers 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, carbamates, 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, poly-L-lysine, etc.), those with intercalators (*e.g.*, acridine,

IDT Exhibit 1020.2155

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, α-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 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.

[0054] A "labeled dNTP," or "labeled rNTP," as used herein, refers, respectively, to a dNTP or rNTP, or analogs thereof, that is directly or indirectly attached with a label. For example, a "labeled" dNTP or rNTP, may be directly labeled with, for example, a dye and/or a detectable moiety, such as a member of a specific binding pair (such as biotin-avidin). A "labeled" dNTP or rNTP, may also be indirectly labeled by its attachment to, for example, a moiety to which a label is/can be attached. A dNTP or rNTP, may comprise a moiety (for example, an amine group) to which a label may be attached following incorporation of the dNTP or rNTP into an extension product. Useful labels in the present invention include fluorescent dyes (*e.g.*, fluorescein isothiocyanate, Texas red, rhodamine, green fluorescent protein and the like), radioisotopes (*e.g.*, ^{3}H, ^{35}S, ^{32}P, ^{33}P, ^{125}I, or ^{14}C), enzymes (*e.g.*, LacZ, horseradish peroxidase, alkaline phosphatase,), digoxigenin, and

IDT Exhibit 1020.2156

colorimetric labels such as colloidal gold or colored glass or plastic (*e.g.*, polystyrene, polypropylene, latex, etc.) beads. Various anti-ligands and ligands can be used (as labels themselves or as a means for attaching a label). In the case of a ligand that has a natural anti-ligand, such as biotin, thyroxine and cortisol, the ligand can be used in conjunction with labeled anti-ligands.

[0055] The "type" of dNTP or rNTP, as used herein, refers to the particular base of a nucleotide, namely adenine, cytosine, thymine, uridine, or guanine.

[0056] "Nucleotide terminator," "polynucleotide chain terminator," "chain terminator" or "terminator" used herein refers to a nucleotide that when incorporated into a primer extension product prevents the further extension of such primer extension product. When the nucleotide terminator includes a ribofuranose sugar, the 3'-position must not have a hydroxyl group capable of being utilized by a polymerase enzyme to incorporate additional nucleotides. Examples of such nucleotide terminators include dideoxyadenosine triphoshate (ddATP), dideoxycytosine triphosphate (ddCTP), dideoxyguanosine triphosphate (ddGTP), dideoxythymidine triphosphate (ddTTP), and dideoxyuridine triphosphate (ddUTP). Alternatively, a ribofuranose analog may be used, such as arabinose. Reversible nucleotide terminators and acyclic terminators also may be used. A "labeled" nucleotide terminator includes a detectable label, as described above.

[0057] "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. Oligonucleotides in the invention include composite primers and auxiliary primers. The terms "oligonucleotide" and "polynucleotide" are not mutually exclusive. The description above for polynucleotides is equally and fully applicable to oligonucleotides.

[0058] A "primer," as used herein, refers to a nucleotide sequence, generally with a free 3'-OH group, that hybridizes with a template sequence (such as a target polynucleotide, target DNA, or a primer extension product) and is capable of promoting polymerization of a polynucleotide complementary to the template. A "primer" can be, for example, an oligonucleotide.

[0059] To "inhibit" is to decrease or reduce an activity, function, and/or amount as compared to a reference.

[0060] A "random primer," 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

IDT Exhibit 1020.2157

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 sequence of a random primer (or its complement) may or may not be naturally-occurring, or may or may not be present in a pool of sequences in a sample of interest. The amplification of a plurality of nucleic acid 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 primer) which collectively are designed to hybridize to a desired and/or a significant number of target sequences. A random primer may 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. Random primers may comprise a random sequence at the 3' end, but are not limited to 3' random sequences.

[0061] An "arbitrary primer" refers to a primer that hybridizes to a multiplicity of target sequences. The arbitrary hybridization of such a primer to target sequences may be achieved via partial complementarity to a multiplicity of sequences along the polynucleotide target.

[0062] "Protopromoter sequence," and "propromoter sequence," as used herein, refer to a single-stranded DNA sequence region which, in double-stranded form is capable of mediating RNA transcription. In some contexts, "protopromoter sequence," "protopromoter," "propromoter sequence," "propromoter," "promoter sequence," and "promoter" are used interchangeably.

[0063] A "propromoter polynucleotide," as used herein, refers to a polynucleotide comprising a propromoter sequence. Example of a propromoter polynucleotide is a propromoter template oligonucleotide (PTO).

[0064] "Propromoter template oligonucleotide (PTO)"and "promoter template oligonucleotide (PTO)" as used herein, refer to an oligonucleotide that comprises a propromoter sequence and a portion, generally a 3' portion, that is hybridizable (under a given set of conditions) to the 3' region of a primer extension product. The propromoter sequence and the hybridizable portion may be the same, distinct or overlapping nucleotides of an oligonucleotide.

IDT Exhibit 1020.2158

[0065] As used herein, "complex comprising an RNA/DNA partial heteroduplex" generally refers to a population of intermediate complexes that generally includes (a) copies of template polynucleotide and/or copies of the complement of template polynucleotide sequence appended to composite primer sequences; and (b) copies of template polynucleotide and/or copies of the complement of template polynucleotide appended to the complement of composite primer sequences. By virtue of the presence of composite primer sequence in the intermediate complexes, the complexes comprise at least a RNA/DNA partial heteroduplex. The RNA portion of the partial heteroduplex generally is introduced (via extension) by the RNA portion of the composite primer, and the DNA portion of the partial heteroduplex comprises the complement of the RNA portion of the composite primer. As discussed herein, the complex comprising an RNA/DNA partial heteroduplex functions as a substrate for further amplification during the single primer isothermal amplification phase of the methods. Generally, RNA in the RNA/DNA partial heteroduplex is cleaved, generating a 3' single stranded portion with sequences that are complementary to RNA in a composite amplification primer (and thus forming a binding site for a composite amplification primer). Thus, reference to "a complex comprising a 3' single stranded portion" generally refers to the complex comprising an RNA/DNA partial heteroduplex when its RNA is cleaved.

[0066] A "complex" is an assembly of components. A complex may or may not be stable and may be directly or indirectly detected. For example, as is described herein, given certain components of a reaction, and the type of product(s) of the reaction, existence of a complex can be inferred. For purposes of this invention, a complex is generally an intermediate with respect to the final amplification product(s). An example of a complex is a complex of composite primer extension product and second composite primer extension product, as described herein.

[0067] A "portion" or "region," used interchangeably herein, of a polynucleotide or oligonucleotide is a contiguous sequence of 2 or more bases. In other embodiments, a region or portion is at least about any of 3, 5, 10, 15, 20, 25 or more contiguous nucleotides.

[0068] A region, portion, or sequence which is "adjacent" to another sequence directly abuts that region, portion, or sequence. For example, an RNA portion which is adjacent to a 5' DNA portion of a composite primer directly abuts that region.

IDT Exhibit 1020.2159

[0069] A "reaction mixture" is an assemblage of components, which, under suitable conditions, react to form a complex (which may be an intermediate) and/or a product(s).

[0070] "A", "an" and "the", and the like, unless otherwise indicated include plural forms. "A" fragment means one or more fragments.

[0071] Conditions that "allow" an event to occur or conditions that are "suitable" for an event to occur, such as hybridization, strand extension, 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 nucleotide sequence, temperature, and buffer conditions. These conditions also depend on what event is desired, such as hybridization, cleavage, and/or strand extension.

[0072] Sequence "mutation," as used herein, refers to any sequence alteration in a sequence of interest in comparison to a reference sequence. A reference sequence can be a wild type sequence or a sequence to which one wishes to compare a sequence of interest. A sequence mutation includes single nucleotide changes, or alterations of more than one nucleotide in a sequence, due to mechanisms such as substitution, transversion, deletion or insertion. Single nucleotide polymorphism (SNP) is also a sequence mutation as used herein.

[0073] "Microarray" and "array," as used interchangeably herein, comprise a surface with an array, preferably ordered array, of putative binding (*e.g.*, by hybridization) sites for a biochemical sample (target) which often has undetermined characteristics. In a preferred embodiment, a microarray refers to an assembly of distinct polynucleotide or oligonucleotide probes immobilized at defined locations on a substrate. Arrays are formed on substrates fabricated with materials such as paper, glass, plastic (*e.g.*, polypropylene, nylon, polystyrene), polyacrylamide, nitrocellulose, silicon, optical fiber or any other suitable solid or semi-solid support, and configured in a planar (*e.g.*, glass plates, silicon chips) or three-dimensional (*e.g.*, pins, fibers, beads, particles, microtiter wells, capillaries) configuration. Probes forming the arrays may be attached to the substrate by any number of ways including (i) *in situ* synthesis (*e.g.*, high-density oligonucleotide arrays) using photolithographic techniques (see, Fodor et al., *Science* (1991), 251:767-773; Pease et al., *Proc. Natl. Acad. Sci. U.S.A.* (1994), 91:5022-5026; Lockhart et al., *Nature Biotechnology* (1996), 14:1675; U.S. Pat. Nos. 5,578,832; 5,556,752; and 5,510,270); (ii) spotting/printing at medium to low-density (*e.g.*, cDNA probes) on glass, nylon or nitrocellulose (Schena et

IDT Exhibit 1020.2160

al, *Science* (1995), 270:467-470, DeRisi et al, *Nature Genetics* (1996), 14:457-460; Shalon et al., *Genome Res.* (1996), 6:639-645; and Schena et al., *Proc. Natl. Acad. Sci. U.S.A.* (1995), 93:10539-11286); (iii) by masking (Maskos and Southern, *Nuc. Acids. Res.* (1992), 20:1679-1684) and (iv) by dot-blotting on a nylon or nitrocellulose hybridization membrane (see, *e.g.*, Sambrook et al., Eds., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., Vol. 1-3, Cold Spring Harbor Laboratory (Cold Spring Harbor, N.Y.)). Probes may also be noncovalently immobilized on the substrate by hybridization to anchors, by means of magnetic beads, or in a fluid phase such as in microtiter wells or capillaries. The probe molecules are generally nucleic acids such as DNA, RNA, PNA, and cDNA but may also include proteins, polypeptides, oligosaccharides, cells, tissues and any permutations thereof which can specifically bind the target molecules.

[0074] The term "3'" generally refers to a region or position in a polynucleotide or oligonucleotide 3' (downstream) from another region or position in the same polynucleotide or oligonucleotide.

[0075] The term "5'" generally refers to a region or position in a polynucleotide or oligonucleotide 5' (upstream) from another region or position in the same polynucleotide or oligonucleotide.

[0076] The term "3'-DNA portion," "3'-DNA region," "3'-RNA portion," and "3'-RNA region," refer to the portion or region of a polynucleotide or oligonucleotide located towards the 3' end of the polynucleotide or oligonucleotide, and may or may not include the 3' most nucleotide(s) or moieties attached to the 3' most nucleotide of the same polynucleotide or oligonucleotide. The 3' most nucleotide(s) can be preferably from about 1 to about 50, more preferably from about 10 to about 40, even more preferably from about 20 to about 30 nucleotides.

[0077] The term "5'-DNA portion," "5'-DNA region," "5'-RNA portion," and "5'-RNA region," refer to the portion or region of a polynucleotide or oligonucleotide located towards the 5' end of the polynucleotide or oligonucleotide, and may or may not include the 5' most nucleotide(s) or moieties attached to the 5' most nucleotide of the same polynucleotide or oligonucleotide. The 5' most nucleotide(s) can be preferably from about 1 to about 50, more preferably from about 10 to about 40, even more preferably from about 20 to about 30 nucleotides.

IDT Exhibit 1020.2161

[0078] "Absent" or "absence" of product, and "lack of detection of product" as used herein includes insignificant, or *de minimus* levels, generally due to lack of significant accumulation of product.

Amplification methods of the invention

[0079] The following are examples of the nucleic acid amplification methods of the invention. It is understood that various other embodiments may be practiced, given the general description provided above. For example, reference to using a composite primer means that any of the composite primers described herein may be used.

[0080] In one embodiment, the invention provides nucleic acid amplification methods in which an RNA/DNA composite primer is used. In some embodiments, the method includes global amplification using a composite primer that is capable of binding to multiple sites within a template polynucleotide, including DNA and RNA template polynucleotides. In some embodiments, the method includes amplification of mRNA using a composite primer containing a 3' poly-dT sequence. In some embodiments, the method includes amplification of a nucleic acid template with a composite primer containing at least one random nucleotide at the 3' end or an arbitrary nucleotide sequence. In some embodiments, a multiplicity of first primers is used. In some embodiments, the first primer is a tailed primer comprising a 5' portion which is not hybridizable to the template under conditions in which the first primer hybridizes to the template.

[0081] Generally, the methods of the invention involve two phases: (a) hybridization of a first primer to a polynucleotide target sequence, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, primer extension to form a first primer extension product, and extension of a second primer to form a second primer extension product, whereby a complex comprising the first and second extension products in an RNA/DNA partial heteroduplex is generated; and (b) composite-primer dependent single primer isothermal amplification using a composite amplification primer which comprises a sequence that is hybridizable to a polynucleotide comprising a complement of the first composite primer, wherein the composite amplification primer comprises an RNA portion and a 3' DNA portion. In the methods of the invention, an enzyme capable of cleaving single-stranded RNA is added to remove or reduce the amount of excess first composite primer which was not incorporated into the first primer extension product and

IDT Exhibit 1020.2162

subsequently into the partial heteroduplex of first and second primer extension products, before addition of the composite amplification primer.

Formation of a complex of first and second primer extension products

[0082] Using the first primer extension product as a template, a second primer extension product complementary to the first primer extension product is generated by extension of a second primer with a DNA-dependent DNA polymerase along the DNA portion of the first primer extension product, and extension by an RNA-dependent DNA polymerase along the 5' RNA portion of the first primer extension product, generating a double stranded complex comprising a RNA/DNA complex at the end. Generation of second primer extension product may be primed with a second primer comprising a random sequence or may be primed by the 3' end of a different composite primer extension product. In some embodiments, second strand production is primed by exogenous (added) primers and/or by fragments of template RNA (endogenous primers). In some embodiments, second primer extension product is primed with specific exogenous primer(s), which may be composite primer(s) or non-composite primer(s).

Incubation with an enzyme capable of cleaving single-stranded RNA

[0083] RNA sequences of unhybridized first primer are degraded by incubation of a reaction mixture containing complexes of first and second primer extension products and excess first primer, with an enzyme capable of cleaving single-stranded RNA, such as RNase I. The enzyme is inactivated or removed from the reaction mixture prior to addition of a second composite primer for amplification.

Single primer isothermal amplification using a complex comprising an RNA/DNA partial heteroduplex as a template

[0084] In a further phase of the methods, after cleavage of RNA of the first primer with an enzyme capable of cleaving single-stranded RNA and inactivation or removal of the enzyme, the complex comprising an RNA/DNA partial heteroduplex is a substrate for amplification using a single primer isothermal amplification procedure as follows: An enzyme which cleaves RNA sequence from an RNA/DNA hybrid (such as RNase H) cleaves RNA from the partial heteroduplex of first and second primer extension products, leaving a partially double stranded polynucleotide complex comprising a 3' single stranded

IDT Exhibit 1020.2163

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 RNA sequence in the first composite primer, and thus forms a specific binding site for a composite amplification primer (which may or may not be the same as the first composite primer, *i.e.*, the first primer and the composite amplification primer may be the same or different). In some embodiments, the first primer comprises a random sequence in the 3' DNA portion, and the composite amplification primer does not comprise the random sequence. In some embodiments, the composite amplification primer comprises a shorter total polynucleotide sequence length than the first primer, for example, a shorter length 3' DNA portion. In some embodiments, the composite amplification primer comprises a sequence that is the same as the RNA portion of the first composite primer. In some embodiments, a plurality of first composite primers are used wherein the RNA portions of the first composite primers are identical and the DNA portions of the first composite primers comprise different sequences.

[0085] Extension of a bound composite amplification 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. Patents Nos. 6,251,639, 6,692,918, and 6,946,251.

[0086] 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 amplification primer hybridization, primer extension, cleavage of RNA from an RNA/DNA hybrid and strand displacement. Insofar 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. Thus, in some embodiments, the reaction conditions permit stringent hybridization (*i.e.*, hybridization of sequences that are generally complementary). As is evident from the description herein, in other embodiments, the reaction conditions are less stringent (*i.e.*, permit hybridization of sequences that are less than fully complementary).

IDT Exhibit 1020.2164

[0087] Generally, 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 50, at least 100, at least 200, at least 300, or more distinct sequences (such as a gene or other subsegment of a polynucleotide, a marker (such as a SNP or other polymorphism) are amplified, and 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 chromosomes or portion of a chromosome for genomic DNA 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 embodiments, the template polynucleotide sequences that are amplified comprise a multiplicity of target polynucleotides representing non-related nucleic acid sequences, such as mixtures of bacterial or viral DNA or RNA targets.

[0088] For convenience, reference is made to a polynucleotide product. In some embodiments, amplified product is a mixture of sense and antisense copies of a given template polynucleotide. For example, if the template polynucleotide is double stranded DNA, the amplification product will correspond to each strand. If the template polynucleotide is single stranded (*e.g.*, RNA or single stranded DNA), amplification product may be produced that is the copy of template polynucleotide (sense copy) and the complement of the template polynucleotide (antisense copy). In the case of global amplification of genomic DNA employing arbitrary primer(s), amplification to produce copies of both strands of the genomic targets is expected. The amplification product of different senses can be annealed to form a double stranded (or partially double stranded) complex, or 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.

[0089] As illustrated in these embodiments, the methods of the invention are composite-primer dependent (*i.e.*, amplification is not observed in the absence of the composite primer) and 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

IDT Exhibit 1020.2165

provided above. For example, as described and exemplified herein, certain steps may be performed as temperature is changed (*e.g.*, raised, or lowered).

Amplification of RNA sequences

[0090] In one aspect, the invention provides methods for amplification of a single RNA species or a pool of RNA species. Some methods provide for generation of multiple copies of DNA comprising sequences complementary to an RNA sequence of interest. Other methods provide for generation of multiple copies of an RNA sequence of interest. These methods are suitable for, for example, generation of cDNA libraries and subtractive hybridization probes. Single stranded DNA or RNA products are generated, which are readily suitable for a variety of uses including expression profiling, *e.g.*, by multiplex analysis by microarray technologies, as well as electrophoresis-based technologies such as differential display. The methods are amenable to automation and do not require thermal cycling.

[0091] The methods of the invention are directed to the amplification of one or more species of RNA, such as a pool of RNA sequences, and are especially suitable for the amplification of all RNA (such as mRNA or the whole transcriptome) sequences in a preparation of total RNA from a biological sample. Thus, one of the major advantages of the methods of the invention is the ability to amplify an entire pool of sequences, which is essential for the ability to analyze the gene expression profile in cells, such as the cells in a biological sample of interest. The methods of the invention may also be used to amplify a specific RNA sequence of interest, or a multiplicity of RNAs, for example, members of a family of related RNAs. The methods of the invention also are suitable for amplifying a large multiplicity, and most preferably all RNA (such as mRNA) sequences in a sample.

[0092] Insofar as many mRNAs have a unique poly-A 3'-end, amplification initiated from the 3'-end sequence of mRNAs is commonly performed for preparation of cDNA libraries and subsequent sequence analysis for determination of gene expression profiling or other applications. The methods of the invention are similarly suited for preparation of libraries of amplified 3'-portions of mRNAs. A composite primer used in the methods of invention can be designed to be hybridizable to a multiplicity, or all, of the RNA or all mRNA species in the sample by using a composite primer comprising random sequences, according to methods known in the art. Alternatively, if a selected RNA or family of

IDT Exhibit 1020.2166

related RNAs are to be amplified, the composite primer will comprise sequence(s) hybridizable to the selected RNA or family of related RNAs.

[0093] The methods generally comprise using specially-designed primers, generally one or more RNA/DNA composite primers, to generate a complex of first and second strand cDNAs that comprise a portion with a particular characteristic (*e.g.*, cleavable by an enzyme). As used herein, it is understood that "cDNA" refers to a polynucleotide primer extension product. Generally, the complex comprises an RNA/DNA heteroduplex at an end of the double stranded cDNA complex. The RNA/DNA heteroduplex at an end of the double stranded cDNA complex may comprise a defined end sequence, generally introduced by the RNA portion of the composite primer. The composite primer according to the methods of the invention comprises a 3'-DNA portion that generally is designed to be hybridizable to a target RNA(s). The remaining portion(s) (such as the 5' RNA portion) of the composite primer generally, but not necessarily, comprises a sequence that is not hybridizable to a target RNA (which would constitute a tail when the primer is bound to a target RNA). Thus, and as the description herein makes clear, reference to a primer that hybridizes to a sequence (or hybridization template) encompasses embodiments in which at least a portion of the primer is hybridized, as well as those embodiments in which an entire primer is hybridized.

[0094] After production of a complex of first and second primer extension products, the reaction mixture containing the complex and excess first primer, if any, is contacted with an enzyme capable of cleaving single-stranded RNA. In one embodiment, the enzyme is RNase I. The enzyme capable of cleaving single-stranded RNA is inactivated or removed prior to addition of a composite amplification primer. In one embodiment, the enzyme is inactivated by heat.

[0095] The double stranded cDNA complex is a substrate for linear amplification as follows: An enzyme which cleaves RNA from an RNA/DNA hybrid (such as RNase H) cleaves RNA sequence from the hybrid, leaving a 3' DNA sequence on the second strand cDNA available for binding by a composite amplification primer which may or may not be the same as the first composite primer. Extension of a bound composite amplification primer by DNA polymerase with strand displacement activity produces a primer extension product, which displaces the previously bound cleaved first primer extension product, whereby single stranded DNA product accumulates. The single stranded DNA product is a

IDT Exhibit 1020.2167

copy of the complement of the target RNA (or "antisense" DNA). This linear amplification is described in Kurn et al., U.S. Patent Nos. 6,251,639, 6,692,918, and 6,946,251.

[0096] In one aspect, the invention works as follows: A composite RNA/DNA primer forms the basis for replication of the template RNA The composite primer (also referred to herein as "first composite primer" or "first primer") hybridizes to template RNA which comprises the RNA sequence(s) of interest, and the composite primer is extended by an RNA-dependent DNA polymerase to form a first primer extension product (interchangeably called "composite primer extension product", or "first-strand cDNA"). After cleavage of the template RNA, a second primer extension product (interchangeably called "second-strand cDNA") is formed (as described below) in a complex with the first primer extension product. The complex of first and second primer extension products is composed of an RNA/DNA hybrid at one end due to the presence of the composite primer in the first primer extension product. An agent such as an enzyme which cleaves RNA from an RNA/DNA hybrid (such as RNase H) cleaves RNA sequence from the hybrid, leaving a sequence on the second primer extension product available for binding by another composite primer (also referred to as "second composite primer" or "composite amplification primer" herein), which may comprise a sequence that is hybridizable to a polynucleotide comprising a complement of the first composite primer. Prior to addition of the composite amplification primer, excess first composite primer (*i.e.*, unhybridized and unincorporated into first primer extension product) is degraded by an enzyme capable of cleaving single-stranded RNA, and the enzyme capable of cleaving single-stranded RNA is then inactivated or removed. The composite amplification primer hybridizes to the single-stranded DNA sequence produced by treatment of the complex of first and second extension products with the agent which cleaves RNA from an RNA/DNA hybrid (*e.g.*, RNase H), and is extended with a DNA polymerase having strand displacement activity, which displaces the previously bound cleaved first primer extension product, resulting in displaced cleaved first primer extension product.

[0097] In some embodiments of the invention, the second primer extension product is formed as follows: Following cleavage of the RNA template, a second primer is then hybridized to the first primer extension product and extended to form a second primer extension product in a complex with the first primer extension product. The complex of first and second primer extension products is composed of an RNA/DNA hybrid at one end due to the presence of the composite primer in the first primer extension product. The

IDT Exhibit 1020.2168

second primer is any sequence that is hybridizable to the first DNA strand such that it is capable of being extended by a DNA polymerase along a first primer extension product to create a second primer extension product. Thus, in some embodiments, the second primer is an oligonucleotide, which may or may not comprise a random sequence (*i.e.*, a sequence designed to be hybridizable (under a given set of conditions) to one or more sequences in the sample). In other embodiments, the second primer comprises a sequence of the first DNA strand (generally at the 3' end) that is hybridized to a sequence in the first DNA strand (for example, a hairpin or self-annealed structure).

[0098] In another aspect of the amplification methods, one or more fragments of the target RNA serves as the primer of the second primer extension product. The target RNA in the initial complex comprising the target RNA and first primer extension product is cleaved with an agent (such as RNase H or heat) such that at least one fragment of the template RNA remains hybridized to the first primer extension product. In this aspect of the invention, one (or more) template RNA fragment(s) serves as a second "primer" in the manner described above, to generate a fragment extension product which has the same function as the second primer extension product in the amplification methods described above. A suitable RNA fragment in the methods of the invention is long enough such that it does not dissociate from the first strand cDNA, preferably from about 3 to about 30, more preferably from about 5 to about 25, even more preferably from about 10 to about 20, and most preferably from about 12 to about 17, nucleotides in length.

[0099] In embodiments involving transcription (referred to herein as "enhanced" methods), the second primer may further comprise a sequence such that displaced first primer extension products (other than the very first composite primer extension product) contain a sequence to which a polynucleotide comprising a propromoter (also referred to herein as "propromoter polynucleotide") is capable of hybridizing. Hybridization of the propromoter polynucleotide to a displaced primer extension product and extension of the 3' end of the displaced first primer extension product (if there is an overhang) results in a double stranded promoter region that drives transcription (via DNA-dependent RNA polymerase) to produce sense RNA products. This "enhanced" approach is described in Kurn et al., U.S. Patent Nos. 6,251,639, 6,692,918, and 6,946,251.

[0100] Accordingly, the invention provides methods of producing at least one copy of a polynucleotide sequence complementary to an RNA sequence of interest comprising combining and reacting the following: (a) a target RNA comprising an RNA sequence of

IDT Exhibit 1020.2169

interest; (b) a first (composite) primer comprising an RNA portion and a 3' DNA portion; (c) a second primer that is hybridizable to an extension product of the composite primer; (d) an RNA-dependent DNA polymerase; and (e) a DNA-dependent DNA polymerase; whereby a complex of first and second primer extension products is produced. The reaction mixture comprising the complexes of first and second primer extension products and excess first primer, if any, is contacted with an enzyme capable of cleaving single stranded RNA to reduce or remove single-stranded RNA of the first primer. The enzyme capable of cleaving single stranded RNA is then inactivated or removed from the reaction mixture. The following are then combined and reacted: (f) an enzyme that cleaves RNA from an RNA/DNA hybrid; (g) deoxyribonucleoside triphosphates or suitable analogs (which may or may not be labeled); (h) a composite amplification primer comprising an RNA portion and a 3' DNA portion; and (i) a DNA polymerase with strand displacement activity. This combination is subjected to suitable conditions for composite amplification primer hybridization, extension of the primer, RNA cleavage, and displacement of the first primer extension product, wherein RNA from the first composite primer is cleaved and a composite amplification primer binds in the site vacated by the cleaved RNA.

[0101] In embodiments that include transcription (*i.e.*, the enhanced methods), the following are also included in the amplification reaction (either at the same time as the components listed above or added separately): (j) a propromoter polynucleotide comprising a propromoter and a region which hybridizes to an extension product of the first primer (a displaced primer extension product); (k) an RNA polymerase; and (l) ribonucleoside triphosphates or suitable analogs (which may or may not be labeled), in place of (g). In embodiments that include transcription, conditions employed are also suitable for hybridization of the propromoter polynucleotide to the displaced cleaved first primer extension product, extension of the 3' end of the cleaved first primer extension product (if necessary) to generate a double-stranded promoter region, and RNA transcription driven by the promoter.

[0102] As described and exemplified herein, the above-described reaction mixtures may be subdivided into two or more different reaction mixtures for separate, generally sequential, incubations that correspond to different aspects of the amplification process.

[0103] In some embodiments, the invention provides methods of producing at least one copy of a polynucleotide sequence complementary to an RNA sequence of interest comprising combining and reacting the following: (a) complex of first and second primer

IDT Exhibit 1020.2170

extension products comprising an RNA/DNA heteroduplex at one end, produced as described above, wherein the reaction mixture comprising the complex of first and second extension products and first primer is contacted with an enzyme capable of cleaving single stranded RNA and the enzyme is inactivated or removed prior to contact with a composite amplification primer; (b) a composite amplification primer comprising an RNA portion and a 3' DNA portion; (c) a DNA-dependent DNA polymerase having strand displacement activity; (d) an enzyme that cleaves RNA from an RNA/DNA hybrid; and (e) deoxyribonucleoside triphosphates or suitable analogs (which may or may not be labeled). In embodiments that include transcription, the following are also included in the amplification reaction (either at the same time as the components listed above or added separately): (f) a propromoter polynucleotide comprising a propromoter and a region which hybridizes to an extension product of the first primer (a displaced primer extension product); (g) an RNA polymerase; and (h) ribonucleoside triphosphates or suitable analogs (which may or may not be labeled) in place of (e). The combination is subjected to suitable conditions for composite amplification primer hybridization, extension of the primer, RNA cleavage, and displacement of the first primer extension product wherein its RNA from the in the RNA/DNA heteroduplex of the complex of first and second primer extension products is cleaved and another composite amplification primer binds in the site on the second primer extension product vacated by the cleaved RNA. In embodiments that include transcription, conditions employed are also suitable for hybridization of the propromoter polynucleotide to the displaced first primer extension product, extension of the 3' end of the first primer extension product (if necessary) to generate a double-stranded promoter region, and RNA transcription driven by the promoter.

[0104] In another aspect, the invention provides methods of producing single stranded antisense and sense DNA copies of an RNA sequence of interest using a first composite primer, a second composite primer (termed the "reverse composite" primer), and a target RNA fragment. The method involves the following: (a) formation of a double stranded cDNA comprising a RNA-DNA heteroduplex at each end of the double stranded cDNA; and (b) linear amplification of first strand (sense) cDNA and second strand (antisense) cDNA by primer extension from two composite primers and strand displacement. Single stranded first and second strand cDNA product is produced. This product is useful in, *e.g.*, producing cDNA libraries. As is evident, in this aspect of the invention, the second primer extension product is primed by a composite primer.

IDT Exhibit 1020.2171

Use of enzyme that cleaves single stranded RNA to cleave ribose residues of amplification products or non-specific primer extension products

[0105] In one embodiment, the invention provides a method for removing ribose residues from the 5' ends of amplification products. In methods for amplification of a polynucleotide using a double stranded polynucleotide with a heteroduplex at one end and cleavage of the RNA portion of the heteroduplex to provide a binding site for an amplification primer, for example, as described in U.S. Patent Nos. 6,251,639, 6,692,918, and 6,946,251, the cleavage may be incomplete, resulting in an amplification product comprising one or more ribose residues at the 5' end. An enzyme that cleaves single-stranded RNA may be used to cleave the remaining ribose residues from the 5' end of the amplification products. Treatment of the amplification products with an enzyme capable of cleaving single-stranded RNA, such as RNase I, will advantageously remove the residual ribose residues from the 5' end to generate amplification products with a defined 5' end sequence. This is particularly advantageously for nucleic acid sequencing applications, where the sequencing ladders are better resolved when the 5' ends of the amplification products are well defined, or any other applications which similarly depend on analyzing the amplification products where the size of the amplification products is essential for the analysis.

[0106] In another embodiment, the invention provides methods for degrading the 5' RNA portion of non-specific primer extension products. In methods in which a composite primer is used for generation of a primer extension product from a template polynucleotide, undesirable non-specific hybridization of the primer to a polynucleotide template other than the target polynucleotide can occur. For example, the first primer could hybridize to the first primer extension product. Such undesired hybridization, followed by extension by a polymerase, results in formation of a double-stranded polynucleotide with an RNA/DNA heteroduplex at one end and single-stranded RNA at the other end. The single-stranded RNA end could be copied in the downstream amplification steps to form a double-stranded product with a heteroduplex at both ends. These products will serve as amplification substrates and thus lead to generation of undesired amplification products. Treatment with an enzyme that specifically cleaves single-stranded RNA will cleave the single-stranded RNA portion of the undesired first and second primer extension products and prevent the generation of undesired amplification products in the downstream amplification steps.

IDT Exhibit 1020.2172

Thus, the RNA portion of the nonspecifically hybridized primers can be cleaved with an enzyme that cleaves single stranded RNA, minimizing or eliminating the ability of these undesirable products to interfere with the amplification reaction by generating nonspecific amplification products.

Applications

[0107] The methods of the invention include methods using the amplified products (so-called "applications"). In some embodiments, the invention provides methods of sequencing polynucleotide sequences. For sequencing methods based on methods described herein wherein the amplified product is DNA, the appropriate dNTPs, or analogs thereof, which may be labeled or unlabeled, are used. For sequencing methods based on methods described herein wherein the amplified product is RNA, the appropriate rNTPs, or analogs thereof, which may be labeled or unlabeled, may be used. In some embodiments, a nucleotide terminator, such as a labeled nucleotide terminator, is incorporated during amplification.

[0108] In other embodiments, the invention provides methods of detecting nucleic acid sequence mutations. In one embodiment, the amplified products are used to detect and/or identify single strand conformation polymorphisms in a target polynucleotide.

[0109] The invention provides methods to characterize (for example, detect presence or absence of and/or quantify) a polynucleotide (*e.g.*, RNA) sequence of interest by generating DNA or RNA products using amplification methods of the invention, and analyzing the products by detection/quantification methods such as those based on array technologies or solution phase technologies. These amplified products may be labeled or unlabeled.

[0110] In yet another embodiment, the invention provides methods for immobilizing nucleic acids, including methods for generating microarrays of nucleic acids (DNA or RNA) using amplified products of the amplification methods of the invention.

[0111] In another embodiment, the invention provides methods of generating cDNA libraries, methods of generating subtractive hybridization probes, and methods of generating subtractive hybridization libraries.

[0112] Various methods for the detection and quantification of gene expression levels are known in the art. For example, microarrays, in which either defined cDNAs or oligonucleotides are immobilized at discrete locations on, for example, solid or semi-solid

IDT Exhibit 1020.2173

substrates, or on defined particles, enable the detection and/or quantification of the expression of a multitude of genes in a given specimen.

[0113] Using these previously known methods to detect presence of absence of and/or quantify, and analyzing, multiple mRNA species in a sample, which in turn is used as a measure of gene expression profiling, generally requires direct labeling of cDNA, which requires a large amount of input total RNA, in part because mRNA represents only a small subset of the total RNA pool. Thus, when using total RNA preparations, or isolated mRNA, from a given cell or tissue sample, the analysis of gene expression in the sample using methods such as arrays requires a substantial amount of input RNA, which generally ranges from 50 to 200 μg. Similarly, 2 to 5 μg of mRNA purified from a total RNA preparation would generally be required for a single analysis of gene expression profiling using array technologies. This is a clear limitation of methods such as those based on array technology, insofar as the number of cells, or size of tissue specimen required is very large, and these cells or tissue specimens are often scarcely available for testing or are too precious. This limitation is especially severe in the study of clinical specimens, where the cells to be studied are rare and/or difficult to cultivate *in vitro*, and in high throughput screening of libraries of effector molecules. Also, previous transcription-based methods of amplification of mRNA (described in, for example, Lockhart et al, *Nature Biotechnology* (1996), 14, 1675-1680); van Gelder et al., U.S. Patent No. 5,716,785), are limited to the amplification efficiency of DNA-dependent RNA polymerases and some of these methods require multiple steps. Moreover, the process by which the polymerase promoter sequence is incorporated is prone to result in non-specific amplification.

[0114] The methods of the invention offer the ability to efficiently amplify mRNA under conditions that provide for high specificity of target amplification and which is generally reflective of the distribution in the input RNA. Thus, the utility of the detection/quantification methods which can be used with the amplification products of the invention, such as those based on the powerful array technology, real time PCR, quantitative TaqMan, quantitative PCR using molecular beacons and the like, should be greatly enhanced.

[0115] The linear aspect of the amplification methods of the invention significantly increases the specificity of target amplification. Since generation of multiple copies of a sequence of interest is not dependent nor based on cyclical, exponential, amplification of

IDT Exhibit 1020.2174

amplification products, the specificity of products obtained is greatly enhanced. The distribution of the various species of amplified products is generally reflective of the distribution in the input polynucleotide, *e.g.*, RNA.

[0116] The methods of the invention do not require thermocycling and all of the steps can be performed isothermally, although the various steps may be carried out a different temperatures. This feature provides numerous advantages, including facilitating automation and adaptation for high through-put procedures. The isothermal reaction is faster than that afforded by thermal cycling and is suitable for performing the methods of the invention in miniaturized devices.

[0117] The intermediate double stranded cDNA complex comprising an RNA/DNA heteroduplex provides a substrate for linear amplification. Cleavage of the RNA portion of the RNA/DNA heteroduplex permits further amplification without the need to denature the double stranded cDNA intermediate complex. Moreover, since the cleaved double stranded cDNA complex is mostly double stranded, it is less likely that the secondary structure of a single stranded template will interfere with subsequence amplification.

[0118] Finally, most of the methods of the invention produce products that are single stranded, thus rendering them more accessible to binding to probes, either in a homogeneous manner, *i.e.* in solution, or binding to probes immobilized on solid supports. The double stranded products of the methods of the invention are useful for, *e.g.*, production of cDNA libraries.

[0119] The ability to efficiently amplify polynucleotides, such as mRNA (or any other desired RNA species or population) under conditions that provides for high specificity of target amplification and which will generally not alter the representation of the various polynucleotide, *e.g.*, mRNA, species in the preparation, will greatly enhance the utility of the detection/quantification methods such as those based on the powerful array technology.

IDT Exhibit 1020.2175

Components and reaction conditions used in the methods of the invention

Template nucleic acid

[0120] The nucleic acid (NA) target to be amplified includes 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, yeasts, viruses, viroids, molds, fungi, plants, animals, humans, and fragments thereof. Preferred target polynucleotide includes DNA (*e.g.*, genomic DNA, including human genomic DNA, and mammalian genomic DNA (such as mouse, rat)) and RNA (*e.g.*, mRNA, ribosomal RNA, and total RNA). It should be understood that template RNA includes coding and non-coding RNA. The sequences can be naturally occurring or recombinant nucleic acid targets, including cloned nucleic fragments of interest.

[0121] The target nucleic acid 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. Nucleic acid can be obtained from sources containing very small quantities of nucleic acid, such a single cell, small numbers of cells, patient samples, forensic samples, and archeological samples. Obtaining and purifying nucleic acids use standard techniques in the art, including methods designed to isolate one or a very small number of cells, such a cell sorting or laser capture micro-dissection. The methods of the invention are particularly suited for use with genomic DNA (*e.g.*, human and other mammalian genomic DNA), as well as RNA (*e.g.*, total RNA or mRNA samples). Amplification of an RNA target may be accomplished by initial cDNA synthesis, as known in the art, followed by amplification from the cDNA template.

[0122] The target polynucleotide(s) can be known or unknown and may contain more than one desired specific nucleic acid sequence of interest, each of which may be the same or different from each other. If the target polynucleotide is double stranded (*e.g.*, double stranded DNA or a double stranded DNA/RNA hybrid, such as is produced by first strand cDNA synthesis), the target may first be treated to render it single stranded (*e.g.*, by denaturation or by cleavage of the RNA portion of a DNA/RNA hybrid). Denaturation may also be carried out to remove secondary structure present in a single stranded target

IDT Exhibit 1020.2176

molecule (*e.g.*, RNA). In some cases, double stranded DNA target polynucleotide may be first cleaved by one or more restriction endonuclease enzymes.

[0123] When the target polynucleotide is double stranded (ds) DNA, the initial step of the amplification of a target nucleic acid sequence may be to render the target single stranded. If the target nucleic acid is a double stranded (ds) DNA, the initial step can be target denaturation, for example, by target denaturation. The denaturation step may be thermal denaturation or any other method known in the art, such as alkali treatment. If the target nucleic acid is present in an DNA-RNA hybrid, the initial step can be denaturation of the hybrid to obtain a DNA, or removal of the RNA strand using other means known in the art, such as thermal treatment, digestion with an enzyme that cleaves RNA from an RNA/DNA hybrid (such as RNase H) or alkali treatment, to generate single stranded DNA. When the target is RNA, the initial step may be the synthesis of a single stranded cDNA. Techniques for the synthesis of cDNA from RNA are known in the art, and include reverse transcription of RNA strand using a primer that binds to a specific target, such as the poly-A tail of eukaryotic mRNAs or other specific or consensus sequences. In addition, reverse transcription can be primed by a population of degenerate or partially degenerate primers. First strand cDNA can be separated from the complex of RNA and first strand cDNA as described herein.

[0124] RNAs can be from any source in purified or unpurified form, which can be RNA such as total RNA, tRNA, mRNA, rRNA, mitochondrial RNA, chloroplast RNA, DNA-RNA hybrids, or mixtures thereof, from any source and/or species, including human, animals, plants, and microorganisms such as bacteria, yeasts, viruses, viroids, molds, fungi, plants, and fragments thereof. It is understood that the RNA can be coding or noncoding RNA (such as untranslated small RNAs). RNAs can be obtained and purified using standard techniques in the art. Use of a DNA target (including genomic DNA target) would require initial transcription of the DNA target 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. Thus, RNA template can be itself generated from a DNA source (such as genomic DNA), using methods known in the art, including Kurn, U.S. Patent No. 6,251,639. RNA copies of genomic DNA would generally include untranscribed sequences generally not found in mRNA, such as introns, regulatory and control elements, etc. RNA targets may also be generated from cloned genomic DNA sequences that can be subjected to in vitro transcription. Use of a DNA-

IDT Exhibit 1020.2177

RNA hybrid would require denaturation of the hybrid to obtain a single stranded RNA, denaturation followed by transcription of the DNA strand to obtain an RNA, or other methods known in the art such as digestion with an RNase H to generate single stranded DNA.

Composite primer

[0125] The methods of the invention generally employ two primers each having at least one single-stranded RNA portion, a first primer that hybridizes to template polynucleotide and is extended with a polymerase to form a primer extension product and an a composite amplification primer. In some embodiments, the primers are composite primers that contain RNA and DNA portions, *e.g.*, an RNA portion and a 3' DNA portion. One or both of the composite primers may comprise a 5' RNA portion and a 3' DNA portion. One or both of the composite primers may comprise a 5' RNA portion that is adjacent to the 3' DNA portion.

[0126] In some embodiments, the first primer is a composite primer having partial nucleic acid sequence homology to a multiplicity of genomic DNA sequences, particularly in the 3' sequences of the composite primer, when analyzed using standard nucleic acid comparison algorithms. In some embodiments, the first composite primer contains a random sequence, for example, in the 3' DNA portion. In some embodiments, for example, for amplification of mRNA sequences, the primer is a composite primer having a 3' poly-dT sequence, and optionally containing at least one random nucleotide at the 3' end. In some embodiments, the first primer comprises a multiplicity of different sequences. In some embodiments, the 3' DNA portion of the first primer is a random sequence, for example, a random hexamer, or a longer random sequence, for example, between 5 and 12 nucleotides.

IDT Exhibit 1020.2178

[0127] For development of a first primer having partial nucleic acid sequence homology to a multiplicity of genomic DNA sequences, a composite primer sequence can be used as a query sequence in Blast, to search the human genomic DNA database (or other suitable database, such as a mammalian genomic DNA database). Generally, the search is performed using search parameters suitable for identification of partial or "low stringency" alignments, generally the least stringent conditions provided by the program. Such parameters are known in the art and include use of the NCBI Blast program for searching "short, nearly exact matches," with word size = 7 (conditions permitting as few as 7 consecutive nucleotide perfect matches at any position in the primer sequence). *See, e.g.,* http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?ALIGNMENTS=50&ALIGNMENT_VIEW=Pairwise&AUTO_FORMAT=Semiauto&CLIENT=web&DATABASE=nr&DESCRIPTIONS=100&ENTREZ_QUERY=>(none)&EXPECT=1000&FORMAT_BLOCK_ON_RESPAGE=None&FORMAT_ENTREZ_QUERY=(none)&FORMAT_OBJECT=Alignment&FORMAT_TYPE=HTML&LAYOUT=TwoWindows&NCBI_GI=on&PAGE=Nucleotides&PROGRAM=blastn&SERVICE=plain&SET_DEFAULTS.x=16&SET_DEFAULTS.y=8&SHOW_OVERVIEW=on&WORD_SIZE=7&END_OF_HTTPGET=Yes. First primers useful in the methods of the invention (*i.e.*, that randomly hybridize to template polynucleotide) generally exhibit high partial homology rate with genomic DNA sequences, for example homology of stretches of 7 nucleotides with about 100 genomic DNA sequences, with about 70% of the hits located at the 3' end of the composite primer. A first primer with a very unique sequence (*i.e.*, low levels of homology with target genomic DNA sequences) did not function efficiently in the methods of the invention when used with genomic DNA template.

[0128] As is evident from the discussion above, reference to a primer that binds (hybridizes to) a sequence (or template) encompasses embodiments in which at least a portion of the primer is hybridized, embodiments in which two (or more portions) of the primer are hybridized (separated by unhybridized (looped out) portions of the primer), and embodiments in which the entire primer is hybridized. In certain embodiments, a 5'-portion, commonly the 5'-most portion, of the composite primer is designed such that the particular 5'-portion is not expected to randomly hybridize to template polynucleotide (composite primers of this configuration are referred to as "tailed" primers, in reference to the 'tail' of unhybridized primer). In some embodiments, the tail portion of the composite primer is the entire 5' RNA portion of the composite primer. Thus, according to the

IDT Exhibit 1020.2179

methods of the invention, only a portion of the 3'-end of the composite primer must be hybridized in order for initiation of primer extension by DNA polymerase. In some embodiments, for example, only 1, 2, 3, 4, 5, 6, 7 or more nucleotides of the 3' end of the primer must hybridize in order for primer extension to be initiated. It is understood that hybridization of the 3'-most portion of the composite primer may be stabilized to various extents by further hybridization of another portion of the primer (with or without looping out of intervening primer portions). A DNA polymerase can be included during primer hybridization to enhance (*e.g.*, stabilize) hybridization of composite primer by initiation of primer extension, and thus, stabilization of primer hybridization.

[0129] Random composite primer hybridization and/or generation of composite primer extension product is promoted by use of conditions designed to permit random (non-specific) primer hybridization. Such conditions are well known in the art, and are further discussed below, and include: decreased stringency during primer hybridization and/or first strand synthesis (including reduced temperature and/or buffer conditions of reduced stringency, such as reduced ionic strength); composite primer selection and/or design (discussed further herein); composite primer and template concentration, presence or absence of an agent that stabilizes a 3' hybridized primer (such as a DNA polymerase), and presence or absence of agents such as DMSO that lower the temperature requirements for stable hybridization. It is understood that the selection of reaction conditions may be used to control the frequency of composite primer hybridization, and thus control coverage and/or representation of template polynucleotide sequences in amplification product.

[0130] Generally, the composite primer is also designed so that there is no primer-dimer formation capability, as determined using software routinely available to one of ordinary skill in the art, *e.g.* Oligo Primer Analysis Software from Molecular Biology Insight, and references therein. One of skill in the art will understand that other factors affect nucleic acid hybridization affinities. For example, any and all of the guanosine-cytosine content of the primer-target and primer-primer duplexes, minor groove binders, O-methylation or other modification of nucleotides, temperature, and salt are potentially important factors in constructing primers with the requisite differences in binding energies. Another factor in designing and constructing primers is the free energy parameters of hybridization of given sequences under a given set of hybridization conditions. The free energy parameters for the formation of a given hybrid may be calculated by methods known in the art (see, *e.g.*, Tinoco et al. *Nature* (1973) 246:40-41 and Freier et al., *Proc.*

IDT Exhibit 1020.2180

Natl. Acad. Sci. USA (1986) 83:9373-9377; computer programs, *e.g.*, Oligo Primer Analysis Software from Molecular Biology Insight, and references therein), and it is possible to predict, for a given oligonucleotide template, primer sequences for which the required free energy changes for formation of various complexes will be met.

[0131] The composite primers should be extendable by DNA polymerase, *i.e.*, a DNA-dependent DNA polymerase for extension of the primer along a DNA target polynucleotide, or an RNA-dependent DNA polymerase for extension of the primer along an RNA target polynucleotide.

[0132] Generation of primers suitable for extension by polymerization is well known in the art, such as described in PCT Pub. No. WO99/42618 (and references cited therein). Generally, the primer should permit high efficiency of amplification of a synthetic target that contains a specific primer target binding site (*e.g.*, the complementary sequence to the primer), for example, permitting amplification of about 10^6 to 10^9 using methods described in Kurn, U.S. Patent 6,251,639. The composite amplification primer is designed such that subsequent displacement of the primer extension product by binding of a new (additional) composite primer and the extension of the new primer by the polymerase can be achieved. In addition, cleavage of the RNA portion of the primer extension product leads to generation of amplification product which is not a substrate for amplification by the composite amplification primer. It is understood that, in the following section that generally describes aspects of the composite primers used in the methods of the invention, characteristics described may be applicable to the primers if used for hybridizing and initiating the polynucleotide amplification (production of composite extension product) and/or for single primer amplification as described herein.

[0133] In some embodiments, first and second different composite primers are used in the methods of the invention. In some embodiments, the second composite primer comprises a sequence that is hybridizable to a polynucleotide comprising a complement of the first composite primer.

[0134] For use in single primer amplification and/or composite primer extension product formation, a composite amplification primer comprises at least one RNA portion that is capable of (a) binding (hybridizing) to a sequence on the single stranded portion of the complex (formed by cleavage of RNA in the complex comprising a RNA/DNA partial heteroduplex) (in some embodiments, on second primer extension product) independent of hybridization of the DNA portion(s) to a sequence on the same single stranded portion; and

IDT Exhibit 1020.2181

(b) being cleaved with an agent such as a ribonuclease when hybridized to the single stranded portion. The composite primers bind to the single stranded portion, and are extended by DNA polymerase to form a RNA/DNA partial heteroduplex in which only the RNA portion of the primer is cleaved upon contact with an agent which cleaves RNA in an RNA/DNA hybrid, such as an enzyme, such as a ribonuclease (such as RNase H), while the composite primer extension product remains intact, thus enabling annealing of another composite primer.

[0135] When used for the single primer amplification described herein, the composite primers also comprise a 3' DNA portion that is capable of hybridization to a sequence on the 3' single stranded portion of the complex such that its hybridization to the 3' single stranded portion is favored over that of the nucleic acid strand that is displaced from the complex 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 a preferred embodiment, hybridization of the 3' DNA portion of the composite primer to its complementary sequence in the complex (*e.g.*, in the second primer extension product) favored over the hybridization of the homologous sequence in the 5' end of the displaced strand to the composite primer extension product.

[0136] The composite primer comprises a combination of RNA and DNA (see definition above), with the 3'-end nucleotide being a nucleotide suitable for nucleic acid extension. The 3'-end nucleotide can be any nucleotide or analog that when present in a primer, is extendable by a DNA polymerase. Generally, the 3'-end nucleotide has a 3'-OH. Suitable primers include those that comprise at least one portion of RNA and at least one portion of DNA. For example, composite primers can comprise a 5'-RNA portion and a 3'-DNA portion (in which the RNA portion is adjacent to the 3'-DNA portion); or 5'- and 3'-DNA portions with an intervening RNA portion. Accordingly, in one embodiment, the composite primer comprises a 5' RNA portion and a 3'-DNA portion, preferably wherein the RNA portion is adjacent to the 3'-DNA portion. In another embodiment, the composite primer comprises 5'- and 3'-DNA portions with at least one intervening RNA portion (*i.e.*, an RNA portion between the two DNA portions). In yet another embodiment, the composite primer of the invention comprises a 3'-DNA portion and at least one intervening RNA portion (*i.e.*, an RNA portion between DNA portions).

IDT Exhibit 1020.2182

[0137] The length of an RNA portion in a composite primer comprising a 3'-DNA portion and an RNA portion can be preferably from about 1 to about 50, preferably about 18 to about 30, more preferably from about 3 to about 20, even more preferably from about 4 to about 15, and most preferably from about 5 to about 10 nucleotides. In some embodiments of a composite primer comprising a 3'-DNA portion and an RNA portion, an RNA portion can be at least about any of at least about 1, 3, 4, 5, 10, 15, or 18 nucleotides, with an upper limit of about any of 10, 14, 15, 20, 25, 30, 50 nucleotides. In certain embodiments, the composite primer has an RNA portion of about 5, 10, 15, 20, 25, 30, 35, 40, 45, or 50 nucleotides.

[0138] The length of the 5'-RNA portion in a composite primer comprising a 5'-RNA portion and a 3'-DNA portion can be preferably from about 3 to about 50 nucleotides, preferably about 18 to about 30 nucleotides, more preferably from about 5 to about 20 nucleotides, even more preferably from about 7 to about 18 nucleotides, preferably from about 8 to about 17 nucleotides, and most preferably from about 5 to about 50 nucleotides. In other embodiments of a composite primer comprising a 5'-RNA portion and a 3'-DNA portion, the 5'-RNA portion can be at least about any of 3, 5, 7, 8, 10, 15, or 18 nucleotides, with an upper limit of about any of 14, 15, 17, 18, 20, 30, or 50 nucleotides. In certain embodiments, the composite primer has an RNA portion of about 5, 10, 15, 20, 25, 30, 35, 40, 45, or 50 nucleotides.

[0139] In embodiments of a composite primer comprising a 5'-RNA portion and a 3'-DNA portion further comprising non-5'-RNA portion(s), a non-5'-RNA portion can be preferably from about 1 to about 7 nucleotides, more preferably from about 2 to about 6 nucleotides, and most preferably from about 3 to about 5 nucleotides. In certain embodiments of a composite primer comprising a 5'-RNA portion and a 3'-DNA portion further comprising non-5'-RNA portion(s), a non-5'-RNA portion can be at least about any of 1, 2, 3, 5, with an upper limit of about any of 5, 6, 7, 10 nucleotides.

[0140] In embodiments of a composite primer comprising a 5'-RNA portion and a 3'-DNA portion, in which the 5'-RNA portion is adjacent to the 3'-DNA portion, the length of the 5'-RNA portion can be preferably from about 3 to about 50 nucleotides, about 18 to about 30, more preferably from about 5 to about 20 nucleotides, even more preferably from about 7 to about 18 nucleotides, preferably from about 8 to about 17 nucleotides, and most preferably from about 5 to about 50 nucleotides. In certain embodiments of a composite primer comprising a 5'-RNA portion and a 3'-DNA portion, in which the 5'-RNA portion

IDT Exhibit 1020.2183

is adjacent to the 3'-DNA portion, the 5'-RNA portion can be at least about any of 3, 5, 7, 8, 10, 15, or 18 nucleotides, with an upper limit of about any of 14, 15, 17, 18, 20, 30, or 50 nucleotides. In certain embodiments, the composite primer has an RNA portion of about 5, 10, 15, 20, 25, 30, 35, 40, 45, or 50 nucleotides.

[0141] The length of an intervening RNA portion in a composite primer comprising 5'- and 3'-DNA portions with at least one intervening RNA portion can be preferably from about 1 to about 7 nucleotides, more preferably from about 2 to about 6 nucleotides, and most preferably from about 3 to about 5 nucleotides. In some embodiments of a composite primer comprising 5'- and 3'-DNA portions with at least one intervening RNA portion, an intervening RNA portion can be at least about any of 1, 2, 3, 5 nucleotides, with an upper limit of about any of 5, 6, 7, 10 nucleotides. The length of an intervening RNA portion in a composite primer comprising a 3'-DNA portion and at least one intervening RNA portion can be preferably from about 1 to about 7 nucleotides, more preferably from about 2 to about 6 nucleotides, and most preferably from about 3 to about 5 nucleotides. In some embodiments of a composite primer comprising a 3'-DNA portion and at least one intervening RNA portion, an intervening RNA portion can be at least about any of 1, 2, 3, 5 nucleotides, with an upper limit of about any of 5, 6, 7, 10 nucleotides. In a composite primer comprising a 3'-DNA portion and at least one intervening RNA portion, further comprising a 5'-RNA portion, the 5'-RNA portion can be preferably from about 3 to about 25 nucleotides, more preferably from about 5 to about 20 nucleotides, even more preferably from about 7 to about 18 nucleotides, preferably from about 8 to about 17 nucleotides, and most preferably from about 10 to about 15 nucleotides. In some embodiments of a composite primer comprising a 3'-DNA portion and at least one intervening RNA portion, further comprising a 5'-RNA portion, the 5'-RNA portion can be at least about any of 3, 5, 7, 8, 10 nucleotides, with an upper limit of about any of 15, 17, 18, 20 nucleotides.

[0142] The length of the 3'-DNA portion in a composite primer comprising a 3'-DNA portion and an RNA portion can be preferably from about 1 to about 20, preferably from about 6 to about 12, preferably from about 5 to about 20, more preferably from about 3 to about 18, even more preferably from about 5 to about 15, and most preferably from about 7 to about 12 nucleotides. In some embodiments of a composite primer comprising a 3'-DNA portion and an RNA portion, the 3'-DNA portion can be at least about any of 1, 3, 5, 6, 7, or 10 nucleotides, with an upper limit of about any of 10, 12, 15, 18, 20, 22

IDT Exhibit 1020.2184

nucleotides. In one embodiment, the composite primer has a 3'-DNA portion of 1 nucleotide, or any of about 5, 6, 10, 12, 15, or 20 nucleotides.

[0143] The length of the 3'-DNA portion in a composite primer comprising a 5'-RNA portion and a 3'-DNA portion can be preferably from about 1 to about 20 nucleotides, preferably from about 6 to about 12, preferably from about 5 to about 20, more preferably from about 3 to about 18 nucleotides, even more preferably from about 5 to about 15 nucleotides, and most preferably from about 7 to about 12 nucleotides. In some embodiments of a composite primer comprising a 5'-RNA portion and a 3'-DNA portion, the 3' DNA portion can be at least about any of 1, 3, 5, 6, 7, 10 nucleotides, with an upper limit of about any of 10, 12, 15, 18, 20, 22 nucleotides. In one embodiment, the composite primer has a 3'-DNA portion of 1 nucleotide, or any of about 5, 6, 10, 12, 15, or 20 nucleotides.

[0144] In embodiments of a composite primer comprising a 5'-RNA portion and a 3'-DNA portion, further comprising non-3'-DNA portion(s), a non-3'-DNA portion can be preferably from about 1 to about 10 nucleotides, more preferably from about 2 to about 8 nucleotides, and most preferably from about 3 to about 6 nucleotides. In some embodiments of a composite primer comprising a 5'-RNA portion and a 3'-DNA portion, further comprising non-3'-DNA portion(s), a non-3'-DNA portion can be at least about any of 1, 2, 3, 5 nucleotides, with an upper limit of about any of 6, 8, 10, 12 nucleotides.

[0145] In embodiments of a composite primer comprising a 5'-RNA portion and a 3'-DNA portion in which the 5'-RNA portion is adjacent to the 3'-DNA portion, the length of the 3'-DNA portion can be preferably from about 1 to about 20 nucleotides, preferably from about 6 to about 12, preferably from about 5 to about 20 nucleotides, more preferably from about 3 to about 18 nucleotides, even more preferably from about 5 to about 15 nucleotides, and most preferably from about 7 to about 12 nucleotides. In certain embodiments of the primer comprising a 5'-RNA portion and a 3'-DNA portion in which the 5'-RNA portion is adjacent to the 3'-DNA portion, the 3'-DNA portion can be at least about any of 1, 3, 5, 6, 7, 10 nucleotides, with an upper limit of about any of 10, 12, 15, 18, 20, 22 nucleotides. In one embodiment, the composite primer has a 3'-DNA portion of about 1 nucleotide, or any of about 5, 6, 10, 12, 15, or 20 nucleotides.

[0146] The length of a non-3'-DNA portion in a composite primer comprising 5'- and 3'-DNA portions with at least one intervening RNA portion can be preferably from about 1 to about 10 nucleotides, more preferably from about 2 to about 8 nucleotides, and most

IDT Exhibit 1020.2185

preferably from about 3 to about 6 nucleotides. In some embodiments of a primer comprising 5'- and 3'-DNA portions with at least one intervening RNA portion, a non-3'-DNA portion can be at least about any of 1, 2, 3, 5 nucleotides, with an upper limit of about any of 6, 8, 10, 12 nucleotides.

[0147] The length of the 3'-DNA portion in a composite primer comprising 5'- and 3'-DNA portions with at least one intervening RNA portion can be preferably from about 1 to about 20 nucleotides, more preferably from about 3 to about 18 nucleotides, even more preferably from about 5 to about 15 nucleotides, and most preferably from about 7 to about 12 nucleotides. In some embodiments of a composite primer comprising 5'- and 3'-DNA portions with at least one intervening RNA portion, the 3'-DNA portion can be at least about any of 1, 3, 5, 7, 10 nucleotides, with an upper limit of about any of 10, 12, 15, 18, 20, 22 nucleotides. In one embodiment, the composite primer has a 3'-DNA portion of about 7 nucleotides.

[0148] The length of a non-3'-DNA portion (*i.e.*, any DNA portion other than the 3'-DNA portion) in a composite primer comprising a 3'-DNA portion and at least one intervening RNA portion can be preferably from about 1 to about 10 nucleotides, more preferably from about 2 to about 8 nucleotides, and most preferably from about 3 to about 6 nucleotides. In some embodiments of a composite primer comprising a 3'-DNA portion and at least one intervening RNA portion, a non-3'-DNA portion can be at least about any of 1, 3, 5, 7, 10 nucleotides, with an upper limit of about any of 6, 8, 10, 12 nucleotides. The length of the 3'-DNA portion in a composite primer comprising a 3'-DNA portion and at least one intervening RNA portion can be preferably from about 1 to about 20 nucleotides, more preferably from about 3 to about 18 nucleotides, even more preferably from about 5 to about 15 nucleotides, and most preferably from about 7 to about 12 nucleotides. In some embodiments of a composite primer comprising a 3'-DNA portion and at least one intervening RNA portion, the 3'-DNA portion can be at least about any of 1, 3, 5, 7, 10 nucleotides, with an upper limit of about any of 10, 12, 15, 18, 20, 22 nucleotides. In one embodiment, the composite primer has a 3'-DNA portion of about 7 nucleotides. It is understood that the lengths for the various portions can be greater or less, as appropriate under the reaction conditions of the methods of this invention.

[0149] In some embodiments, the 5'-DNA portion of a composite primer includes the 5'-most nucleotide of the primer. In some embodiments, the 5'-RNA portion of a composite primer includes the 5' most nucleotide of the primer. In other embodiments, the

IDT Exhibit 1020.2186

3'-DNA portion of a composite primer includes the 3' most nucleotide of the primer. In other embodiments, the 3'-DNA portion is adjacent to the 5'-RNA portion and includes the 3' most nucleotide of the primer (and the 5'-RNA portion includes the 5' most nucleotide of the primer).The total length of the composite primer can be preferably from about 10 to about 50 nucleotides, more preferably from about 15 to about 30 nucleotides, and most preferably from about 20 to about 25 nucleotides. In some embodiments, the length can be at least about any of 10, 15, 20, 25 nucleotides, with an upper limit of about any of 25, 30, 50, 60 nucleotides. In certain embodiments, the composite primer is about 21 or about 27 nucleotides in length. It is understood that the length can be greater or less, as appropriate under the reaction conditions of the methods of this invention.

[0150] As described herein, one or more different composite primers may be used in an amplification reaction.

DNA polymerase, and an agent capable of cleaving an RNA-DNA hybrid

[0151] The amplification methods of the invention often employ the following enzymes: an RNA-dependent DNA polymerase, a DNA-dependent DNA polymerase, and an agent capable of cleaving an RNA strand of an RNA-DNA hybrid (for example, a ribonuclease such as RNase H). One or more of these activities may be found and used in a single enzyme. For example, RNase H activity may be supplied by an RNA-dependent DNA polymerase (such as reverse transcriptase) or may be provided in a separate enzyme. Reverse transcriptases useful for this method may or may not have RNase H activity. Many reverse transcriptases, such as those from avian myeloblastosis 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 reverse transcriptase which lacks the RNase H activity. Reverse transcriptase 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. In these cases, the addition of an RNase H from other sources, such as that isolated from *E. coli*, can be employed for the formation of the double stranded cDNA. The RNA-dependent DNA polymerase activity and DNA-dependent DNA polymerase activity may

IDT Exhibit 1020.2187

be provided by the same enzyme (for example, Bst polymerase), or these activities may be provided in separate enzymes.

[0152] Often, amplification methods of the invention involve formation of a complex comprising an RNA/DNA partial heteroduplex. This process generally utilizes the enzymatic activities of an RNA-dependent DNA polymerase and a DNA-dependent DNA polymerase. Generally, RNA in the RNA/DNA partial heteroduplex is cleaved by an agent (such as an enzyme, such as a ribonuclease) capable of cleaving RNA from an RNA/DNA hybrid, generating a 3' single stranded portion with sequences that are complementary to RNA in a composite primer (and thus forming a binding site for a composite primer).

[0153] RNA-dependent DNA polymerases for use in the methods and compositions of the invention are capable of effecting extension of a primer according to the methods of the invention. Accordingly, a preferred RNA-dependent DNA polymerase is 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 and compositions of the invention include reverse transcriptase and, for example, a DNA polymerase that possesses both DNA-dependent and RNA-dependent DNA polymerase activity, such as Bst DNA polymerase.

[0154] DNA-dependent DNA polymerases for use in the methods and compositions of the invention are capable of effecting extension of the composite primer according to the methods of the invention. Accordingly, a preferred polymerase is one that is capable of extending a nucleic acid primer along a nucleic acid template that is comprised at least predominantly of deoxyribonucleotides. The formation of the complex comprising the RNA/DNA partial heteroduplex can be carried out by a DNA polymerase which comprises both RNA-dependent DNA polymerase and DNA-dependent DNA polymerase activities (such as Bst DNA polymerase, or a reverse transcriptase). Amplification of an RNA sequence according to methods of the invention involves the use of a DNA polymerase that is able to displace a nucleic acid strand from the polynucleotide to which the displaced strand is bound, and, generally, the more strand displacement capability the polymerase exhibits (*i.e.*, compared to other polymerases which do not have as much strand displacement capability) is preferable. Preferably, the DNA polymerase has high affinity for binding at the 3'-end of an oligonucleotide hybridized to a nucleic acid strand. Preferably, the DNA polymerase does not possess substantial nicking activity. Generally, the DNA polymerase preferably has little or no 5'->3' exonuclease activity so as to

IDT Exhibit 1020.2188

minimize degradation of primer, or primer extension polynucleotides. Generally, this exonuclease activity is dependent on factors such as pH, salt concentration, whether the template is double stranded or single stranded, and so forth, all of which are familiar to one skilled in the art. Mutant DNA polymerases in which the 5'->3' exonuclease activity has been deleted, are known in the art and are suitable for the amplification methods described herein. Mutant DNA polymerases which lack both 5' to 3' nuclease and 3' to 5' nuclease activities have also been described , for example, exo$^{-/-}$Klenow DNA polymerase. It is preferred that the DNA polymerase displaces primer extension products from the template nucleic acid in at least about 25%, more preferably at least about 50%, even more preferably at least about 75%, and most preferably at least about 90%, of the incidence of contact between the polymerase and the 5' end of the primer extension product. In some embodiments, the use of thermostable DNA polymerases with strand displacement activity is preferred. Such polymerases are known in the art, such as described in U.S. Pat. No. 5744312 (and references cited therein). Preferably, the DNA polymerase has little to no proofreading activity.

[0155] Suitable DNA polymerases for use in the methods and compositions of the invention include those disclosed in U.S. Pat. Nos. 5648211 and 5744312, which include exo$^-$ Vent (New England Biolabs), exo$^-$ Deep Vent (New England Biolabs), Bst (BioRad), exo$^-$ Pfu (Stratagene), Bca (Panvera), sequencing grade Taq (Promega), exo$^{-/-}$Klenow DNA polymerase, and thermostable DNA polymerases from thermoanaerobacter thermohydrosulfuricus.

[0156] In some embodiments, the agent that is capable of cleaving RNA in an RNA/DNA hybrid for use in the methods and compositions of the invention is a ribonuclease that is capable of cleaving ribonucleotides in an RNA/DNA hybrid but is not capable of cleaving single-stranded RNA. Preferably, the ribonuclease cleaves ribonucleotides in an RNA/DNA hybrid regardless of the identity and type of nucleotides adjacent to the ribonucleotide to be cleaved. It is preferred that the ribonuclease cleaves independent of sequence identity. Examples of suitable ribonucleases for the methods and compositions of the invention are well known in the art, including ribonuclease H (RNase H), *e.g.*, Hybridase.

[0157] As is well known in the art, DNA-dependent DNA polymerase activity, RNA-dependent DNA polymerase activity, and the ability to cleave RNA from a RNA/DNA hybrid may be present in different enzymes, or two or more activities may be present in the

IDT Exhibit 1020.2189

same enzyme. Accordingly, in some embodiments, the same enzyme comprises RNA-dependent DNA polymerase activity and cleaves RNA from an RNA/DNA hybrid. In some embodiments, the same enzyme comprises DNA-dependent DNA polymerase activity and cleaves RNA from an RNA/DNA hybrid. In some embodiments, the same enzyme comprises DNA-dependent DNA polymerase activity, RNA-dependent DNA polymerase activity and cleaves RNA from an RNA/DNA hybrid. In some embodiments, different enzymes comprise RNA-dependent DNA polymerase activity and DNA-dependent DNA polymerase activity. In some embodiments, different enzymes comprise RNA -dependent DNA polymerase activity and cleave RNA from an RNA/DNA hybrid. In some embodiments, different enzymes comprise DNA-dependent DNA polymerase activity and cleave RNA from an RNA/DNA hybrid.

Enzyme capable of cleaving single-stranded RNA

[0158] Methods of the invention employ an enzyme capable of cleaving single-stranded RNA. The enzyme may be used to cleave RNA in a first primer prior to addition of a second primer containing single-stranded RNA to a reaction mixture and/or to cleave single-stranded RNA from amplification products as described herein. In embodiments in which the enzyme is used to cleave single stranded RNA of a first composite primer, the enzyme capable of cleaving single-stranded RNA is generally added in an amount that is sufficient to degrade all or substantially all of the single-stranded RNA in the first RNA-containing primer.

[0159] RNase enzymes have been previously described in procedures for detecting mutations (Grange et al. (1990) *Nucleic Acids Res.* 18(14):4227-36; Goldrick (2001) *Hum Mutat.* 18(3):190-204; Murthy et al. (1995) *DNA Cell Biol.* 14(1):87-94; U.S. Pat. No. 4,946,773), degradation of excess mRNA to enrich cDNA/mRNA hybrid after first strand cDNA synthesis (Carninci et al. (1999) *Methods Enzymol.* 303:19; PCT Application No. WO 98/20122), and for prevention of cross contamination in a PCR amplification reaction (Walder et al. (1993) *Nucleic Acids Res.* 21(18):4339-43, for the preparation of full length cDNA libraries (Carninci et al. (2000) *Genome Res.* 10(10):1617-30.

[0160] RNase I works well in the methods of the invention as described herein, since it has little or no sequence specificity, and is known to cleave the bond between any combination of ribo-nucleotides in a single-stranded RNA. RNase T may also be used although it is known to exhibit sequence specificity. Combinations of RNase enzymes may

IDT Exhibit 1020.2190

also be used. An important aspect in choosing a single-stranded RNA specific RNase enzyme suitable for use in the methods of the invention is the ability to easily inactivate the RNase activity prior to addition of a second RNA-containing primer, so as not to compromise the integrity of the RNA in the second primer, for example a DNA/RNA amplification primer. RNase I is easily inactivated at elevated temperature. In addition, RNase I is known to be insensitive to ion composition of the reaction mixture, and thus its use in the methods of the invention generally do not require buffer modification or exchange (*i.e.*, to change to a buffer composition suitable for the cleavage of the RNA portion of the first strand RNA-containing primer).

[0161] RNase I may be inactivated by elevation of the temperature of the reaction mixture, for example, to 70^0C to 90^0C for 5 to 20 minutes.

[0162] RNase I may be used at a concentration of about 0.01 to about 0.11 units per µl. In some embodiments, RNase I is used at a concentration of any of about 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.10, or 0.11 units per µl. The reaction is usually carried out at 37°C for about 30 minutes. Reaction buffer commonly comprises 10mM Tris-HCl, pH 7.5, 100mM NaCl. The enzyme is active at a wide range of concentration of divalent ions such as MgCl in the range of 0 to 6 mM or higher.

Reaction conditions and detection

[0163] Appropriate reaction media and conditions for carrying out the methods of the invention are those that permit nucleic acid 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^{2+} or Mn^{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.

IDT Exhibit 1020.2191

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 or acetylated BSA, single strand binding proteins (for *e.g.*, 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. Preferably, the amplification reactions (particularly, primer extension other than the composite and second primer extension product synthesis steps, and strand displacement) are performed isothermally, which avoids the cumbersome thermocycling process. The amplification reaction is carried out at a temperature that permits hybridization of the oligonucleotides (primer) of the invention to the template polynucleotide and primer extension products, and that does not substantially inhibit the activity of the enzymes employed. The temperature can be in the range of 0° C to about 85° C, about 25° C to about 85° C, about 30°C to about 80°C, and about 37°C to about 75°C.

[0164] Random priming and/or primer extension and/or isothermal amplification can be conducted under conditions of reduced stringency (*i.e.*, permitting hybridization of sequences that are not fully complementary). For a given set of reaction conditions, the ability of two nucleotide sequences to hybridize with each other is based on the degree of complementarity of the two nucleotide sequences, which in turn is based on the fraction of matched complementary nucleotide pairs. The more nucleotides in a given sequence that are complementary to another sequence, the more stringent the conditions can be for hybridization and the more specific will be the binding of the two sequences. Conversely, the lower the stringency of the conditions for hybridization, the lower the complementarity necessary for binding between the hybridizing and/or partially hybridizing composite primer and template polynucleotide. Decreased stringency is achieved by any one or more of the following: reducing the temperature, decreasing the ratio of co-solvents, lowering the salt concentration, and the like. Conditions that increase or reduce the stringency of a

IDT Exhibit 1020.2192

hybridization reaction are widely known and published in the art. See, for example, Sambrook et al. (1989), and in Ausubel (1987), supra. Useful hybridization conditions are also provided in, *e.g.*, Tijessen, 1993, Hybridization With Nucleic Acid Probes, Elsevier Science Publishers B.V. and Kricka, 1992, Nonisotopic DNA Probe Techniques, Academic Press San Diego, Calif. The hybridization conditions chosen depend on a variety of factors known in the art, for example the length and type (*e.g.*, RNA, DNA, PNA) of primer and primer binding region of the oligonucleotide template, as well as the concentration of primer and template polynucleotides.

[0165] Insofar as it is convenient to use buffer conditions that are compatible with DNA polymerase activity and/or ribonuclease activity, stringency of hybridization of composite primers can be controlled by altering temperature of the reaction. Examples of relevant conditions include (in order of increasing stringency): incubation temperatures of approximately 15°C, 20°C, 25°C, 30°C, 37°C, 40°C, 45°C, 50°C, 60°C, or more. Accordingly, in some embodiments, composite primer random hybridization occurs at a reduced temperature, for example at 25°C - 37°C, followed at incubation at increased temperature(s) suitable for the amplification phase of the methods (such as about 50°C). In some embodiments, temperature is increased at 5°C increments. In other embodiments, temperature is shifted from low to high temperature.

[0166] Nucleotide and/or nucleotide analogs, such as deoxyribonucleoside triphosphates, that can be employed for synthesis of the primer extension products in the methods of the invention are provided in the amount of from preferably about 50 to about 2500 μM, more preferably about 100 to about 2000 μM, even more preferably about 200 to about 1700 μM, and most preferably about 250 to about 1500 μM. In some embodiments, a nucleotide or nucleotide analog whose presence in the primer extension strand enhances displacement of the strand (for example, by causing base pairing that is weaker than conventional AT, CG base pairing) is included. Such nucleotide or nucleotide analogs include deoxyinosine and other modified bases, all of which are known in the art.

[0167] The oligonucleotide components of the amplification reactions of the invention are generally in excess of the number of target nucleic acid sequence to be amplified. They can be provided at about or at least about any of the following: 10, 10^2, 10^4, 10^6, 10^8, 10^{10}, 10^{12} times the amount of target nucleic acid. Composite primers can each be provided at about or at least about any of the following concentrations: 50 nM, 100 nM, 500 nM, 1 uM,

IDT Exhibit 1020.2193

2.5 uM , 5uM, 10 uM. Composite primer concentration also impacts frequency and/or position of composite primer hybridization. Generally, increased primer concentrations increased frequency of primer hybridization. Auxiliary primers can be provided at about or at least about any of the following concentrations: about 25 nM, about 50nM, about 100 nM, about 500 nM, about 1 uM, about 2.5 uM, about 5 uM, about 10 uM, or more.

[0168] In one embodiment, the foregoing components are added simultaneously at the initiation of the amplification process. In another embodiment, components are added in any order prior to or after appropriate timepoints during the amplification process, as required and/or permitted by the amplification reaction. Such timepoints, some of which are noted below, can be readily identified by a person of skill in the art. The enzymes used for nucleic acid amplification according to the methods of the invention can be added to the reaction mixture either prior to the target nucleic acid denaturation step, following the denaturation step, or following hybridization of the primer to the target polynucleotide, as determined by their thermal stability and/or other considerations known to the person of skill in the art. In these embodiments, the reaction conditions and components may be varied between the different reactions.

[0169] The amplification process can be stopped at various timepoints, and resumed at a later time. Said timepoints can be readily identified by a person of skill in the art. One timepoint is at the end of random composite primer hybridization. Another timepoint is at the end of random composite primer hybridization and composite primer extension product synthesis. Another timepoint (in some embodiments) is following cleavage of template RNA. Another timepoint is following cleavage of excess first RNA-containing primer with an enzyme capable of cleaving single-stranded RNA, such as RNase I. Another timepoint is immediately prior to initiation of single primer amplification (which in some embodiments, may be initiated by addition of an enzyme (such as RNase H) that cleaves RNA from RNA/DNA heteroduplex, and optionally, DNA polymerase). Another timepoint is at the end of second primer extension product synthesis. Methods for stopping the reactions are known in the art, including, for example, cooling the reaction mixture to a temperature that inhibits enzyme activity or heating the reaction mixture to a temperature that destroys an enzyme. Methods for resuming the reactions are also known in the art, including, for example, raising the temperature of the reaction mixture to a temperature that permits enzyme activity, replenishing a destroyed (depleted) enzyme, or adding reagent(s) necessary for initiation of a step (for example, addition of RNase H and/or DNA

polymerase to initiate the single primer amplification phase of the methods). In some embodiments, one or more of the components of the reactions is replenished prior to, at, or following the resumption of the reactions. For example, it may be necessary to replenish the composite primer prior to beginning the single primer amplification reaction if the same composite primer is being used. Alternatively, the reaction can be allowed to proceed (*i.e.*, from start to finish) without interruption.

[0170] The reaction can be allowed to proceed without purification of intermediate complexes, for example, to remove primer. Products can be purified at various timepoints, which can be readily identified by a person of skill in the art. One timepoint is at the end of formation of the complex comprising an RNA/DNA partial heteroduplex. Another timepoint is at the end of composite primer hybridization.

[0171] The detection of the amplification product is indicative of the presence of the target sequence. Quantitative analysis is also feasible. Direct and indirect detection methods (including quantitation) are well known in the art. For example, by comparing the amount of product amplified from a test sample containing an unknown amount of a polynucleotide containing a target sequence to the product of amplification of a reference sample that has a known quantity of a polynucleotide that contains the target sequence, the amount of target sequence in the test sample can be determined.

Compositions and kits of the invention

[0172] The invention also provides compositions and kits used in the methods described herein. The compositions may be any component(s), reaction mixture and/or intermediate described herein, as well as any combination.

[0173] In one embodiment, the invention provides an enzyme capable of cleaving single-stranded RNA and one or more of: a composite primer comprising an RNA portion and a 3'-DNA portion; a DNA-dependent and/or RNA-dependent polymerase; an enzyme that cleaves RNA from an RNA/DNA duplex; and auxiliary primers (for example, a population of random hexamer primers).

[0174] In some embodiments, the composition comprises a composite primer comprising 5'- and 3'-DNA portions with at least one intervening RNA portion. In other examples, the invention provides a composition comprising a composite primer as

IDT Exhibit 1020.2195

described herein, and auxiliary primers (for example, a population of randomized hexamer primers). In other examples, the composition comprises a composite primer that is derivatized by attachment of a moiety capable of effecting attachment of a polynucleotide comprising the composite primer to a solid substrate used in preparing nucleic acid microarrays. In some embodiments, the composite primer is further by attachment of a positively charged moiety such as an amine. In other embodiments, the composite primer is labeled, for example by derivatizing the composite primer with a detectable moiety, such as a label, or a moiety that can be covalently or non-covalently attached to a label. Labeled composite primers are further described herein.

[0175] In some embodiments, the composition further comprises a labeled dNTP, labeled rNTP and/or a nucleotide terminator, *e.g.*, a labeled nucleotide terminator.

[0176] In still other embodiments, the composition comprises a non-canonical nucleotide (such as dUTP), and reagents suitable for labeling and/or fragmenting abasic sites, as described in co-pending U.S. Patent Application Publication Nos. 2004/0005614 and 2005/0208538, and U.S. provisional patent application o. 60/817,890.. In methods of the invention in which amplification is performed in the presence of one or more non-canonical nucleotides, the amplified products may be contacted with an enzyme such as a glycosylase to cleave the base portion of the non-canonical nucleotide to generate an abasic site. The polynucleotides containing abasic sites may be contacted with an agent that cleaves polynucleotides at abasic sites to generate polynucleotide fragments. The polynucleotide fragments may be labeled at or near the abasic site.

[0177] In some embodiments, the compositions further comprise amplification products produced as described herein. Accordingly, the invention provides compositions comprising a population of polynucleotides which are copies or the complement of a target sequence, which are produced by any of the methods described herein (or compositions comprising the products). The invention also includes compositions and various configurations (such as arrays) of these populations, which may be homogeneous (same sequence) or heterogeneous (different sequence). These populations may be any assembly of sequences obtained from the methods described herein.

[0178] The compositions are generally in lyophilized or aqueous form, preferably in a suitable buffer.

IDT Exhibit 1020.2196

[0179] The compositions are generally in a suitable medium, although they can be in lyophilized form. Suitable media include, but are not limited to, aqueous media (such as pure water or buffers).

[0180] The invention provides kits for carrying out the methods of the invention. Accordingly, a variety of kits are provided in suitable packaging. The kits may be used for any one or more of the uses described herein, and, accordingly, may contain instructions for any one or more of the following uses: methods of amplification; genotyping, nucleic acid mutation detection (including methods of genotyping), determining the presence or absence of a sequence of interest, quantitating a sequence of interest, preparation of an immobilized nucleic acid (which can be a nucleic acid immobilized on a microarray), comparative genomic hybridization, and characterizing nucleic acids using the amplified nucleic acid products generated by the methods of the invention, methods of expression profiling, subtractive hybridization and the preparation of probes for subtractive hybridization, and methods of preparing libraries (which may be cDNA and/or differential hybridization libraries).

[0181] The kits of the invention comprise one or more containers comprising an enzyme capable of cleaving single-stranded RNA, such as RNase I, RNase T, RNase A, or a combination thereof) and any combination of the components described herein. A kit may comprise any of the composite primers described herein. In some embodiments, a kit further comprises auxiliary primers, which may or may not be separately packaged. The composite primer may be labeled or unlabeled. Kits may also optionally further include any of one or more of the enzymes described herein (for example, DNA-dependent DNA polymerase, RNA-dependent DNA polymerase, a DNA polymerase that provides both DNA-dependent and RNA-dependent DNA polymerase activities, an enzyme capable of cleaving RNA from an RNA/DNA hybrid, such as RNase H, as well as deoxynucleoside triphosphates (labeled or unlabeled or derivatized). Kits may also include one or more suitable buffers (for example, as described herein). Kits may also include a labeled dNTP(s) and/or a non-canonical nucleotide (such as dUTP), as described in Kurn et al, co-pending U.S. patent application No. 2004-0005614.

[0182] One or more reagents in the kit can be provided as a dry powder, usually lyophilized, including excipients, which on dissolution will provide for a reagent solution having the appropriate concentrations for performing any of the methods described herein.

IDT Exhibit 1020.2197

Each component can be packaged in separate containers or some components can be combined in one container where cross-reactivity and shelf life permit.

[0183] The kits of the invention may optionally include a set of instructions, generally written instructions, although electronic storage media (*e.g.*, magnetic diskette or optical disk) and/or reference to a website containing instructions are also acceptable, relating to the use of components of the methods of the invention for the intended nucleic acid amplification, and/or, as appropriate, for using the amplification products for purposes such as detection of sequence mutation. The instructions included with the kit generally include information as to reagents (whether included or not in the kit) necessary for practicing the methods of the invention, instructions on how to use the kit, and/or appropriate reaction conditions.

[0184] In another example, any of these kits further comprises one or more controls (which can be, for example, template polynucleotide (*e.g.*, DNA template such as genomic DNA or RNA template such as total RNA or mRNA), composite primers, and/or auxiliary primer(s).

[0185] The component(s) of the kit may be packaged in any convenient, appropriate packaging. The components may be packaged separately, or in one or multiple combinations.

[0186] The relative amounts of the various components in the kits can be varied widely to provide for concentrations of the reagents that substantially optimize the reactions that need to occur to practice the methods disclosed herein and/or to further optimize the sensitivity of any assay.

[0187] The invention also provides systems for effecting the methods described herein. These systems comprise an enzyme capable of cleaving single-stranded RNA and various combinations of the components discussed above.

[0188] Any of the systems embodiments may also comprise a template (target) sequence, as described herein. A system generally includes one or more apparatuses for performing the amplification methods of the invention. Such apparatuses include, for example, heating devices (such as heating blocks or water baths) and apparatuses which effect automation of one or more steps of the methods described herein. The methods of the invention are particularly suitable for use with miniaturized devices, as thermal cycling is not required for any of the steps. A non-limiting example of suitable devices includes the BioAnalyzer (Agilent and Caliper) and the eSensor.

IDT Exhibit 1020.2198

[0189] The invention also provides reaction mixtures (or compositions comprising reaction mixtures) which contain an enzyme capable of cleaving single-stranded RNA and various combinations of components described herein. Examples of reaction mixtures have been described. In some embodiments, the invention provides reaction mixtures comprising (a) a target polynucleotide; (b) a composite primer comprising a 3' DNA portion and an RNA portion; (c) a polymerase; and (d) an enzyme capable of cleaving single-stranded RNA, such as RNase I. As described herein, any of the composite primers may be in the reaction mixture (or a plurality of composite primers), including a composite primer that comprises a 5' RNA portion which is adjacent to the 3' DNA portion. The reaction mixture could also further comprise an enzyme which cleaves RNA from an RNA/DNA hybrid, such as RNase H.

[0190] Other reaction mixtures are described herein and are encompassed by the invention.

[0191] The invention also includes compositions comprising any of the complexes (which are intermediates in the methods described herein) described herein. Examples of such complexes are schematically depicted in Figures 1-3. As an example, one complex of the invention is a complex comprising: (a) a target polynucleotide strand; and (b) a composite primer, said composite primer comprising a 3' DNA portion and an RNA portion. The composite primer may have an RNA portion which is 5' and adjacent to the 3' DNA portion. As another example, a complex of the invention is a complex comprising: (a) a composite primer extension product; and (b) a target polynucleotide.

[0192] In yet another example, a complex of the invention is a complex comprising a RNA/DNA partial heteroduplex, prepared by any of the methods described herein. In some embodiments, the complex further comprises a second RNA/DNA partial heteroduplex at a second end. In yet another example, the complex of the invention is a complex comprising a 3' single stranded DNA portion produced by any of the methods described herein. In some embodiments, the complex further comprises a second 3' single stranded region. In another example, the complex of the invention is (a) a complex comprising a 3' single stranded DNA portion, and (b) a composite primer hybridized to the 3' single stranded portion.

[0193] The following Example is intended to illustrate, but not limit, the invention.

IDT Exhibit 1020.2199

EXAMPLE

[0194] An RNA target is amplified using a method involving two chimeric RNA/DNA primers with degradation of the RNA portion of the first chimeric primer prior to addition of the second chimeric primer for improved amplification efficiency as described below.

First strand cDNA synthesis

[0195] First stand cDNA synthesis is initiated randomly across the length of RNA transcripts using chimeric primers having a 3'-DNA portion that includes a random, for example, a random hexamer sequence, and a 5'-RNA portion that includes a sequence that is not complementary to the target RNA. 1 to 2μl of total RNA sample (0.5 to 50ng total RNA purified from a sample to be analyzed), is mixed with 1 to 2μl chimeric first strand cDNA primer mixture (5 to 40μM final concentration), and the mixture was transferred to a heating block or a thermal cycler, and incubated at 50°C to 65°C for 5 minutes, for denaturation. The mixture is removed from the heating block or thermocycler and placed on ice.

[0196] A mixture of reverse transcriptase and a buffer containing Mg, salt, RNase inhibitor (Rnasin or any other commercially available RNase inhibitor commonly used for cDNA synthesis) and all four dNTPs, in volume of up to 10μl, is added to the chimeric primer mixture and the reaction is incubated at 25^0C (range of 16 to 30^0C) for 10 minutes (range of 5 to 20 minutes), followed by incubation at 42^0C (range 37^0C to 48^0C) for 15 minutes (range 5 to 60 minutes), and a short incubation at 65°C for 2 to 5 minutes for inactivation of the reverse transcriptase and nicking of the RNA in the newly synthesized RNA-cDNA heteroduplex, as previously described (see, for example, U.S. Patent Application No. 2003/0087251). The reaction mixture is placed on ice.

Second strand synthesis

[0197] A mixture of DNA polymerase (Klenow exo-, Bst large fragment, BCA or any other polymerase suitable for second strand cDNA synthesis) in a suitable buffer containing a suitable ion composition for activity of the polymerase used, RNase inhibitor (Rnasin or any other commercially available RNase inhibitor commonly used for cDNA synthesis), and all four dNTPs is added to the first strand synthesis reaction mixture (to a total volume

IDT Exhibit 1020.2200

of 10 to 20µl). The reaction mixture is incubated at 25^{0}C for 5 to 10 minutes followed by incubation at a temperature appropriate for the DNA polymerase used (37^{0}C to 50^{0}C) for 30 to 60 minutes. At the end of this second strand cDNA synthesis step, the reaction mixture is incubated at 75^{0}C (65 to 85^{0}C) for 5 to 15 minutes for inactivation of the polymerase and RNase inhibitor (Rnasin). The inactivation of the RNase inhibitor prior to progressing to the next step is particularly important when using an RNase which is known to be inhibited by the inhibitor used in this step of the reaction.

Degradation of the RNA portion of the first strand cDNA chimeric primers by treatment with single-strand specific RNase

[0198] The degradation of the 5'-RNA portion of the chimeric first strand cDNA primers which have not been incorporated into the double stranded cDNA to form an RNA/DNA heteroduplex at one end of the double-stranded cDNA products can be carried out with an enzyme that is capable of degrading single-stranded RNA, for example, RNase I.

[0199] Cleavage of single-stranded RNA is initiated by the addition of RNase I to the second strand cDNA reaction mixture (following the heat inactivation step at the end of the second strand cDNA synthesis reaction) and incubation of the reaction mixture at 37^{0}C for 15 minutes (10 to 30 minutes range). For convenience and accuracy of pipetting, the stock RNase I (which is usually supplied in storage buffer comprising 50% glycerol) can be diluted with the buffer used in the second strand cDNA synthesis step. 1 to 5µl of the diluted mixture is added to the reaction mixture for a final concentration of 0.01 to 0.11 Units per µl reaction volume. The enzyme is inactivated by incubation of the reaction mixture at elevated temperature (70^{0}C to 90^{0}C) for 5 to 20 minutes.

Amplification

[0200] Amplification is carried out according to the previously described method for RNA amplification using chimeric primers, for example, as described in U.S. Patent Application Nos. 2003/0087251or 2004/0005614, or at www.nugeninc.com. An example of amplification reaction conditions is as follows:

IDT Exhibit 1020.2201

5 μl of the second strand cDNA reaction mixture is mixed with 20 μl reaction mixture containing:

2 μl of 10X buffer (200 mM Tris-HCl, pH 8.5, 50 mM $MgCl_2$, 1% NP-40)

0.2 μl of dATP, dGTP, dCTP and dTTP (25 mM)

0.2 μl of amplification chimeric primer (100 μM)

5 μl of amplification enzyme mixture comprising DNA polymerase, RNase H, T4 Gene 32 protein and Rnasin.

DEPC-treated water to a total volume of 20 μl.

The reaction is carried out at 50°C for 60 minutes, and amplification is stopped by heating at 80°C for 5 minutes.

[0201] Although the foregoing invention has been described in some detail by way of illustration and examples for purposes of clarity of understanding, it will be apparent to those skilled in the art that certain changes and modifications may be practiced without departing from the spirit and scope of the invention. Therefore, the description should not be construed as limiting the scope of the invention.

[0202] All publications, patents and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes to the same extent as if each individual publication, patent or patent application were specifically and individually indicated to be so incorporated by reference.

IDT Exhibit 1020.2202

CLAIMS

We claim:

1. A method of generating multiple copies of a polynucleotide sequence complementary to an RNA sequence of interest, said method comprising the steps of:

(a) extending a first primer hybridized to a target RNA with at least one enzyme comprising RNA-dependent DNA polymerase activity, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, whereby a complex comprising a first primer extension product and the target RNA is produced;

(b) cleaving the target RNA in the complex of step (a);

(c) extending a second primer hybridized to the first primer extension product with at least one enzyme comprising DNA-dependent DNA polymerase activity and at least one enzyme comprising RNA-dependent DNA polymerase activity, whereby a second primer extension product is produced to form a complex of first and second primer extension products;

(d) incubating the reaction mixture comprising the products of step (c) and excess first primer, if any, with an enzyme that is capable of cleaving single-stranded RNA;

(e) inactivating the enzyme that is capable of cleaving single-stranded RNA;

(f) cleaving RNA from the first primer in the complex of first and second primer extension products with at least one enzyme that cleaves RNA from an RNA/DNA hybrid such that a composite amplification primer hybridizes to the second primer extension product, wherein the composite amplification primer comprises an RNA portion and a 3' DNA portion; and

(g) extending the composite amplification primer hybridized to the second primer extension product with at least one enzyme comprising DNA-dependent DNA polymerase activity;

whereby said first primer extension product is displaced, RNA is cleaved from the composite amplification primer and another composite amplification primer hybridizes such that primer extension and strand displacement are repeated, and whereby multiple copies of a polynucleotide sequence complementary to the RNA sequence of interest are generated.

IDT Exhibit 1020.2203

2. A method of generating multiple copies of an RNA sequence of interest, said method comprising the steps of:

(a) extending a first primer hybridized to a target RNA with at least one enzyme comprising RNA-dependent DNA polymerase activity, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, whereby a complex comprising a first primer extension product and the target RNA is produced;

(b) cleaving the target RNA in the complex of step (a);

(c) extending a second primer hybridized to the first primer extension product with at least one enzyme comprising DNA-dependent DNA polymerase activity and at least one enzyme comprising RNA-dependent DNA polymerase activity, whereby a second primer extension product is produced to form a complex of first and second primer extension products;

(d) incubating the reaction mixture comprising the products of step (c) and excess first primer, if any, with an enzyme that is capable of cleaving single-stranded RNA;

(e) inactivating the enzyme that is capable of cleaving single-stranded RNA;

(f) cleaving RNA from the first primer in the complex of first and second primer extension products with at least one enzyme that cleaves RNA from an RNA/DNA hybrid such that a composite amplification primer hybridizes to the second primer extension product, wherein the composite amplification primer comprises an RNA portion and a 3' DNA portion;

(g) extending said composite amplification primer hybridized to the second primer extension product with at least one enzyme comprising DNA-dependent DNA polymerase activity, whereby said first primer extension product is displaced, RNA is cleaved from the composite amplification primer and another composite amplification primer hybridizes such that primer extension and strand displacement are repeated; and

(h) hybridizing the displaced first primer extension product with a polynucleotide comprising a propromoter and a region which is hybridizable to the displaced first primer extension product under conditions which allow transcription to occur by RNA polymerase, such that RNA transcripts are produced comprising sequences complementary to the displaced first primer extension product,

IDT Exhibit 1020.2204

whereby multiple copies of the RNA sequence of interest are generated.

3. A method of generating multiple copies of a polynucleotide sequence complementary to an RNA sequence of interest comprising:

(a) incubating a reaction mixture, said reaction mixture comprising:

(i) a target RNA;

(ii) a first primer that is hybridizable to a target RNA, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion; and

(iii) at least one enzyme comprising RNA-dependent DNA polymerase activity

wherein the incubation is under conditions that permit primer hybridization and primer extension, whereby a complex comprising a first primer extension product and the target RNA is formed;

(b) incubating a reaction mixture, said reaction mixture comprising:

(i) the first primer extension product;

(ii) a second primer;

(iii) at least one enzyme comprising DNA-dependent DNA polymerase activity;

(iv) at least one enzyme comprising RNA-dependent DNA polymerase activity; and

(v) optionally, at least one enzyme capable of cleaving RNA from an RNA/DNA hybrid;

wherein the incubation is under conditions permitting formation of a complex comprising the first primer extension product and a second primer extension product;

(c) incubating at least a portion of the reaction mixture of (b) with an enzyme capable of cleaving single-stranded RNA;

(d) inactivating the enzyme capable of cleaving single-stranded RNA; and

(e) incubating a reaction mixture, said reaction mixture comprising:

IDT Exhibit 1020.2205

(i) the reaction products generated according to steps (a), (b), (c), and (d);

(ii) at least one enzyme capable of cleaving RNA from an RNA/DNA hybrid;

(iii) a composite amplification primer, wherein the composite amplification primer comprises a RNA portion and a 3' DNA portion;

(iv) at least one enzyme comprising DNA-dependent DNA polymerase activity, wherein the incubation is under conditions that permit cleavage of RNA, composite primer hybridization, primer extension, and displacement of the first primer extension product from the complex comprising the first primer extension product and the second primer extension product, whereby another composite amplification primer hybridizes and primer extension and strand displacement are repeated;

whereby multiple copies of a polynucleotide sequence complementary to the RNA sequence of interest are generated.

4. A method of generating multiple copies of a polynucleotide sequence complementary to an RNA sequence of interest, said method comprising:

(i) contacting a reaction mixture comprising a complex of first and second primer extension products and a first primer with an enzyme capable of cleaving single-stranded RNA, wherein the first primer extension product is produced by extension of the first primer hybridized to a target RNA with at least one enzyme comprising RNA-dependent DNA polymerase activity, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, wherein the second primer extension product is generated by extension of a second primer hybridized to the first primer extension product;

(ii) inactivating the enzyme capable of cleaving single-stranded RNA; and

(iii) contacting the complex of first and second primer extension products with a composite amplification primer and at least one enzyme that cleaves RNA from an RNA/DNA hybrid, said composite amplification primer comprising an RNA portion and a 3' DNA portion, wherein the composite amplification primer hybridizes to the second primer extension product;

whereby said first primer extension product is displaced, RNA is cleaved from the composite amplification primer and another composite amplification primer hybridizes

IDT Exhibit 1020.2206

such that primer extension and strand displacement are repeated; and whereby multiple copies of a polynucleotide sequence complementary to the RNA sequence of interest are generated.

5. The method of any of claims 1 to 4, wherein the RNA portion of the first primer is 5' with respect to the 3' DNA portion.

6. The method of claim 5, wherein the 5' RNA portion of the first primer is adjacent to the 3' DNA portion.

7. The method of any of claims 1 to 6, wherein the first primer is a tailed primer that comprises a 5' portion that is not hybridizable to the target RNA under conditions in which the first primer hybridizes to the target RNA.

8. The method of any of claims 1 to 7, wherein the first primer comprises a random primer sequence.

9. The method of any of claims 1 to 7, wherein the 3' DNA portion of the first primer comprises a poly-dT sequence.

10. The method of claim 9, wherein the 3' DNA portion of the first primer comprises at least 1 random nucleotide at the 3' end.

11. The method of any of claims 1 to 10, wherein the RNA portion of the composite amplification primer is 5' with respect to the 3' DNA portion.

12. The method of claim 11, wherein the 5' RNA portion of the composite amplification primer is adjacent to the 3' DNA portion.

13. The method of any of claims 1 to 12, wherein the target RNA is mRNA.

14. The method of any of claims 1 to 13, wherein the second primer comprises a fragment of the target RNA hybridized to the first primer extension product.

IDT Exhibit 1020.2207

15. The method of any of claims 1 to 13, wherein the second primer comprises DNA.

16. The method of any of claims 1 to 13 and 15, wherein the second primer is a random primer.

17. The method of any of claims 1 to 16, wherein the at least one enzyme that cleaves RNA from an RNA/DNA hybrid is RNase H.

18. The method of any of claims 1 to 17, wherein the enzyme capable of cleaving single-stranded RNA is RNase I.

19. The method of claim 18, wherein said inactivating said RNase I comprises inactivation by heat.

20. The method of claim 1, wherein step (g) comprises extension of said composite amplification primer in the presence of a non-canonical nucleotide.

21. The method of claim 3, wherein the reaction mixture of step (e) further comprises a non-canonical nucleotide.

22. The method of claim 4, wherein step (iii) comprises extension of said composite amplification primer in the presence of a non-canonical nucleotide.

23. The method of any of claims 20 to 22, wherein the non-canonical nucleotide is dUTP.

24. The method of claim 1, wherein step (g) comprises extension of said composite amplification primer in the presence of a labeled nucleotide.

25. The method of claim 1, wherein step (g) comprises extension of said composite amplification primer in the presence of a labeled nucleotide terminator.

IDT Exhibit 1020.2208

26. The method of claim 3, wherein the reaction mixture of step (e) further comprises a labeled nucleotide.

27. The method of claim 3, wherein the reaction mixture of step (e) further comprise a labeled nucleotide terminator.

28. The method of claim 4, wherein step (iii) comprises extension of said composite amplification primer in the presence of a labeled nucleotide.

29. The method of claim 4, wherein step (iii) comprises extension of composite amplification primer in the presence of a labeled nucleotide terminator.

30. The method of any of claims 1, 3, or 4, wherein said multiple copies of a polynucleotide sequence complementary to a sequence of interest are contacted with an enzyme that is capable of cleaving single-stranded RNA, wherein RNA from the composite amplification primer is cleaved.

31. The method of claim 30, wherein the enzyme that is capable of cleaving single-stranded RNA is RNase I.

32. A method for amplification of a polynucleotide template, comprising:

(a) extending a first primer hybridized to the polynucleotide template with at least one enzyme comprising a DNA polymerase activity, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, whereby a complex comprising a first primer extension product and the template polynucleotide is produced;

(b) extending a second primer hybridized to the first primer extension product to produce a complex comprising a first primer extension product and a second primer extension product;

(c) incubating the reaction mixture comprising the products of step (b) and excess first primer, if any, with an enzyme that is capable of cleaving single-stranded RNA, whereby unhybridized single-stranded RNA of the RNA portion of the first primer is cleaved;

(d) inactivating the enzyme that is capable of cleaving single-stranded RNA;

IDT Exhibit 1020.2209

(e) cleaving RNA from the first primer in the complex of first primer extension product and second primer extension product with at least one enzyme that cleaves RNA from an RNA/DNA hybrid such that a composite amplification primer hybridizes to the second primer extension product, wherein the composite amplification primer comprises an RNA portion and a 3' DNA portion; and

(f) extending the composite amplification primer hybridized to the second primer extension product with at least one enzyme comprising DNA-dependent DNA polymerase activity,

whereby the first primer extension product is displaced, RNA is cleaved from the composite amplification primer and another composite amplification primer hybridizes such that primer extension and strand displacement are repeated, and whereby multiple copies of a polynucleotide amplification product are generated.

33. A method for amplification of a template polynucleotide, comprising:

(a) incubating a reaction mixture, said reaction mixture comprising:

(i) a template polynucleotide;

(ii) a first primer, wherein the first primer is a composite primer that is hybridizable to at least one template site, wherein the composite primer comprises an RNA portion and a 3' DNA portion; and

(iii) a DNA polymerase;

wherein the incubation is under conditions that permit primer hybridization and primer extension, whereby a complex comprising a first primer extension product and the template polynucleotide is formed;

(b) incubating a reaction mixture, said reaction mixture comprising:

(i) the first primer extension product;

(ii) a second primer;

(iii) at least one enzyme comprising DNA-dependent DNA polymerase activity;

(iv) at least one enzyme comprising RNA-dependent DNA polymerase activity; and

(v) optionally, at least one enzyme capable of cleaving RNA from an RNA/DNA hybrid;

IDT Exhibit 1020.2210

wherein the incubation is under conditions permitting formation of a complex comprising the first primer extension product and a second primer extension product;

(c) incubating at least a portion of the reaction mixture of (b) with an enzyme capable of cleaving single-stranded RNA;

(d) inactivating the enzyme capable of cleaving single-stranded RNA; and

(e) incubating a reaction mixture, said reaction mixture comprising:

(i) the reaction products generated according to steps (a), (b), (c), and (d);

(ii) at least one enzyme capable of cleaving RNA from an RNA/DNA hybrid;

(iii) a composite amplification primer, wherein the composite amplification primer comprises an RNA portion and a 3' DNA portion;

(iv) at least one enzyme comprising DNA-dependent DNA polymerase activity, wherein the incubation is under conditions that permit cleavage of RNA, composite primer hybridization, primer extension, and displacement of the first primer extension product from the complex comprising the first primer extension product and the second primer extension product, whereby another composite amplification primer hybridizes and primer extension and strand displacement are repeated;

whereby multiple copies of a polynucleotide amplification product are generated.

34. A method for amplification of a template polynucleotide, said method comprising:

(i) contacting a reaction mixture comprising a first primer and a complex of first primer extension product and second primer extension product with an enzyme capable of cleaving single-stranded RNA, wherein the complex of first primer extension product and second primer extension product is produced by extension of a first primer hybridized to the template polynucleotide with a DNA polymerase, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, wherein the second primer extension product is generated by extension of a second primer hybridized to the first primer extension product;

(ii) inactivating the enzyme capable of cleaving single-stranded RNA; and

(iii) contacting the complex of first primer extension product and second primer extension product with at least one enzyme that cleaves RNA from an RNA/DNA hybrid and a composite amplification primer, said composite amplification primer comprising an

IDT Exhibit 1020.2211

RNA portion and a 3' DNA portion, wherein the composite amplification primer hybridizes to the second primer extension product;

whereby said first primer extension product is displaced, RNA is cleaved from the composite amplification primer, and another composite amplification primer hybridizes such that primer extension and strand displacement are repeated; and whereby multiple copies of a polynucleotide amplification product are generated.

35. The method of any of claims 32 to 34, wherein the RNA portion of the first primer is 5' with respect to the 3' DNA portion.

36. The method of any of claim 35, wherein the 5' RNA portion of the first primer is adjacent to the 3' DNA portion.

37. The method of any of claims 32 to 36, wherein the first primer is hybridizable to a multiplicity of template polynucleotide sites.

38. The method of any of claims 32 to 37, wherein the RNA portion of the composite amplification primer is 5' with respect to the 3' DNA portion.

39. The method of claim 38, wherein the 5' RNA portion of the composite amplification primer is adjacent to the 3' DNA portion.

40. The method of any of claims 32 to 39, wherein the template polynucleotide is DNA.

41. The method of any of claims 32 to 39, wherein the template polynucleotide is RNA.

42. The method of any of claims 32 to 41, wherein the first primer is at least two different primers.

43. The method of any of claims 32 to 42, wherein the first primer comprises a random sequence.

IDT Exhibit 1020.2212

44. The method of any of claims 32 to 43, wherein the at least one enzyme that cleaves RNA from an RNA/DNA hybrid is RNase H.

45. The method of any of claims 32 to 44, wherein the enzyme capable of cleaving single-stranded RNA is RNase I.

46. The method of claim 45, wherein said inactivating said RNase I comprises inactivation by heat.

47. The method of claim 32, wherein step (f) comprises extension of said composite amplification primer in the presence of a non-canonical nucleotide.

48. The method of claim 33, wherein the reaction mixture of step (e) further comprises a non-canonical nucleotide.

49. The method of claim 34, wherein step (iii) comprises extension of said composite amplification primer in the presence of a non-canonical nucleotide.

50. The method of any of claims 47 to 49, wherein the non-canonical nucleotide is dUTP.

51. The method of claim 32, wherein step (f) comprises extension of said composite amplification primer in the presence of a labeled nucleotide.

52. The method of claim 32, wherein step (f) comprises extension of said composite amplification primer in the presence of a labeled nucleotide terminator.

53. The method of claim 33, wherein the reaction mixture of step (e) further comprises a labeled nucleotide.

54. The method of claim 33, wherein the reaction mixture of step (e) further comprises a labeled nucleotide terminator.

IDT Exhibit 1020.2213

55. The method of claim 34, wherein step (iii) comprises extension of said composite amplification primer in the presence of a labeled nucleotide.

56. The method of claim 34, wherein step (iii) comprises extension of said composite amplification primer in the presence of a labeled nucleotide terminator.

57. The method of any of claims 32 to 56, wherein said multiple copies of a polynucleotide amplification product are contacted with an enzyme that is capable of cleaving single-stranded RNA, wherein RNA from the composite amplification primer is cleaved.

58. The method of claim 57, wherein the enzyme that is capable of cleaving single-stranded RNA is RNase I.

59. A method for amplifying a polynucleotide sequence complementary to a target polynucleotide sequence, comprising:

(a) extending a composite primer in a complex comprising (i) a polynucleotide template comprising the target sequence; and (ii) the composite primer, said composite primer comprising an RNA portion and a 3' DNA portion, wherein the composite primer is hybridized to the polynucleotide template;

(b) cleaving the RNA portion of the annealed composite primer with an enzyme that cleaves RNA from an RNA/DNA hybrid such that another composite primer hybridizes to the template and repeats primer extension and strand displacement,

whereby multiple copies of the complementary sequence of the target sequence are produced; and

(c) incubating said multiple copies of the complementary sequence of the target sequence with an enzyme that is capable of cleaving single-stranded RNA, wherein RNA from the composite primer is cleaved.

60. A method of generating multiple copies of a polynucleotide sequence complementary to an RNA sequence of interest, said method comprising:

(a) extending a composite amplification primer in a complex comprising:

IDT Exhibit 1020.2214

(i) a complex of first and second primer extension products, wherein the first primer extension product is produced by extension of a first primer hybridized to a target RNA with at least one enzyme comprising RNA-dependent DNA polymerase activity, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, wherein the second primer extension product is generated by extension of a second primer hybridized to the first primer extension product, and wherein RNA from the complex of first and second primer extension products is cleaved with at least one enzyme that cleaves RNA from an RNA/DNA hybrid; and

(ii) a composite amplification primer, said composite amplification primer comprising an RNA portion and a 3' DNA portion, wherein the composite amplification primer is hybridized to the second primer extension product;

whereby said first primer extension product is displaced, RNA is cleaved from the composite amplification primer and another composite amplification primer hybridizes such that primer extension and strand displacement are repeated; and whereby multiple copies of a polynucleotide sequence complementary to the RNA sequence of interest are generated; and

(b) contacting said multiple copies of the complementary sequence of the RNA sequence of interest with an enzyme that is capable of cleaving single-stranded RNA, wherein RNA from the composite primer is cleaved.

61. The method of claim 59 or 60, wherein the enzyme capable of cleaving single-stranded RNA is RNase I.

62. A kit comprising a first composite primer, a second composite primer, and an enzyme capable of cleaving single-stranded RNA, wherein the first composite primer comprises an RNA portion and a 3' DNA portion, wherein the second composite primer comprises an RNA portion and a 3' DNA portion, and wherein the second composite primer comprises a sequence that is hybridizable to a polynucleotide comprising a complement of the first composite primer.

63. The kit of claim 62, wherein the RNA portion of the first composite primer is 5' with respect to the 3' DNA portion of the first composite primer and the RNA portion

IDT Exhibit 1020.2215

of the second composite primer is 5' with respect to the 3' DNA portion of the second composite primer.

64. The kit of claim 63, wherein the 5' portion of the first composite primer is adjacent to the 3' DNA portion of the first composite primer and the 5' RNA portion of the second composite primer is adjacent to the 3' DNA portion of the second composite primer.

65. The kit of any of claims 62 to 64, wherein the first composite primer is a tailed primer that comprises a 5' portion that is not hybridizable to a target RNA under conditions wherein the first composite primer hybridizes to the target RNA.

66. The kit of any of claims 62 to 65, wherein the3' DNA portion of the first composite primer comprises a poly-dT sequence.

67. The kit of claim 66, wherein the 3' DNA portion of the first composite primer comprises at least 1 random nucleotide at the 3' end.

68. The kit of any of claims 62 to 67, wherein the enzyme capable of cleaving single-stranded RNA is RNase I.

69. The kit of any of claims 62 to 68, further comprising an enzyme that cleaves RNA form an RNA/DNA hybrid.

70. The kit of claim 69, wherein the enzyme that cleaves RNA from an RNA/DNA hybrid is RNaseH.

71. The kit of any of claims 62 to 70, further comprising an RNA-dependent DNA polymerase.

72. The kit of any of claims 62 to 71, further comprising a non-canonical nucleotide.

IDT Exhibit 1020.2216

73. The kit of claim 72, wherein the non-canonical nucleotide is dUTP.

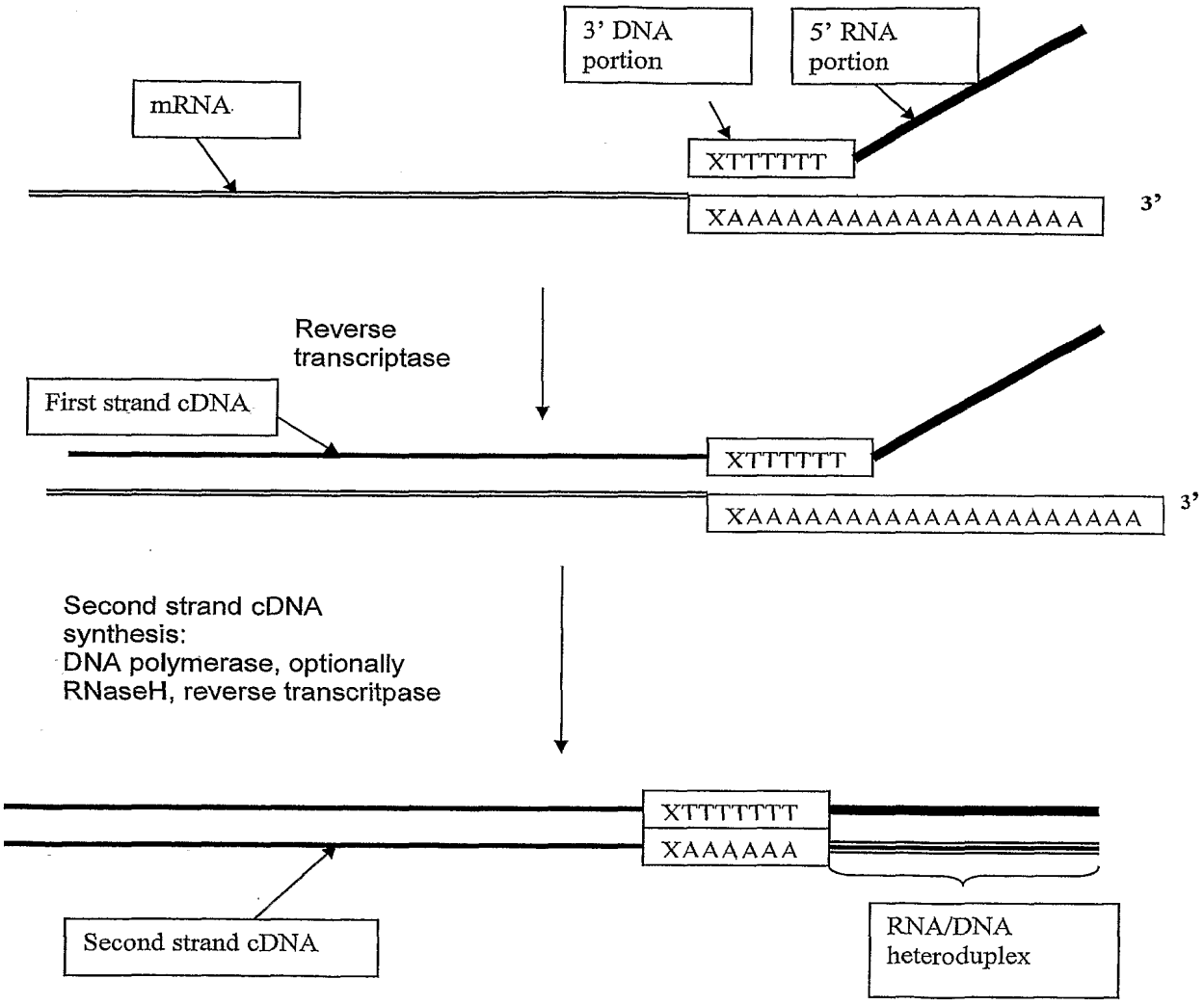
74. The kit of any of claims 62 to 71, further comprising a labeled nucleotide.

75. The kit of any of claims 62 to 71, further comprising a labeled nucleotide terminator.

IDT Exhibit 1020.2217

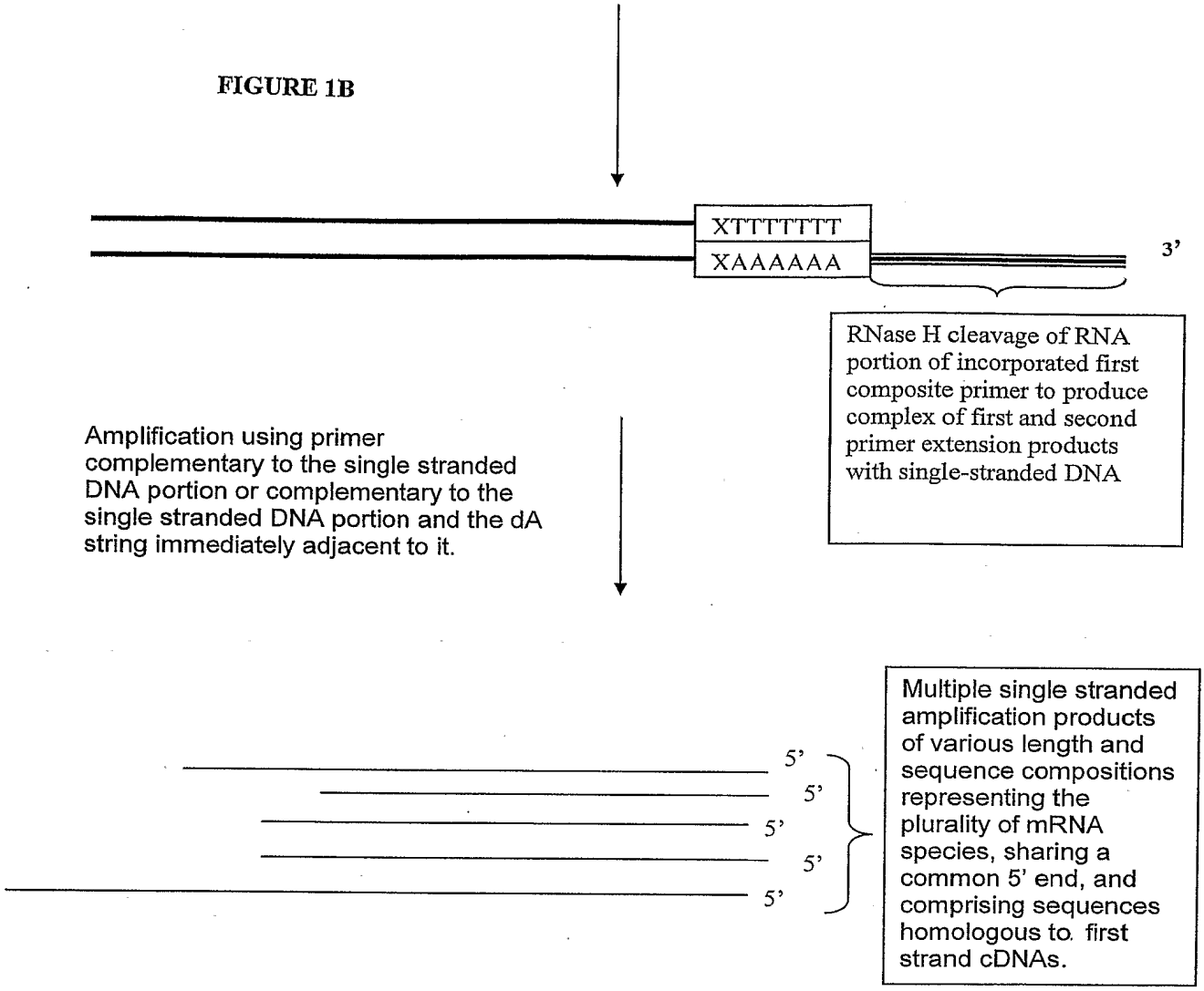
FIGURE 1

FIGURE 1A



IDT Exhibit 1020.2218

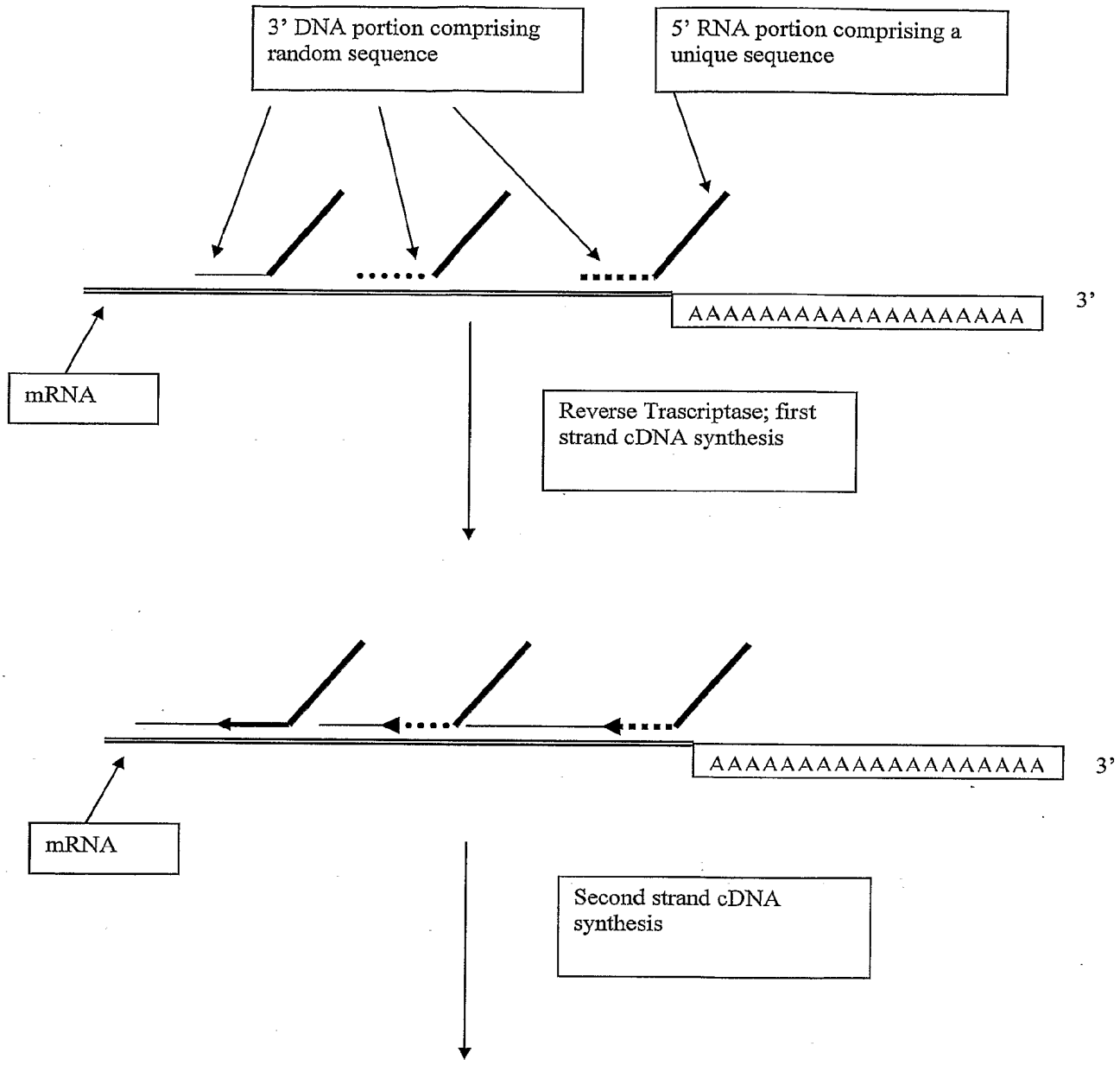
FIGURE 1B



IDT Exhibit 1020.2219

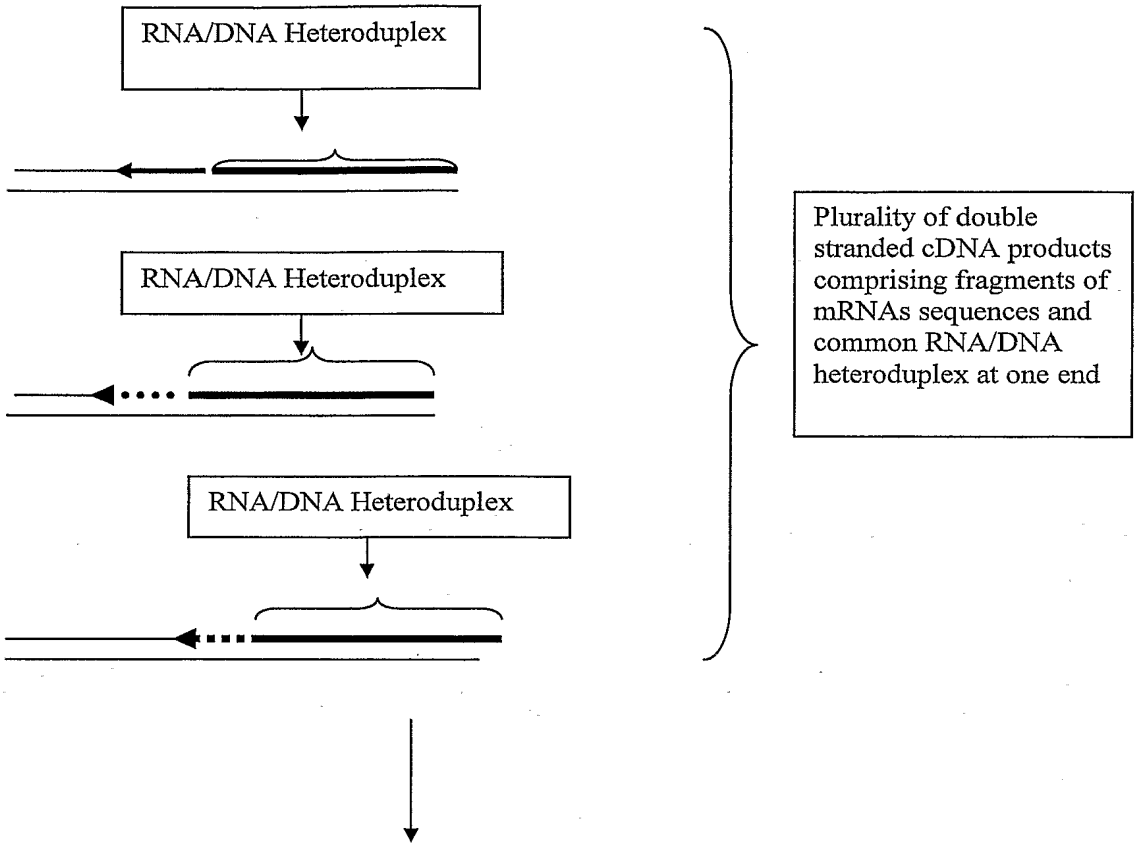
FIGURE 2

FIGURE 2A



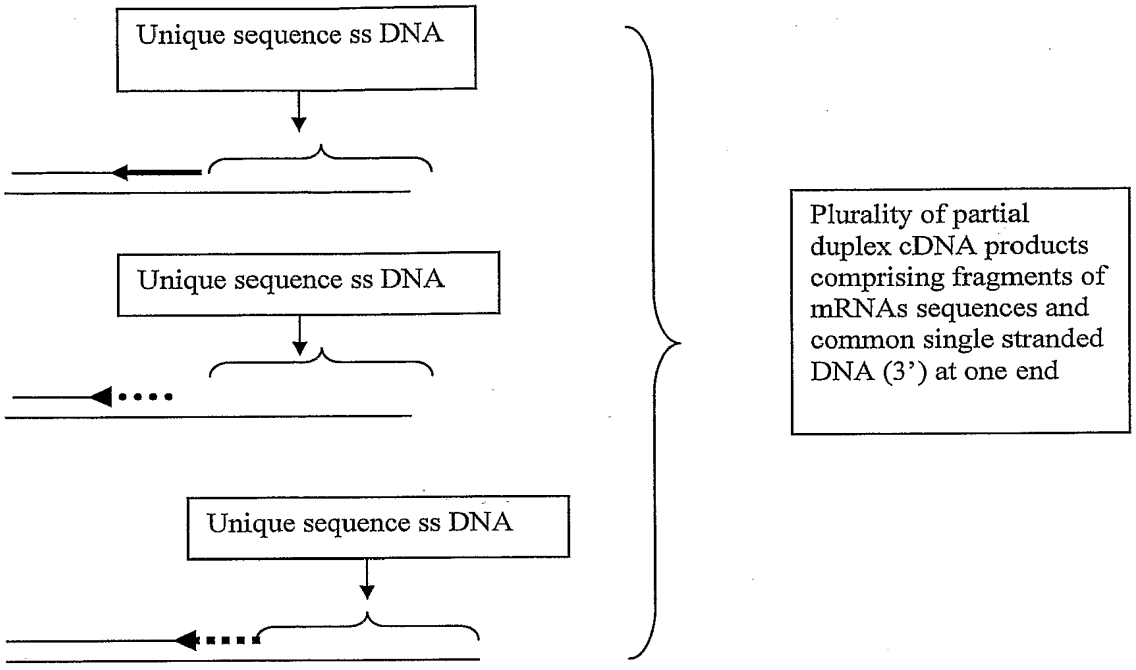
IDT Exhibit 1020.2220

FIGURE 2B



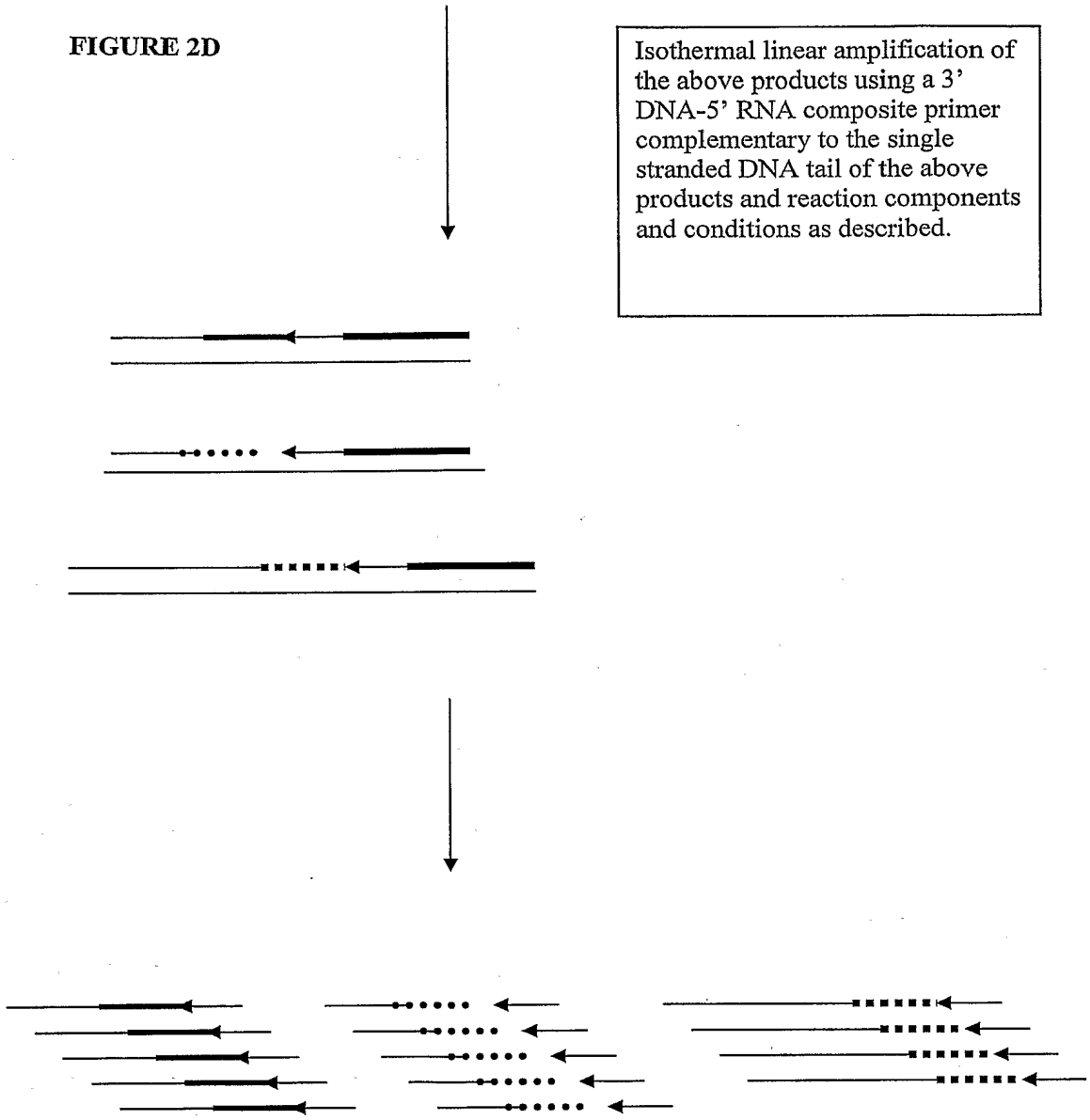
IDT Exhibit 1020.2221

FIGURE 2C



IDT Exhibit 1020.2222

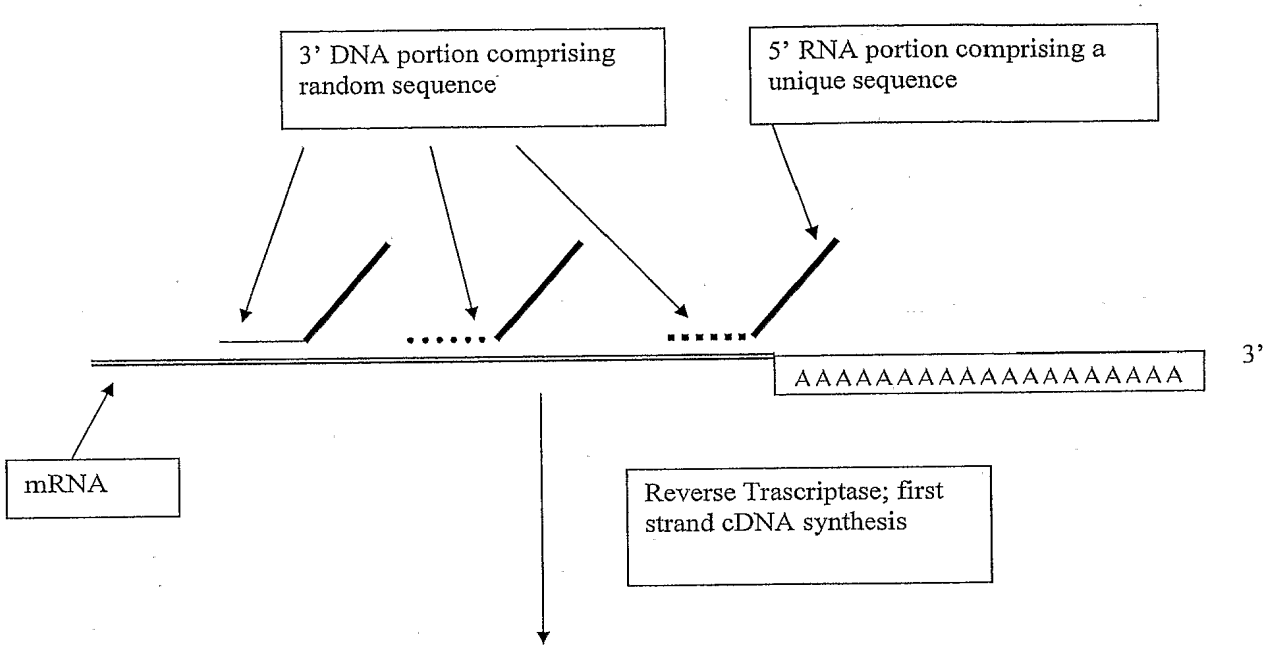
FIGURE 2D



IDT Exhibit 1020.2223

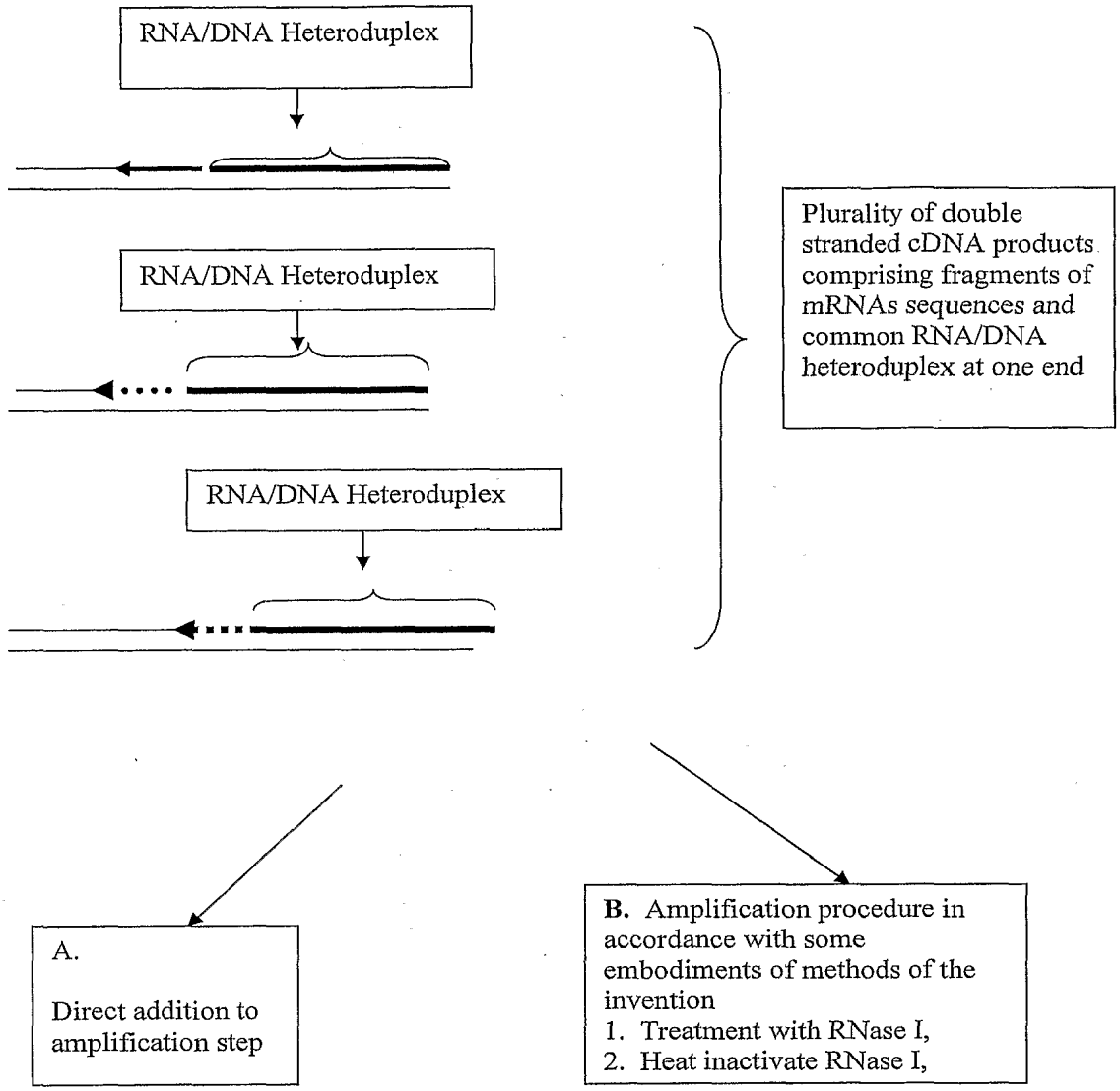
Figure 3

A. Hybridization of a plurality of first strand chimeric primers



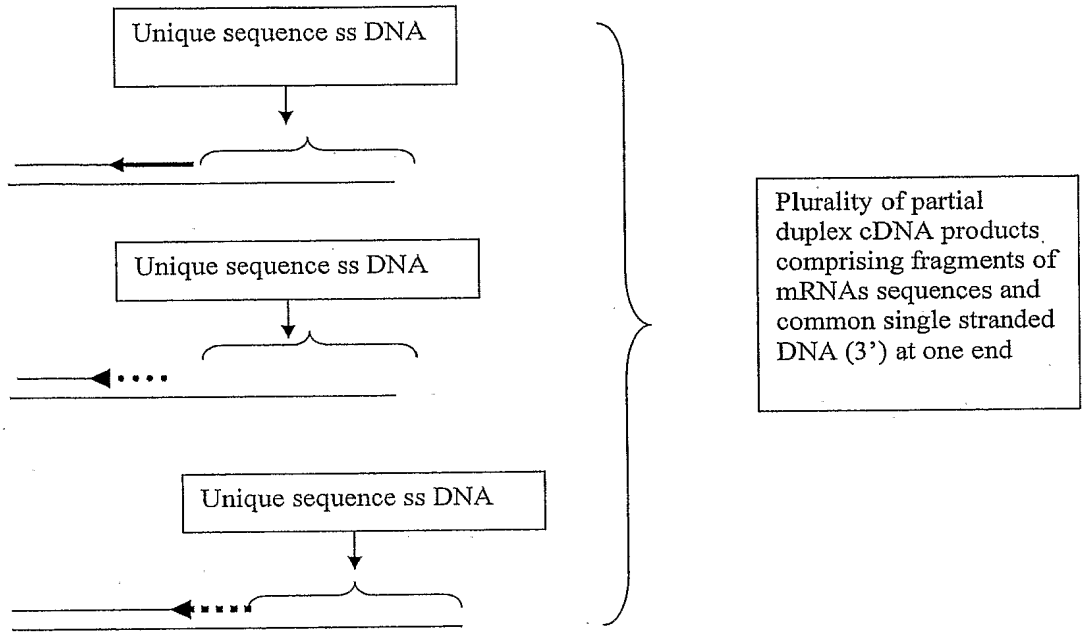
IDT Exhibit 1020.2224

B. Formation of unique double strand cDNA with an RNA/DNA heteroduplex following second strand cDNA synthesis step, and subsequent RNase I treatment



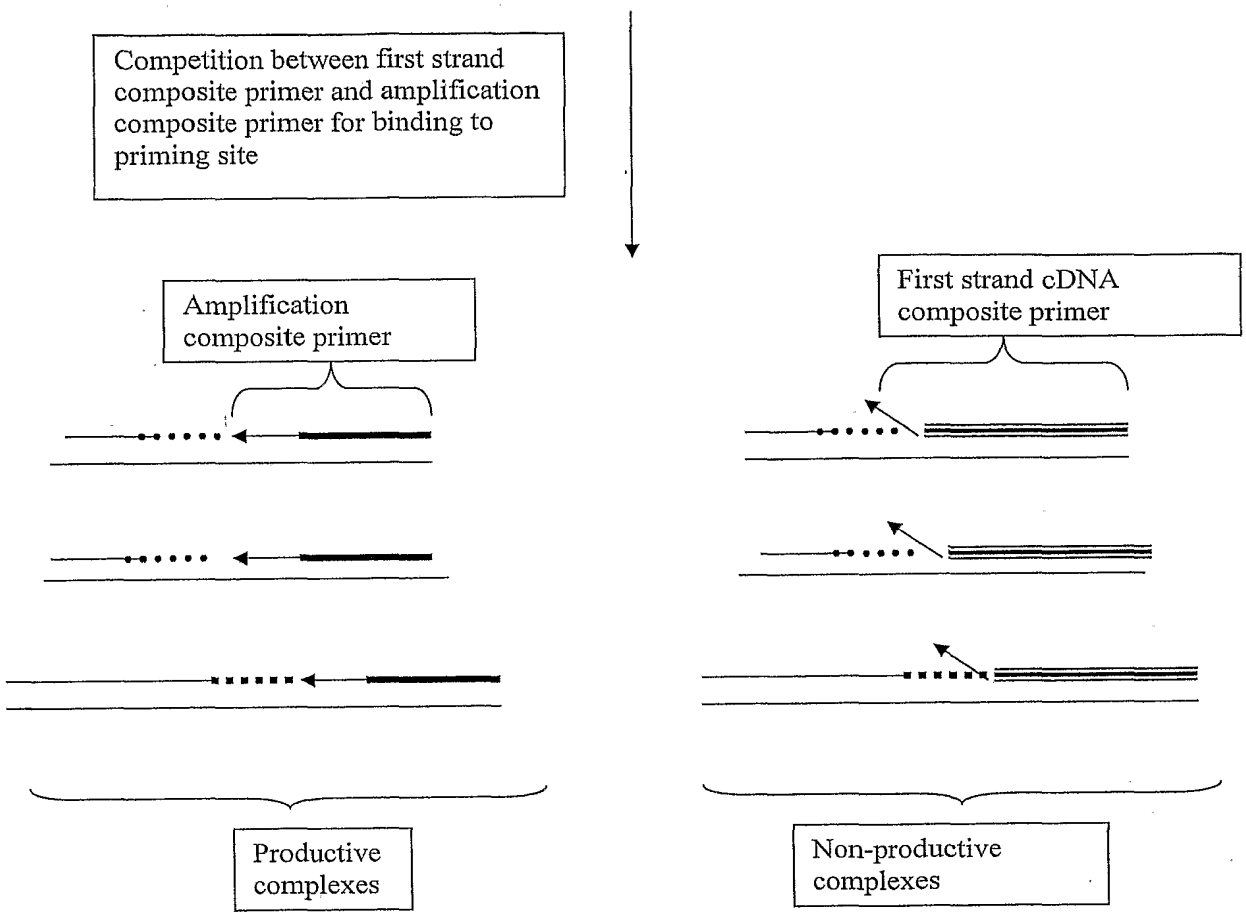
IDT Exhibit 1020.2225

C. Amplification step:
Initial degradation of the RNA portion of the heteroduplex of the double-stranded cDNA (similar reaction with or without RNase I treatment)



IDT Exhibit 1020.2226

D.
Amplification step in the presence of amplification composite primer and first strand cDNA synthesis chimeric primer that is carried over from the first and second strand cDNA synthesis.

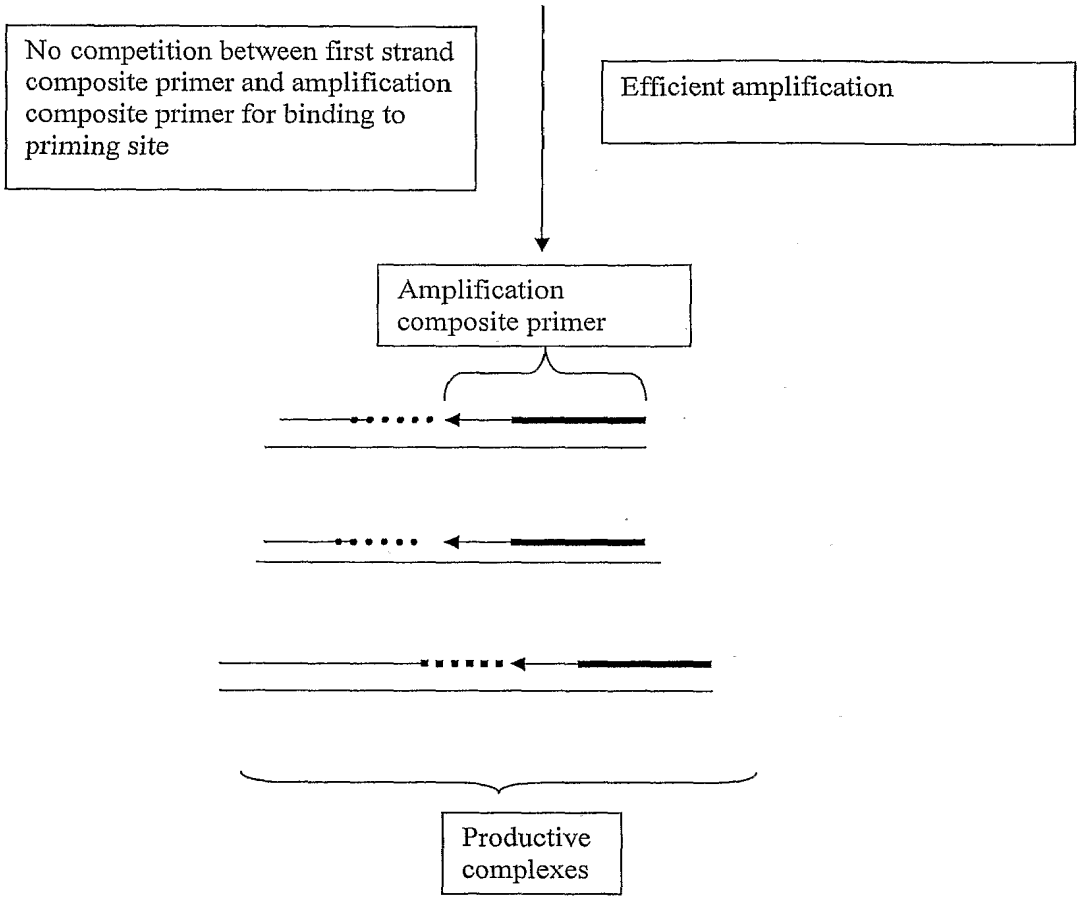


IDT Exhibit 1020.2227

E.

Amplification step following treatment with RNase I:

High efficiency amplification in the absence of competition from first strand-cDNA composite primer.



IDT Exhibit 1020.2228

Espacenet

Bibliographic data: WO2009117698 (A2) — 2009-09-24

METHODS OF RNA AMPLIFICATION IN THE PRESENCE OF DNA

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

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

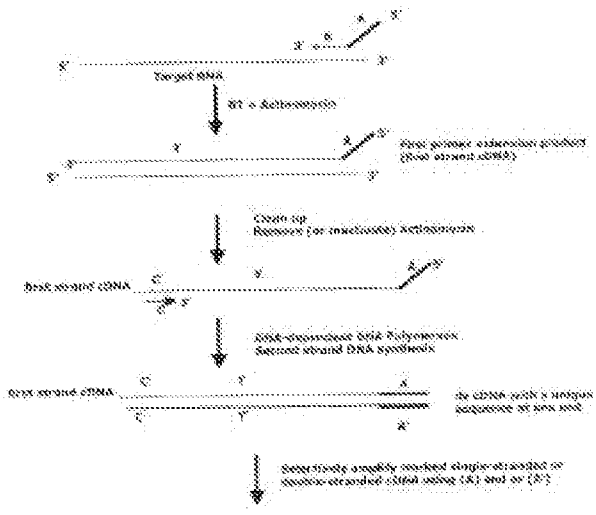
Classification: - international: *C07H21/00; C12N15/10; C12Q1/68*
- cooperative: **C12N15/1096**; **C12P19/34**; **C12Q1/6809**; **C12Q1/6848** more

Application number: WO2009US37870 20090320 Global Dossier

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

Also published as: WO2009117698 (A3) GB2470672 (A) GB2470672 (B) US2009239232 (A1) US7846666 (B2)

Abstract of WO2009117698 (A2)



The invention provides methods for amplification of RNA The methods are particularly suitable for specifically amplifying RNA in the presence of DNA The methods involve producing a marked first primer extension product from a target RNA in the presence of a DNA-dependent DNA polymerase inhibitor, which prevents replication of DNA by the reverse transcriptase enzyme The marked nucleic acid products are subsequently selectively amplified in the presence on non-marked nucleic acids The methods are useful for production and analysis of polynucleotide sequences complementary to an RNA sequence The methods are useful for preparation of nucleic acid libraries and substrates for analysis of gene expression of cells in biological samples The invention also provides

IDT Exhibit 1020.2229

compositions and kits for practicing the amplification methods, as well as methods which use the amplification products.

IDT Exhibit 1020.2230

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 September 2009 (24.09.2009)

PCT

(10) International Publication Number
WO 2009/117698 A2

(51) **International Patent Classification**:
C12N 15/10 (2006.01) ***C07H 21/00*** (2006.01)
C12Q 1/68 (2006.01)

(21) **International Application Number**:
PCT/US2009/037870

(22) **International Filing Date**:
20 March 2009 (20.03.2009)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:

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) **Inventor; and**
(75) **Inventor/Applicant** *(for US only)*: **KURN, Nurith** [IL/US]; 2876 Ramona Street, Palo Alto, CA 94306 (US).

(74) **Agents**: **GOETZ, David, H.** 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 2009/117698 A2

(54) **Title**: METHODS OF RNA AMPLIFICATION IN THE PRESENCE OF DNA

(57) **Abstract**: The invention provides methods for amplification of RNA The methods are particularly suitable for specifically amplifying RNA in the presence of DNA The methods involve producing a marked first primer extension product from a target RNA in the presence of a DNA-dependent DNA polymerase inhibitor, which prevents replication of DNA by the reverse transcriptase enzyme The marked nucleic acid products are subsequently selectively amplified in the presence on non-marked nucleic acids The methods are useful for production and analysis of polynucleotide sequences complementary to an RNA sequence The methods are useful for preparation of nucleic acid libraries and substrates for analysis of gene expression of cells in biological samples The invention also provides compositions and kits for practicing the amplification methods, as well as methods which use the amplification products.

IDT Exhibit 1020.2231

METHODS OF RNA AMPLIFICATION IN THE PRESENCE OF DNA

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application Nos. 61/038,696, filed March 21, 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 related to the following co-pending international patent applications PCT/US2009/033936 and PCT/ US2009/033964 filed February 12, 2009, which claim the benefit of the foregoing provisional applications and U.S. Provisional Application No. 61/028,146; filed February 12, 2008; and which applications are all incorporated herein by reference in their entirety.

BACKGROUND OF THE INVENTION

[0002] The detection and/or quantitation of specific nucleic acid sequences is an important technique for identifying and classifying microorganisms, diagnosing infectious diseases, detecting and characterizing genetic abnormalities, identifying genetic changes associated with cancer, studying genetic susceptibility to disease, and measuring response to various types of treatment. Such procedures are also useful in detecting and quantitating microorganisms in foodstuffs, water, industrial and environmental samples, seed stocks, and other types of material where the presence of specific microorganisms may need to be monitored. Other applications are found in the forensic sciences, anthropology, archaeology, and biology where measurement of the relatedness of nucleic acid sequences has been used to identify criminal suspects, resolve paternity disputes, construct genealogical and phylogenetic trees, and aid in classifying a variety of life forms.

[0003] The ability to amplify ribonucleic acid (RNA) is an important aspect of efforts to elucidate biological processes. Amplification of the total cellular mRNAs prepared from any cell or tissue is important for gene expression profiling. 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. Although analysis of non-amplified mRNA is feasible, a significant amount of starting mRNA can be required. Thus, amplification methods capable of amplifying low abundance RNAs, are of great importance.

[0004] RNA amplification is 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 amplification techniques such as polymerase chain reaction (PCR) or linear isothermal amplification to produce multiple copies of single or double stranded DNA, or RNA. However, the total amount of sample RNA that is available is frequently limited by the amount of biological sample from which it is derived.

IDT Exhibit 1020.2232

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.

[0005] RNA in biological samples is often in the presence of DNA. Amplification of the target RNA in the presence of DNA results in unwanted amplification products as described herein. It is desirable to prevent these unwanted products that originate from initiation on DNA because they may interfere with analysis of target RNA amplification products, result in erroneous conclusions and affect the amplification yield from the target RNA. Purification of RNA from DNA results in reduced yield and/or RNA quality. Therefore, it is highly desirable to develop improved amplification methods of target RNA in the presence of DNA. Moreover, the ability to selectively amplify RNA in a sample comprising total nucleic acid from a biological sample will also assist in the development of procedures and methods for selective amplification in situ as well as directly from stabilized cell lysates. This is especially useful when there are minute amounts of sample for analysis.

[0006] Therefore, there is a need for improved RNA amplification methods that overcome drawbacks in existing methods. The invention provided herein fulfills this need and provides additional benefits.

[0007] All references cited herein, including patent applications and publications, are incorporated by reference in their entirety.

SUMMARY OF THE INVENTION

[0008] One aspect of the invention comprises a method of generating multiple copies of a polynucleotide sequence of or complementary to target RNA which is in a sample comprising DNA, said method comprising the steps of: (a) hybridizing to the target RNA a first primer comprising a sequence (A) that is not complementary to the target RNA, and a sequence (B) at the 3'-end which hybridizes to the target RNA; (b) extending the first primer with at least one enzyme comprising RNA-dependent DNA polymerase activity in the presence of at least one compound comprising DNA-dependent DNA polymerase inhibitor activity, whereby a complex comprising a primer extension product and the target RNA is produced, whereby the first primer extension product comprises a sequence (Y) that is complementary to the target RNA and comprises sequence (A). In one embodiment of this aspect of the invention, the first primer extension product is marked or labeled with the sequence (A) by the use of the first primer comprising the sequence (A). The method further comprises the steps of: (c) disabling or removing at least one compound comprising DNA-dependent DNA polymerase inhibitor activity; and (d) producing multiple copies of a polynucleotide sequence complementary to the target RNA and/or

IDT Exhibit 1020.2233

complementary to sequence (Y) using sequence (A). In one embodiment of this aspect of the invention, the method selectively generates multiple copies of the target RNA polynucleotide sequence of interest in the presence of non-target DNA and/or RNA. In some embodiments, the method is suitable for amplifying one or more whole transcriptomes or a substantial fraction thereof in the presence of DNA. In one embodiment of this aspect of the invention, the method generates amplified polynucleotides suitable for downstream analysis by sequencing, expression profiling, Q-PCR, digital PCR, serial analysis of gene expression (SAGE), RFLP analysis, SNP detection or massively parallel analytical methods including for example massively parallel sequencing, massively parallel expression profiling, microarray analysis, high density PCR arrays, pyrosequencing, sequencing by synthesis, sequencing by ligation, sequencing by hybridization, single molecule sequencing, nanopore sequencing, or massarray analysis. In one embodiment of this aspect of the invention, the method generates amplified polynucleotides in a clonal fashion on a bead, in an emulsion, or as polonies.

[0009] One aspect of the invention comprises a method of generating multiple copies of a polynucleotide sequence complementary to a target RNA which is in a sample comprising DNA, said method comprising the steps of: (a) hybridizing to the target RNA a first primer comprising a sequence (A) that does not hybridize to the target RNA, and a sequence (B) at the 3'-end, which hybridizes to the target RNA; (b) extending the first primer with at least one enzyme comprising RNA-dependent DNA polymerase activity in the presence of at least one compound comprising DNA-dependent DNA polymerase inhibitor activity, whereby a complex comprising a first primer extension product and the target RNA is produced, whereby the first primer extension product comprises a sequence (Y) that is complementary to the target RNA and comprises sequence (A). In one embodiment of this aspect of the invention, the first primer extension product is marked or labeled with the sequence (A) by the use of the first primer comprising the sequence (A). The method further comprises the steps of: (c) disabling or removing at least one compound comprising DNA-dependent DNA polymerase inhibitor activity; (d) cleaving the target RNA in the complex of step (b); (e) extending a second primer along the first primer extension product with at least one enzyme comprising DNA-dependent DNA polymerase activity, wherein the second primer comprises sequence (C) that is complementary to a sequence (C') on the first primer extension product to produce a complex comprising the first primer extension product and a second primer extension product, whereby the second primer extension product comprises the sequence (C), a sequence (Y') complementary to sequence (Y) and sequence (A') complementary to sequence (A); (f) producing multiple copies of a polynucleotide sequence complementary to the target RNA using a primer with a sequence (A) and/or sequence (A'). In one embodiment of this aspect of the invention, the method selectively generates multiple copies of the target RNA polynucleotide sequence of interest in the presence of non-target DNA and/or RNA. In some embodiments, the method is suitable for amplifying one or more whole transcriptomes or a substantial fraction thereof in the presence of DNA. In one embodiment of this aspect of the invention, the methods generates amplified polynucleotides suitable for downstream analysis by sequencing, expression profiling, Q-PCR, digital PCR, serial analysis of gene expression (SAGE), RFLP analysis, SNP detection or massively parallel analytical methods including for example massively parallel sequencing, massively

IDT Exhibit 1020.2234

parallel expression profiling, microarray analysis, high density PCR arrays, pyrosequencing, sequencing by synthesis, sequencing by ligation, sequencing by hybridization, single molecule sequencing, nanopore sequencing, or massarray analysis.

[0010] In some embodiments of the methods, the first primer comprises a 5' sequence that is not hybridizable to the target RNA. In some embodiments of the methods, the sequence (A') is used as a priming site for amplification.

[0011] In some embodiments of the methods, the producing of multiple copies in step (f) are produced using amplification methods comprising polymerase chain reaction (PCR). In some embodiments of the methods, a PCR primer comprises sequence (A) of the first primer. In some embodiments of the methods, a PCR primer comprises sequence (C) of the second primer. In some embodiments of the methods the producing of multiple copies in step (f) comprises performing PCR with primer pairs wherein the first PCR primer is complementary to sequence (A') and the second PCR primer is complementary to all or a portion of sequence (C') or all or a portion of sequence (Y). In some embodiments of the methods the producing of multiple copies in step (f) comprises performing single primer, linear PCR using a PCR primer complementary to sequence (A').

[0012] In some embodiments of the methods the producing of multiple copies in step (f) comprises an amplification method comprising the use of an RNA/DNA composite primer, RNase H, and a DNA dependent DNA polymerase with strand displacement activity.

[0013] In some embodiments of the methods, the first primer comprises an RNA and a DNA sequence wherein a 5'-RNA sequence comprises sequence (A) and a 3'-DNA sequence comprises sequence (B); whereby the complex formed in step (e) comprises an RNA/DNA heteroduplex at one end; and wherein the producing of multiple copies in step (f) comprises the steps of (i) cleaving the RNA from the heteroduplex, (ii) hybridizing to sequence (A') an amplification primer comprising a 5' RNA sequence and a 3' DNA sequence, (iii) extending the amplification primer with at least one enzyme comprising DNA dependent DNA polymerase activity and comprising strand displacement activity, (iv) cleaving the RNA from the extended amplification primer in the heteroduplex with at least one enzyme comprising a specificity for cleaving RNA in a DNA-RNA heteroduplex; and (v) repeating steps (ii) through (iv) to produce multiple copies of amplification product comprising sequence (Y) that is complementary to a portion of the RNA template.

[0014] In some embodiments of the methods, the producing of multiple copies in step (f) comprises in-vitro-transcription (IVT). In some embodiments of the methods, sequence (A) comprises a pro-promoter, whereby the complex formed in step (e) comprises a double stranded promoter for a DNA dependent RNA polymerase comprising sequences (A) and (A'), wherein (A) is a DNA sequence.

[0015] In some embodiments of the methods, the second primer comprises a fragment from the cleaved RNA target. In some embodiments of the methods, the fragment is generated by cleaving RNA in the complex of step (b) with an enzyme that cleaves RNA from an RNA/DNA heteroduplex, by heat, or by chemical means.

IDT Exhibit 1020.2235

[0016] In some embodiments of the methods, the DNA-dependent DNA polymerase inhibitor activity is derived from a compound comprising at least one of the following: actinomycin, alpha-amanitin, aphidicolin, BP5, novobiocin, rifampicin, rifamycin, sulfoquinovosylmonoacylglycerol, sulfoquinovosyldiacylglycerol, ursane, oleanane triterpenoids, ursolic acid, oleanolic acid, mikanolide, dihydromikanolide, dehydroaltenusin, catapol, taxinine, cephalomanninine, dipeptide alcohols, corylifolin; bakuchiol; resveratrol; Neobavaisoflavone; daidzein; bakuchicin, levodopa, dopamine, anacardic acid and oleic acid. In some embodiments of the methods, the DNA-dependent DNA polymerase inhibitor activity is derived from a compound comprising actinomycin.

[0017] In some embodiments of the methods, the sample comprises total nucleic acid in a biological fluid. In some embodiments of the methods, the biological fluid is selected from the group consisting of: plasma, serum, urine, saliva. In some embodiments of the methods, the sample comprises viral RNA. In some embodiments of the methods, the sample comprises cell or tissue lysates. In some embodiments of the methods, the sample comprises microbial RNA in the presence or absence of additional cell or tissue components, such as a mixed population or samples suspected of the presence of infectious agent or agents.

[0018] In some embodiments of the methods, the second primer comprises a tailed primer with a 5'-sequence which does not hybridize to the first primer extension product and which comprises a propromoter sequence such that RNA transcripts are produced comprising a sequence homologous to the target RNA; whereby multiple copies of the homologous sequence of the RNA sequence of interest are generated.

[0019] In some embodiments of the methods, the producing of multiple copies in step (f) comprises hybridizing a PTO, comprising a propromoter sequence and a 3'-hybridizing region, to the 3'-end of a first primer extension product under conditions which allow transcription to occur by RNA polymerase, such that RNA transcripts are produced comprising sequences homologous to the target RNA; whereby multiple copies of the RNA sequence of interest are generated.

[0020] In some embodiments of the methods, the producing of multiple copies in step (f) comprises hybridizing a PTO, comprising a propromoter sequence and a 3'-hybridizing region, to the 3'-end of a second primer extension under conditions which allow transcription to occur by RNA polymerase, such that RNA transcripts are produced comprising sequences complementary to the target RNA; whereby multiple copies of the complementary sequence of the RNA sequence of interest are generated.

[0021] In some embodiments of the methods, the target RNA is mRNA. In some embodiments of the methods, the first primer comprises a sequence complementary to poly-A on RNA. In some embodiments of the methods, the sequence (B) which hybridizes to the target RNA comprises a random sequence. In some embodiments of the methods, the first primer comprises a portion of the sequence complementary to poly-A on RNA and a portion comprising random sequences. In some embodiments of the methods, the first primer comprises a portion of a sequence complementary to specific sequence, or sequences, on the RNA target or targets. In some embodiments of the methods, the first primer comprises a mixture of primers comprising a portion of the sequence complementary to multiple sequences on the target RNA,

IDT Exhibit 1020.2236

and may further comprise first primers comprising a sequence complementary to the poly-A sequence, or any combination of primers described herein.

[0022] In some embodiments of the methods, the method comprises generating multiple copies of two or more different sequences of interest. In some embodiments of the methods, the method comprises at least two different first primers.

[0023] In some embodiments of the methods, the sequence (B) which hybridizes to the target RNA comprises a poly-dT sequence and the target RNA is mRNA.

[0024] In some embodiments of the methods, the enzyme that cleaves RNA is RNase H.

[0025] In some embodiments of the methods, the method comprises sequencing an RNA sequence of interest, said method comprising analyzing amplification products to determine a sequence, said amplification products produced by the methods of the invention in the presence of a mixture of NTPs and NTP analogs such that transcription is terminated upon incorporation of an NTP analog. In some embodiments of the methods, the target RNA is mRNA. In some embodiments of the methods, the methods comprise analyzing amplification products for sequence analysis using any sequencing methods including massively parallel sequencing, pyrosequencing, sequencing by synthesis, sequencing by ligation, sequencing by hybridization, single molecule sequencing, or nanopore sequencing. . In some embodiments of the methods, the target RNA is mRNA, a whole transcriptome, or substantial fraction thereof, or a set of whole transcriptomes or substantial fractions thereof.

[0026] In some embodiments of the methods, the method of detecting a mutation in a target RNA, comprising analyzing sequences of amplification products for the presence of a mutation as compared to a reference polynucleotide sequence, said amplification products produced by the methods of the invention. In some embodiments of the methods, the target RNA is mRNA, a whole transcriptome, or substantial fraction thereof, or a set of whole transcriptomes or substantial fractions thereof.

[0027] In some embodiments of the methods, the method of producing a nucleic acid immobilized to a substrate comprising immobilizing amplification products on a substrate, said amplification products produced by the methods of the invention. In some embodiments of the methods, the target RNA is mRNA. In some embodiments of the methods, the substrate is a microarray. . In some embodiments of the methods, the target RNA is mRNA, a whole transcriptome, or substantial fraction thereof, or a set of whole transcriptomes or substantial fractions thereof.

[0028] One aspect of the invention comprises a kit for amplifying a target RNA comprising: a tailed first primer, reverse transcriptase enzyme, a DNA-dependent DNA polymerase inhibitor, and optionally, a second primer. In some embodiments, the first tailed primer comprises a 3' random portion, and the second primer comprises a random primer. In some embodiments, the kit further comprises a composite amplification primer wherein the tailed first primer is a composite primer comprising a 3' DNA portion and a 5' RNA portion. In some embodiments, the kit further comprises instructions for the use of said kit. In some embodiments, the kit further comprises a DNA-dependent DNA polymerase. In some embodiments, the kit further comprises reagents for DNA purification. In some embodiments, the kit

IDT Exhibit 1020.2237

further comprises reagents for forming an emulsion. In some embodiments, the kit further comprises a solid substrate such as a bead or beads.

INCORPORATION BY REFERENCE

[0029] 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

[0030] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0031] Figure 1 is a diagrammatic representation of an embodiment of the invention involving the creation of a marked double-stranded cDNA from a target RNA sequence using a first primer containing a sequence portion (A) that is non-complementary to the target RNA sequence, such that the first primer extension product is created from target RNA in the presence of a DNA-dependent DNA polymerase (actinomycin). A second primer is used to generate a double-stranded cDNA product comprising a sequence complementary to the target RNA and a unique terminal sequence. The marked single stranded or double stranded cDNA are amplified in a manner such that the marked cDNA is selectively amplified.

[0032] Figure 2a and 2b provide diagrammatic representation of an embodiment the invention in which a composite primer is used as the first primer (Figure 2a) to produce a first primer extension product from RNA target in the presence of a DNA-dependent DNA polymerase (actinomycin); and a linear isothermal (SPIA) amplification process using a composite primer, a second primer, and strand displacement to generate multiple copies of the sequence complementary to the target RNA is used to amplify the marked cDNA (Figure 2b).

[0033] Figure 3 is a diagrammatic representation of an embodiment of the invention in which the first primer comprises a promoter sequence as a marker, and the first primer extension product is formed on an RNA target in the presence of a DNA-dependent DNA polymerase (actinomycin). Transcription (IVT) with an RNA polymerase is used to generate multiple copies of the sequence complementary to the target RNA.

[0034] Figure 4a, 4b, and 4c are diagrammatic representations an embodiment of the invention in which PCR is used to amplify the marked cDNA. In Figure 4a, the first strand primer extension product is produced in the presence of a DNA-dependent DNA polymerase (actinomycin). The marked cDNA can be amplified with one primer linear PCR (4b), or with two primer exponential PCR (4c).

IDT Exhibit 1020.2238

DETAILED DESCRIPTION OF THE INVENTION

General description

[0035] The invention provides novel and improved methods, compositions, and kits for amplification of a target RNA sequence or polynucleotide sequence complementary to an RNA sequence of interest from samples comprising total nucleic acids (i.e. RNA and DNA molecules). The invention relates to the generation of a marked primer extension product comprising a sequence complementary to a target RNA using reverse transcriptase in the presence of a compound that inhibits DNA-dependent DNA polymerase activity, and subsequent amplification of a sequence homologous with or complementary to the target RNA using an amplification where only marked sequences are amplified.

[0036] In one aspect of the invention, the marked primer extension product is created using a tailed primer comprising a 3'-portion which is complementary to the target RNA and a 5'-portion that comprises a specific (e.g. unique or universal) sequence that is not complementary to the target RNA. In this aspect, the specific sequence acts as the marker, becoming incorporated into the sequence of the first primer extension product (e.g. first strand cDNA) at the 5'-end. During this step, the DNA dependent DNA polymerase inhibitor, e.g. actinomycin, inhibits or prevents the creation of marked first primer extension product from the DNA in the sample, thus only first primer extension product representing the RNA in the sample is produced. A second primer can then be used to form a second primer extension product (e.g. second strand cDNA) which will comprise a sequence homologous with the target RNA and will also be marked. It will be marked at its 3'-end with a sequence which comprises a sequence that is complementary to the specific sequence of the tailed primer. Generally, the DNA dependent DNA polymerase inhibitor, e.g. actinomycin, is removed or inactivated, for example by an enzyme, in order to allow the formation of second primer extension product using the first primer extension product, which is typically DNA, as template. This process creates a double stranded cDNA comprising the first primer extension product and second primer extension product wherein the primer extension products are marked at their 5' and 3'-ends respectively with specific sequences. Amplification can then be performed on these single stranded or double stranded species by amplification methods which amplify only the species containing the specific sequence. Amplification methods such as PCR, SPIA, and transcription amplification can be employed. The process described herein of marking only species with sequences related to the target RNA and amplifying only the marked species allows for the specific amplification of RNA targets in the presence of DNA without having to first isolate the target RNA or degrade the DNA in the sample, for example, by DNase treatment.

[0037] This invention relates to methods, compositions, and kits for producing multiple copies of a specific nucleic acid sequence or "target sequence" which may be present either alone or as a component, large or small, of a homogeneous or heterogeneous mixture of nucleic acids. More specifically, the target sequence is an RNA sequence. The mixture of nucleic acids may be that found in a sample taken for diagnostic testing, for screening of blood products, for food, water, industrial or environmental testing, for research studies, for the preparation of reagents or materials for other processes such as cloning, or for

IDT Exhibit 1020.2239

other purposes. The invention also relates to methods of using amplification products from a target RNA sequence, for example, for creation of cDNA libraries, for gene expression profiling, for whole transcriptome analysis, for amplification of populations of viral RNA and the like. In order that the present invention may be more readily understood, certain terms are first defined. Additional definitions are set forth throughout the detailed description.

Definitions

[0038] A "target sequence," "target nucleic acid," or "target RNA," as used herein, is a polynucleotide comprising a sequence of interest, for which amplification is desired. The target sequence may be known or not known, in terms of its actual sequence. In some instances, the terms "target," "template," and variations thereof, are used interchangeably. The target polynucleotide may be from about 10 to about 200 nucleotides in length. Alternatively, the target polynucleotide may be from about 25 to about 400 nucleotides in length. In some cases, the target polynucleotide may be from about 100 nucleotides in length to about 1kb. In still other cases, the target polynucleotide may be 1kb to 10kb in length or more. In some cases, the target sequence may be from about 10kb to 100kb in length or more. In some cases the target polynucleotide comprises a plurality of polynucleotide sequences comprising a plurality of lengths. In some cases, the target polynucleotide comprises a plurality of polynucleotide sequences such as for example a whole transcriptome, a fraction of a whole transcriptome, or a substantial fraction of a whole transcriptome. In other cases, the plurality of target sequences may comprise a group of related sequences such as for example kinases, proteases, ion channels, transmembrane proteins, cytochrome P450 enzymes, or MHC proteins. In still other cases, the plurality of target sequences may be any group of sequences including but not limited to mRNA, a transcriptome, or set of transcriptomes. In some cases the target sequences may be viral RNA or mixtures thereof. In some cases, the target sequence(s) may be purified from a biological source such as tissue, lysates, blood, urine, hair follicle, saliva, skin, or buccal smear. In other cases, the target sequence may not be purified prior to amplification by the methods of the present invention. In some of the methods of the present invention, the target sequence is RNA and is present in a reaction mixture comprising DNA.

[0039] "Amplification," or "amplifying", as used herein, generally refers to the process of producing multiple copies of a desired sequence. "Multiple copies" mean at least 2 copies. A "copy" does not necessarily mean perfect sequence complementarity or identity to the template sequence. For example, copies can include nucleotide analogs such as deoxyinosine, intentional sequence alterations (such as sequence alterations introduced through a primer comprising a sequence that is hybridizable, but not complementary, to the template, or a non-target sequence introduced through a primer), and/or sequence errors that occur during amplification. "Amplifying" a sequence may generally refer to making copies of a sequence or its complement, with the understanding that, as stated above, copying does not necessarily mean perfect sequence complementarity or identity with respect to the template sequence.

[0040] "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,

IDT Exhibit 1020.2240

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, carbamates, 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 phosphate is replaced by P(O)S("thioate"), P(S)S("dithioate"), "(O)NR2("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.

[0041] A "labeled dNTP," or "labeled rNTP," as used herein, refers, respectively, to a dNTP or rNTP, or analogs thereof, that is directly or indirectly attached with a label. For example, a "labeled" dNTP or rNTP, may be directly labeled with, for example, a dye and/or a detectable moiety, such as a member of a specific binding pair (such as biotin-avidin). A "labeled" dNTP or rNTP, may also be indirectly labeled by its attachment to, for example, a moiety to which a label is/can be attached. A dNTP or rNTP, may comprise a moiety (for example, an amine group) to which a label may be attached following incorporation of the dNTP or rNTP into an extension product. Useful labels in the present invention include fluorescent dyes (e.g., fluorescein isothiocyanate, Texas red, rhodamine, green fluorescent protein and the like), radioisotopes (e.g., ^{3}H, ^{35}S, ^{32}P, ^{33}P, ^{125}I, or ^{14}C), enzymes (e.g., LacZ, horseradish peroxidase alkaline

IDT Exhibit 1020.2241

phosphatase,), digoxigenin, and colorimetric labels such as colloidal gold or colored glass or plastic (e.g., polystyrene, polypropylene, latex, etc.) beads. Various anti-ligands and ligands can be used (as labels themselves or as a means for attaching a label). In the case of a ligand that has a natural anti-ligand, such as biotin, thyroxine and cortisol, the ligand can be used in conjunction with labeled anti-ligands.

[0042] "NTP", as used herein, generally refers to either dNTP, rNTP, or analogs thereof.

[0043] The "type" of dNTP or rNTP, as used herein, refers to the particular base of a nucleotide, namely adenine, cytosine, thymine, uridine, or guanine.

[0044] "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 or 6 nucleotides, generally from about 10 or 12 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 and include the first primer, second primer, composite amplification primer, and propromoter polynucleotide (such as the PTO). The terms "oligonucleotide" and "polynucleotide" are not mutually exclusive. The description above for polynucleotides is equally and fully applicable to oligonucleotides.

[0045] A "primer," as used herein, refers to a nucleotide sequence, generally with a free 3'-OH group, that hybridizes with a template sequence (such as a target RNA, or a primer extension product) and is capable of promoting polymerization of a polynucleotide complementary to the template. A "primer" can be, 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. A primer may contain a non-hybridizing sequence that constitutes a tail on the primer. A primer may still be hybridizing even though its sequences are not completely complementary to the target.

[0046] The primers of the invention are usually oligonucleotide primers. A primer is generally an oligonucleotide that is employed in an extension by a polymerase along a polynucleotide template such as in PCR or single primer isothermal amplification for example. The oligonucleotide primer is often 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%, preferably 90%, more preferably 95%, most preferably 100%, complementarity to a sequence or primer binding site. Hence, "complementary", as used herein, refers to complementary to all or 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 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 olignonucleotide primer hybridizes to,

IDT Exhibit 1020.2242

namely, at least six nucleotides, at least ten nucleotides, at least about 15 nucleotides and generally from about 6 to about 10 or 12 nucleotides to about 200, usually about 20 to about 50 nucleotides. In general, the target polynucleotide is larger than the oligonucleotide primer or primers as described previously. In some cases, the hybridizable sequence of an oligonucleotide primer is a random sequence. Oligonucleotide primers comprising random sequences may be referred to as random primers, as described herein. In 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 may comprise a mixture of primers designed to hybridize to a plurality (e.g. 2, 3, 4, about 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, 25,000 or more) of target sequences. In some cases, the plurality of target sequences may 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.

[0047] Composite Primers are employed in certain embodiments of the invention. Composite primers are primers that are composed of RNA and DNA portions. In some aspects, the composite primer is a tailed composite primer comprising, for example, a 3'-DNA portion and a 5'-RNA portion. In the tailed composite primer, a 3'-portion, all or a portion of which comprises DNA, is complementary to a polynucleotide; and a 5'-portion, all or a portion of which comprises RNA, is not complementary to the polynucleotide and does not hybridize to the polynucleotide under conditions in which the 3'-portion of the tailed composite primer hybridizes to the polynucleotide target. When the tailed composite primer is extended with a DNA polymerase, a primer extension product with a 5'-RNA portion comprising a defined sequence can be created. This primer extension product can then have a second primer anneal to it, which can be extended with a DNA polymerase to create a double stranded product with an RNA/DNA heteroduplex comprising a defined sequence at one end. The RNA portion can be selectively cleaved from the partial heteroduplex to create a double-stranded DNA with a 3'-single-stranded overhang which can be useful for a various aspects of the present invention including allowing for isothermal amplification using a composite amplification primer.

[0048] In other aspects, the composite primer is an amplification composite primer (interchangeably called composite amplification primer). In the amplification composite primer, both the RNA and the DNA portions are generally complementary and hybridize to a sequence in the polynucleotide to be copied or amplified. In some embodiments, a 3'-portion of the amplification composite primer is DNA and a 5'-portion of the composite amplification primer is RNA. The composite amplification primer is designed such that the primer is extended from the 3'-DNA portion to create a primer extension product. The 5'-RNA portion of this primer extension product, in a RNA/DNA heteroduplex is susceptible to cleavage by RNase H, thus freeing a portion of the polynucleotide to the hybridization of an additional composite amplification primer. The extension of the amplification composite primer by a DNA polymerase with strand displacement activity releases the primer extension product from the original primer and creates another copy of the sequence of the polynucleotide. Repeated rounds of primer hybridization, primer

IDT Exhibit 1020.2243

extension with strand displacement DNA synthesis, and RNA cleavage create multiple copies of the sequence of the polynucleotide. Composite primers are described in more detail below.

[0049] A "random primer," as used herein, is a primer that generally 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. A random primer will generally be an oligonucleotide or a population of oligonucleotides comprising a random sequence(s) in which the nucleotides at a given position on the oligonucleotide can be any of the four nucleotides, or any of a selected group of the four nucleotides (for example only three of the four nucleotides, or only two of the four nucleotides). In some cases all of the positions of the oligonucleotide or population of oligonucleotides can be any of the four nucleotides, in other cases, only a portion of the position, for instance a particular region, of the oligonucleotide will comprise positions which can be any of the four bases. In some cases, the portion of the olignonucleotide which comprises positions which can be any of the four basis is about 6, 7, 8, 9, 10, 11, 12, 13, 14, or about 15-20 nucleotides in length. In some cases, a random primer may comprise a tailed primer having a 3'-region that comprises a random sequence and a 5'-region that is a non-hybridizing sequence that comprises a specific, non-random sequence. The 3'-region may also comprise a random sequence in combination with a region that comprises poly-T sequences. The sequence of a random primer (or its complement) may or may not be naturally-occurring, or may or may not be present in a pool of sequences in a sample of interest. The amplification of a plurality of 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 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 will comprise a random sequence, while the 5'-portion of the primer comprises a defined sequence. In some embodiments a 3'-random portion of the primer will comprise DNA, and a 5'-portion defined portion of the primer will comprise RNA, in other embodiments, both the 3' and 5'-portions will comprise DNA. In some embodiments, the 5'-portion will contain a defined sequence and the 3'-portion will comprise a poly-dT sequence that is hybridizable to a multiplicity of RNAs in a sample (such as all mRNA).

[0050] "Protopromoter sequence," and "propromoter sequence," as used herein, refer to a single-stranded DNA sequence region which, in double-stranded form is capable of mediating RNA transcription by DNA-dependent RNA polymerase. In some contexts, "protopromoter sequence," "protopromoter," "propromoter sequence," "propromoter," "promoter sequence," and "promoter" are used interchangeably.

IDT Exhibit 1020.2244

[0051] "Propromoter polynucleotide," as used herein, refers to a polynucleotide comprising a propromoter sequence. An example of a propromoter polynucleotide is a propromoter template oligonucleotide (PTO).

[0052] "Propromoter template oligonucleotide (PTO)"and "promoter template oligonucleotide (PTO)" as used herein, refer to an oligonucleotide that comprises a propromoter sequence and a portion, generally a 3'-portion, that is hybridizable (under a given set of conditions) to the 3'-region of a primer extension product. The propromoter sequence and the hybridizable portion may be the same, distinct or overlapping nucleotides of an oligonucleotide.

[0053] To "inhibit" or "inactivate" is to decrease or reduce an activity, function, and/or amount as compared to a reference.

[0054] A "complex" is an assembly of components. A complex may or may not be stable and may be directly or indirectly detected. For example, as is described herein, given certain components of a reaction, and the type of product(s) of the reaction, existence of a complex can be inferred. For purposes of this invention, a complex is generally an intermediate with respect to the final amplification product(s). An example of a complex is a nucleic acid duplex comprising a first primer extension product and a second primer extension product.

[0055] A "portion" or "region," used interchangeably herein, of a polynucleotide or oligonucleotide is a contiguous sequence of 2 or more bases. In other embodiments, a region or portion is at least about any of 3, 5, 10, 15, 20, 25 or more contiguous nucleotides.

[0056] A region, portion, or sequence which is "adjacent" to another sequence directly abuts that region, portion, or sequence. For example, an RNA portion which is adjacent to a 5'-DNA portion of a composite primer directly abuts that region.

[0057] A "reaction mixture" is an assemblage of components, which, under suitable conditions, react to form a complex (which may be an intermediate) and/or a product(s).

[0058] Conditions that "allow" an event to occur or conditions that are "suitable" for an event to occur, such as hybridization, strand extension, 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 nucleotide sequence, temperature, and buffer conditions. These conditions also depend on what event is desired, such as hybridization, cleavage, strand extension or transcription.

[0059] Sequence "mutation," as used herein, refers to any sequence alteration in a sequence of interest in comparison to a reference sequence. A reference sequence can be a wild type sequence or a sequence to which one wishes to compare a sequence of interest. A sequence mutation includes single nucleotide changes or alterations of more than one nucleotide in a sequence, due to mechanisms such as substitution, transversion, deletion or insertion. Single nucleotide polymorphism (SNP) is also a sequence mutation as used herein.

[0060] "Microarray" and "array," as used interchangeably herein, comprise a surface with an array, preferably ordered array, of putative binding (e.g., by hybridization) sites for a biochemical sample (target) which often has undetermined characteristics. In a preferred embodiment, a microarray refers to an

IDT Exhibit 1020.2245

assembly of distinct polynucleotide or oligonucleotide probes immobilized at defined positions on a substrate. Arrays are formed on substrates fabricated with materials such as paper, glass, plastic (e.g., polypropylene, nylon, polystyrene), polyacrylamide, nitrocellulose, silicon, optical fiber or any other suitable solid or semi-solid support, and configured in a planar (e.g., glass plates, silicon chips) or three-dimensional (e.g., pins, fibers, beads, particles, microtiter wells, capillaries) configuration. Probes forming the arrays may be attached to the substrate by any number of ways including (i) in situ synthesis (e.g., high-density oligonucleotide arrays) using photolithographic techniques (see, Fodor et al., Science (1991), 251:767-773; Pease et al., Proc. Natl. Acad. Sci. U.S.A. (1994), 91:5022-5026; Lockhart et al., Nature Biotechnology (1996), 14:1675; U.S. Pat. Nos. 5,578,832; 5,556,752; and 5,510,270); (ii) spotting/printing at medium to low-density (e.g., cDNA probes) on glass, nylon or nitrocellulose (Schena et al, Science (1995), 270:467-470, DeRisi et al, Nature Genetics (1996), 14:457-460; Shalon et al., Genome Res. (1996), 6:639-645; and Schena et al., Proc. Natl. Acad. Sci. U.S.A. (1995), 93:10539-11286); (iii) by masking (Maskos and Southern, Nuc. Acids. Res. (1992), 20:1679-1684) and (iv) by dot-blotting on a nylon or nitrocellulose hybridization membrane (see, e.g., Sambrook et al., Eds., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., Vol. 1-3, Cold Spring Harbor Laboratory (Cold Spring Harbor, N.Y.)). Probes may also be non-covalently immobilized on the substrate by hybridization to anchors, by means of magnetic beads, or in a fluid phase such as in microtiter wells or capillaries. The probe molecules are generally nucleic acids such as DNA, RNA, PNA, LNA and cDNA but may also include proteins, polypeptides, oligosaccharides, cells, tissues and any permutations thereof which can specifically bind the target molecules.

[0061] The term "3'" generally refers to a region or position in a polynucleotide or oligonucleotide 3' (downstream) from another region or position in the same polynucleotide or oligonucleotide.

[0062] The term "5'" generally refers to a region or position in a polynucleotide or oligonucleotide 5' (upstream) from another region or position in the same polynucleotide or oligonucleotide.

[0063] The term "3'-DNA portion," "3'-DNA region," "3'-RNA portion," and "3'-RNA region," refer to the portion or region of a polynucleotide or oligonucleotide located towards the 3' end of the polynucleotide or oligonucleotide, and may or may not include the 3' most nucleotide(s) or moieties attached to the 3' most nucleotide of the same polynucleotide or oligonucleotide. The 3' most nucleotide(s) can be from about 1 to about 50, from about 5 to about 40, from about 6 to about 30, from about 10 to about 30, or from about 12 to about 30 nucleotides.

[0064] The term "5'-DNA portion," "5'-DNA region," "5'-RNA portion," and "5'-RNA region," refer to the portion or region of a polynucleotide or oligonucleotide located towards the 5' end of the polynucleotide or oligonucleotide, and may or may not include the 5' most nucleotide(s) or moieties attached to the 5' most nucleotide of the same polynucleotide or oligonucleotide. The 5' most nucleotide(s) can be from about 1 to about 50, from about 10 to about 40, or from about 20 to about 40 nucleotides.

[0065] As is well understood by those skilled in the art, a "tail" sequence of a primer is a sequence not hybridizable, or not substantially hybridizable to the target sequence under conditions (e.g. ion

IDT Exhibit 1020.2246

concentration, temperature, buffer composition, primer concentration, template concentration etc.) in which other region(s) or portion(s) of the primer hybridizes to the target.

[0066] "Absent" or "absence" of product, and "lack of detection of product" as used herein includes insignificant, or de minimus levels, generally due to lack of significant accumulation of product due to cycling.

[0067] A "DNA-dependent DNA polymerase" is an enzyme that synthesizes a complementary DNA copy from a DNA template. Examples are DNA polymerase I from E. coli, bacteriophage T7 DNA polymerase, or DNA polymerases from bacteriophages T4, Phi-29, M2, or T5. DNA-dependent DNA polymerases of the present invention may be the naturally occurring enzymes isolated from bacteria or bacteriophages or expressed recombinantly, or may be modified or "evolved" forms which have been engineered to possess certain desirable characteristics, e.g., thermostability, or the ability to recognize or synthesize a DNA strand from various modified templates. All known DNA-dependent DNA polymerases require a complementary primer to initiate synthesis. It is known that under suitable conditions a DNA-dependent DNA polymerase may synthesize a complementary DNA copy from an RNA template. RNA-dependent DNA polymerases (described below) typically also have DNA-dependent DNA polymerase activity.

[0068] A "DNA-dependent RNA polymerase" or "transcriptase" is an enzyme that synthesizes multiple RNA copies from a double-stranded or partially-double-stranded DNA molecule having a promoter sequence that is usually double-stranded. The RNA molecules ("transcripts") are synthesized in the 5'-to-3' direction beginning at a specific position just downstream of the promoter. Examples of transcriptases are the DNA-dependent RNA polymerase from E. coli and bacteriophages T7, T3, and SP6.

[0069] An "RNA-dependent DNA polymerase" or "reverse transcriptase" ("RT") is an enzyme that synthesizes a complementary DNA copy from an RNA template. All known reverse transcriptases also have the ability to make a complementary DNA copy from a DNA template; thus, they are both RNA- and DNA-dependent DNA polymerases. Reverse transcriptases may also have an RNase H activity. Some examples of reverse transcriptases are reverse transcriptase derived from Moloney murine leukemia virus (MMLV-RT), avian myeloblastosis virus, retroviral reverse transcriptase, retrotransposon reverse transcriptase, hepatitis B reverse transcriptase, cauliflower mosaic virus reverse transcriptase, bacterial reverse transcriptase, E. coli DNA polymerase and klenow fragment, and Tth DNA polymerase, A primer is required to initiate synthesis with both RNA and DNA templates. In other examples a DNA dependent DNA polymerase may also comprise an RNA-dependent DNA polymerase such as Klenow polymerase, Bst DNA polymerase and the like.

[0070] In some cases, the term "polymerase chain reaction" ("PCR") refers to the method of K. B. Mullis, e.g., as described in U.S. Pat. Nos. 4,683,195 4,683,202, and 4,965,188 (each hereby incorporated in its entirety by reference) and any other improved method known in the art. PCR is a method for exponentially increasing the concentration of a segment of a target sequence in a mixture of genomic DNA without cloning or purification. This process for amplifying the target sequence typically consists of introducing a large excess of two oligonucleotide primers to the DNA mixture containing the desired target sequence, followed by a precise sequence of thermal cycling in the presence of a DNA polymerase. The

IDT Exhibit 1020.2247

two primers are complementary to their respective strands of the double stranded target sequence. To effect amplification, the mixture is denatured and the primers then annealed to their complementary sequences within the target molecule. Following annealing, the primers are extended with a polymerase so as to form a new pair of complementary strands. The steps of denaturation, primer annealing and polymerase extension can be repeated many times (i. e., denaturation, annealing and extension constitute one "cycle"; there can be numerous "cycles") to obtain a high concentration of an amplified segment of the desired target sequence. The length of the amplified segment of the desired target sequence is determined by the relative positions of the primers with respect to each other, and therefore, this length is a controllable parameter. By virtue of the repeating aspect of the process, the method is referred to as the "polymerase chain reaction". Because the desired amplified segments of the target sequence become the predominant sequences (in terms of concentration) in the mixture, they are said to be "PCR amplified". In some cases, PCR can be performed in multiplex fashion whereby multiple primer pairs are used to amplify a plurality of double stranded target sequences.

[0071] In some cases, the term "PCR" or "linear PCR" is used herein to describe a process, wherein one primer is used to produce amplification of the complement of a single strand in the same manner as above by use of denaturation, annealing and extension cycles. In this case, the amplification is linear rather than exponential due to the use of a single primer. In some cases, linear PCR can be performed in multiplex fashion whereby multiple primers are used to amplify a plurality of single stranded target sequences or their complement.

[0072] As used herein, the term "RT-PCR" refers to the replication and amplification of RNA sequences. In this method, reverse transcription is coupled to PCR, e.g., as described in U.S. Pat. No. 5,322,770, herein incorporated by reference in its entirety. In RT-PCR, the RNA template is converted to cDNA due to the reverse transcriptase activity of an enzyme, and then amplified using the polymerizing activity of the same or a different enzyme. Both thermostable and thermolabile reverse transcriptase and polymerase can be used.

[0073] The term or "real-time PCR", "quantitative PCR" or "Q-PCR" refers to the replication of polynucleotides such that the number of amplification cycles to reach a certain amount of amplified product is determined and used to calculate the amount of starting product. In some cases, Q-PCR is performed in combination with RT-PCR to first generate a DNA template from a target RNA by RT-PCR and then to use Q-PCR on the DNA template to determine the amount of an RNA such as a transcript or an mRNA, or a set of RNAs that are present in a sample. In some cases, methods other than RT-PCR are used to generate the DNA template from the target RNA. In other cases, Q-PCR is used to quantitate the amount of a DNA or set of DNA targets in a sample. In some cases Q-PCR is performed in a continuous fashion; namely, the amount of amplified product is determined while the amplification is still occurring. Methods and apparatus for performing Q-PCR and variations thereof including but not limited to TaqMan assays are well known in the art and are described for example in US Patent Nos. 5,994,056; and 5,210,015.

IDT Exhibit 1020.2248

[0074] The term "in-vitro transcription" is an RNA amplification technique using a promoter complex and DNA-dependent RNA polymerase to generate multiple copies of the RNA target, e.g. as described in US Pat Nos. 5545522, 5716785, and in the Proceedings of the National Academy of Sciences (1990), 87: 1663-1667, each hereby incorporated in its entirety by reference.

[0075] Various aspects of the invention are described in further detail in the following subsections.

Overview and advantages of the invention

[0076] The invention provides novel and improved methods, compositions, and kits for amplification of a target RNA sequence or polynucleotide sequence complementary to an RNA sequence of interest from samples comprising total nucleic acids (i.e. RNA and DNA molecules).

[0077] Global amplification of RNA currently generally involves purification of RNA from DNA in the sample in order to prevent amplification from DNA when present in the sample. Examples of undesired amplification products include amplification products from DNA sequences, when present in the sample, rather than from RNA targets, which may or may not be products of gene expression., including genomic DNA sequence, plasmid DNA, viral DNA, mitochondrial DNA and the like. Many current global RNA amplification methods employ a reverse transcription step. If such a method is carried out in the presence of DNA, the amplified product from the target RNA will be contaminated by amplified product generated from the single stranded or double stranded DNA in the sample. This is because the reverse transcriptase enzyme does not discriminate between the target RNA and the DNA in the sample, resulting in amplification of both RNA and DNA. Thus, the amplification of sequences from the DNA when present in the sample together with RNA will not be readily distinguishable and will therefore compromise the quality of the amplified RNA product and the amplification efficiency from the RNA target sequences. Purification of RNA from DNA, for example by separating or selectively amplifying only the RNA containing poly-A sequences, or by degrading the DNA (e.g. with DNase) in a sample containing RNA, can result in reduced yield and/or reduced RNA quality. It is highly desirable to develop improved amplification methods of target RNA in the presence of DNA or total nucleic acids. This is especially useful when there are minute amounts of sample for analysis.

[0078] An important aspect of the present invention is the minimal production of unwanted products during amplification of a target RNA or sequence complementary to the target RNA while maintaining a high level of specificity and sensitivity. The methods of this invention take advantage of agents which selectively inhibit the activity of DNA-dependent DNA polymerases toward DNA species, a common unwanted amplification template, while allowing RNA-dependent DNA synthesis from RNA targets via reverse transcriptase. Thus, the target RNA species (or a multiplicity of RNAs, e.g. , members of a family of related RNAs) or sequence complementary to the target RNA may be amplified in the presence of DNA with decreased levels unwanted amplification products. These methods are useful in maximizing efficiency as well as allowing amplification directly from mixed cultures that may or may not contain both RNA and DNA in the reaction chamber. The selective amplification of specific nucleic acid sequences is of value in increasing the sensitivity of diagnostic and other detection assays while maintaining specificity;

IDT Exhibit 1020.2249

increasing the sensitivity, convenience, accuracy and reliability of a variety of research procedures; and providing ample supplies of specific amplification products for various purposes. The methods of the invention are especially suitable for the amplification of all RNA, such as mRNA, sequences in a preparation from a biological sample that may or may not contain DNA. Furthermore, the methods of the present invention may be used to amplify an entire pool of RNA sequences (such as mRNA) without amplifying DNA species, which may be used to analyze the gene expression profile in cells, such as the cells in a biological sample of interest.

[0079] Some methods provide for generation of multiple copies of DNA comprising a sequence complementary to a RNA sequence of interest. Other methods provide for generation of multiple copies of an RNA sequence of interest. These methods are suitable for, for example, generation of cDNA libraries, subtractive hybridization probes, sequence analysis such as massively parallel sequencing methods including pyrosequencing, sequencing by synthesis, sequencing by hybridization, single molecule sequencing, nanopore sequencing, and sequencing by ligation, and gene expression profiling, quantitative PCR, and whole transcriptome analysis, SNP analysis, and massarray analysis. They generate single or double stranded DNA or RNA products, which are readily suitable for a variety of uses including expression profiling, e.g., by multiplex analysis, including hybridization methods such as microarray technologies, by quantitative PCR, serial analysis of gene expression (SAGE), and digital PCR, as well as electrophoresis-based technologies such as differential display. The methods are amenable to automation.

Two part process: preparation of marked DNA and amplification of marked DNA

[0080] The target nucleic acid contains the target RNA sequence, or sequences, of interest to be amplified in the presence of DNA. The invention can be seen as having two general parts: (1) Preparation of marked DNA sequences representative of the target RNA, usually in the presence of DNA; and (2) Amplification of the marked DNA sequences. By amplifying only the marked DNA sequences, the invention selectively amplifies the sequences representative of the target RNA. Unmarked sequences, related to the DNA in the original sample are not amplified. The methods herein are applicable to amplifying any target nucleic acid comprising RNA, including, for example, mRNA, ribosomal RNA, and viral RNA.

[0081] The first part involves preparing a uniquely marked DNA copy of the target RNA sequence. As illustrated in Figure 1, which outlines a general example of a method of the invention, the target RNA sequence is hybridized to a first primer, generally comprising a 5'-unique non-hybridizing sequence (A) portion with a 3'-hybridizing sequence (B) portion, in the presence of a DNA-dependent DNA polymerase inhibitor, such as actinomycin, for example. A RNA-dependent DNA polymerase, such as reverse transcriptase, is used to extend this first primer to form the first primer extension product (first strand cDNA) that contains a sequence (Y), which is complementary to the target RNA. These steps produce the first primer extension product, which is now uniquely marked with the non-hybridizing 5' sequence (A) for further amplification in a second part of the method. In addition, second strand nucleic acid synthesis, which may inadvertently occur in this step, is specifically inhibited by the DNA-dependent DNA polymerase inhibitor. The DNA-dependent DNA polymerase inhibitor is then removed or inactivated and

IDT Exhibit 1020.2250

the target sequence is separated from the first primer extension product, for example, with heat and/or chemical means, using an enzyme that cleaves RNA from an RNA/DNA heteroduplex, or using an enzyme that generally cleaves RNA. A second primer comprising sequence (C) is hybridized to the first primer extension product and is extended by DNA-dependent DNA polymerase to form the second primer extension product containing sequence (Y'), and forming a double stranded cDNA complex with a unique sequence (A/A') at one end. The sequence (C) can comprise a specific sequence, and/or a random sequence. The second primer may be supplied externally as illustrated in Figure 1 or generated internally as fragments of the target RNA via of the RNA, for example, by enzymes such as RNase or by heat treatment. The use of internally generated fragments of the target RNA allows for a type of random priming of the first primer extension product without the addition of random primers. As used herein, it is understood that "cDNA" refers to a polynucleotide primer extension product.

[0082] The second part of the method involves amplification of the marked DNA sequence using the unique sequence (A), (A') or (A/A') of the single or double stranded cDNA as a template for amplification of polynucleotide sequences using amplification methods described herein and known in the art. Nucleic acid amplification methods can be used in the invention, for example, to amplify nucleic acids complementary to the RNA target or to generate multiple copies the transcription product of the cDNA. Amplification methods useful for the methods of the invention include single primer linear isothermal amplification (SPIA), linear or exponential PCR, RNA transcription via a propromoter template oligonucleotide (PTO), and other amplification techniques such as transcription mediated amplification (TMA), strand displacement amplification (SDA), and rolling circle amplification (RCA), which are be further described herein.

[0083] In some embodiments, the steps are performed in the order that is described. In other embodiments, one or more steps described in this application may be combined and/or performed in any order, as long as the requisite product(s) are able to be formed, and, as is evident, the invention includes various combinations of the steps described herein. For example, it is understood that the inactivation of the DNA-dependent DNA polymerase inhibitor may be performed before or after cleavage of the target RNA from the RNA/DNA heteroduplex. It is also evident, and will be briefly described below to highlight this point, that the invention encompasses methods in which the initial, or first, step is any of the steps described herein. For example, the methods of the invention do not require that the first step be production of the first primer extension product from the RNA template. Methods of the invention encompass embodiments in which later, "downstream" steps are an initial step.

[0084] It is understood that various other embodiments may be practiced, given the general description provided above. For example, reference to using a composite primer means that any of the composite primers described herein may be used.

Amplification via SPIA

[0085] In some aspects of the invention, the amplification method that is used to amplify the marked DNA is a single primer isothermal amplification using a complex comprising a RNA/DNA partial

IDT Exhibit 1020.2251

heteroduplex as a template. In this method, termed single primer linear isothermal amplification (SPIA), a complex comprising a RNA/DNA partial heteroduplex is a substrate for further amplification as follows: an enzyme (such as RNase H), which cleaves RNA sequence from an RNA/DNA heteroduplex, cleaves the RNA sequence (A) from the partial heteroduplex, leaving a partially double stranded polynucleotide complex comprising an exposed 3' single-stranded DNA sequence. The 3' single-stranded sequence (formed by cleavage of RNA in the complex comprising a RNA/DNA partial heteroduplex) is generally the complement of a composite amplification primer, and thus forms a specific binding site for the composite amplification primer. This composite amplification primer is the third primer used in this invention and also comprises a unique 5'-sequence. Extension of a bound composite amplification 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 single stranded DNA product accumulates. The single stranded DNA product is a copy of the complement of the target RNA (or "antisense" DNA). The cycle repeats with the removal of the unique sequence on the 5'-tail of the primer extension product, exposing the 3'-end of the second primer extension product for another cycle of amplification. This linear amplification is referred to as "SPIA" (for Single Primer Linear Isothermal Amplification), and is described in Kurn et al., U.S. Pat. No. 6,251,639 and 6,692,918.

[0086] Amplification using a complex comprising a RNA/DNA partial heteroduplex as a template for further amplification by SPIA generally occurs under conditions permitting composite primer hybridization, primer extension by a DNA polymerase with strand displacement activity, cleavage of RNA from a RNA/DNA heteroduplex and strand displacement. In so far as the composite amplification primer hybridizes to the 3'-single-stranded portion (of the partially double stranded polynucleotide which is formed by cleaving RNA in the complex comprising a RNA/DNA partial heteroduplex) comprising, generally, the complement of at least a portion of the composite amplification primer sequence, composite primer hybridization may be under conditions permitting specific hybridization.

[0087] In some embodiments, the methods of the invention result in amplification of a multiplicity, a large multiplicity, or a very large multiplicity of target RNA. In some embodiments, essentially all of the target RNA present in the initial sample (e.g., all of the mRNA) is amplified. In other embodiments, at least 1, at least 5, at least 10, at least 20, at least 50, at least 100, at least 200, at least 300, up to at least 10,000 or more distinct sequences (such as other sub-segments of a RNA) 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. Target RNA sequences that are amplified may be present on the same RNA strand or on different RNA strands. It will be understood by those of skill in the art that the global amplification methods of the invention are suitable for amplification of any pool or subset of RNA.

[0088] In some embodiments, the methods of the invention are used to globally amplify a single stranded RNA target or the double stranded DNA that is initially produced from the RNA target using methods described herein. In these cases, the amplification product will generally be a copy of either the target RNA (sense copy) or of the complement to the target RNA (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

IDT Exhibit 1020.2252

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.

[0089] 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).

[0090] 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 heteroduplex is included in the first reaction mixture).

[0091] Although generally only one composite amplification primer is described above, it is further understood that the amplification methods may be performed in the presence of two or more different first and/or second 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 first and/or second composite primers that randomly prime template polynucleotide can be combined.

[0092] In one aspect, the invention provides methods of generating multiple copies of a polynucleotide sequence complementary to an RNA sequence of interest in the presence of DNA, illustrated in Figure 2, as follows: (a) hybridizing a first primer to the target RNA sequence in the presence of a DNA-dependent DNA polymerase inhibitor, e.g. actinomycin, wherein the first primer comprises a 3'-DNA sequence (B) and a 5'-RNA sequence (A); (b) extending the first primer with an RNA-dependent DNA polymerase in the presence of a DNA-dependent DNA polymerase inhibitor, whereby a complex comprising a first primer extension product (interchangeably called "first strand cDNA") and the target RNA is produced, wherein the 5'-end of the first primer extension product is the first primer sequence (A); (c) inactivating and/or removing the DNA-dependent DNA polymerase inhibitor, e.g. actinomycin, using methods described herein and cleavage of the target RNA; (d) extending a second primer hybridized to the first primer extension product at sequence (C') with a DNA-dependent DNA polymerase, wherein a second primer comprises a hybridizing DNA sequence (C), or extending a fragment of the RNA template, which is hybridizing to sequence (C'), whereby a second primer extension product (interchangeably called "second strand cDNA") is produced to form a complex of first and second primer extension products and comprises the sequence (A') complementary to (A). The complex of first and second primer extension products is composed of a RNA/DNA heteroduplex at one end due to the presence of the composite primer in the first primer extension product. As previously described herein, the second primer may be supplied externally, where it hybridizes to a specific sequence on the first strand cDNA, or may be generated internally as fragments of the target RNA via enzymatic cleavage by enzymes such as RNase or by other processes such

IDT Exhibit 1020.2253

as heat treatment. In the case of the latter source of the second primer, the primer will not be directed to a specific sequence on the first strand cDNA.

[0093] Following generation of the double stranded cDNA product with the marked first strand cDNA sequence, the second part in the process, as illustrated in Figure 2b, may involve SPIA amplification which may be performed as follows: (e) cleaving RNA from the first primer in the complex of first and second primer extension products with an enzyme, such as RNaseH, that cleaves RNA from an RNA/DNA heteroduplex such that a third composite amplification primer hybridizes to the second primer extension product at sequence (A'), wherein the composite amplification primer comprises an RNA portion and a 3′ DNA portion; (f) extending the composite amplification primer hybridized to the second primer extension product with a DNA-dependent DNA polymerase to produce another composite primer extension product; whereby the first primer extension product is displaced, and whereby multiple copies of a polynucleotide sequence complementary to the RNA sequence of interest are generated. This amplification step produces the complement containing the sequence (C') and the sequence (Y), which is downstream of (C '), that is complementary to the sequence of the target RNA.

[0094] As is clear to one skilled in the art, reference to production of copies of a RNA sequence of interest or copies of a polynucleotide sequence complementary to a RNA sequence of interest refers to products that may contain, comprise or consist of such sequences. As is evident to one skilled in the art, aspects that refer to combining and incubating the resultant mixture also encompasses method embodiments which comprise incubating the various mixtures (in various combinations and/or subcombinations) so that the desired products are formed.

[0095] For example, in some aspects, the components of the invention may be combined and reacted in a reaction mixture where SPIA is allowed to proceed. The invention provides methods of generating multiple copies of a polynucleotide sequence complementary to an RNA sequence of interest, said method comprising the steps of: combining and reacting: the complex of first and second primer extension products described above; a composite amplification primer (which may or may not be the same as the first primer) that is hybridizable to the second primer extension product, wherein said composite amplification primer comprises an RNA portion and a 3′ DNA portion; a DNA-dependent DNA polymerase; and an agent (such as an enzyme) that cleaves RNA from an RNA/DNA heteroduplex; wherein the mixture is incubated under conditions (which include necessary substrates and buffer conditions) that permit RNA cleavage from the first strand cDNA, composite amplification primer hybridization to the second strand cDNA, composite amplification primer extension, resulting in displacement of the first primer extension product from the complex.

Amplification via transcription

[0096] In another aspect of this invention, the second part of the method is amplification of the marked DNA sequence through transcription to generate a copy of the target RNA sequence. This transcription step can involve the use of a PTO and the unique sequence (A), (A'), or (A/A') of the single or double stranded cDNA, generated using any of the described methods herein. The PTO comprises a 3′-

IDT Exhibit 1020.2254

hybridizing region and a 5'-non-hybridizing (tail) region with the propromoter sequence. In some aspects, the unique sequence (A/A') may be a double stranded promoter, as further described below. As described herein, the PTO may comprise RNA, DNA, or combinations thereof.

[0097] The use of a first primer (first strand cDNA primer) comprising a 3'-hybridizable region and a 5'-tail region comprising a single strand sequence of a promoter for DNA-dependent RNA polymerase has been described before in US Pat Nos. 5545522, 5716785, and Proceedings of the National Academy of Sciences (1990), 87: 1663-1667, each hereby incorporated in its entirety by reference. When first strand synthesis is carried out in the presence of a DNA-dependent DNA polymerase inhibitor, only copies of the target RNA will be generated. The product of the first and second primer extension products generated by this method is a double stranded DNA with a double stranded promoter sequence at one end, which is the substrate for the DNA dependent RNA polymerase, such as for example T7 RNA polymerase, SP6 RNA polymerase and the like.

[0098] For example, in one aspect involving transcription illustrated in Figure 3, the invention provides methods of generating multiple copies of (amplifying) an RNA sequence of interest. The first primer is the single stranded PTO comprising a 3'-hybridizing sequence (B) and 5'-non-hybridizing sequence (A) which serves as a propromoter for DNA-dependent RNA polymerase. In the presence of actinomycin, the first strand cDNA is generated by reverse transcriptase and the resulting first strand cDNA contains the sequence (Y) and 5'-sequence (A). Following inactivation or removal of actinomycin and removal of the target RNA from the complex, a second primer, which contains sequence (C), hybridizes to the 3'-end of the first strand cDNA and is extended using DNA-dependent DNA polymerase. The second strand cDNA that is generated is in a complex with the first strand cDNA and contains the double stranded promoter sequence (A/A'), thus enabling RNA synthesis through transcription via RNA polymerase. Multiple copies of the sequence complementary to the target RNA with sequence (C') and (Y) are generated in this aspect.

[0099] As previously described, the second primer may also be endogeneous in the form of a fragment or fragments of the target RNA sequence which is cleaved following first strand synthesis using methods described herein. In this case, the second primer is not hybridized to a specific sequence on the cDNA but rather to various regions of the sequence (Y) on the first strand cDNA since the fragments are generated from randomly cleaved portions and therefore are hybridized to random locations on the cDNA. In this aspect, multiple copies generated comprise the various sequences (Y_1, Y_2, $Y_{3....}$) that are homologous to various regions of the target RNA.

[00100] In another aspect, a PTO is used as the second primer to form a double stranded complex composed of a promoter sequence for transcription. One embodiment of a PTO comprises a second primer with a 3' hybridizing end and a 5' non-hybridizing end, which contains a propromoter sequence. The 3'-end (e.g. sequence (C) hybridizes to the 3'-end of the first primer extension product (first strand cDNA), which contains a 5' sequence (A) and was formed from the steps described above. Extension of the PTO by DNA polymerase produces a double stranded complex containing a sequence (A/A') at one end, and a propromoter sequence incorporated into the second strand. Following the denature of the double stranded cDNA, a primer with sequence (A) is hybridized to the second extension product and is extended to form a

IDT Exhibit 1020.2255

double stranded product having a double stranded promoter sequence from which transcription occurs using RNA polymerase at one end, and the sequence (A/A') at the other end. The double stranded promoter region can be used to promote transcription by an RNA polymerase. The result of the transcription can be an amplification step that generates multiple copies of a RNA sequence of interest.

[00101] In yet another aspect involving transcription, the invention provides methods of generating multiple copies of (amplifying) a sequence complementary to an RNA sequence of interest comprising: (a) hybridizing a PTO to the 3'-end of a second strand cDNA, wherein the second strand cDNA comprises a 5'-sequence (C) and a 3'-sequence (A') and is generated using any of the methods described herein and wherein the PTO comprises a 3'-hybridizing end with sequence (A) that is generally, but not necessarily, blocked so that it cannot be extended by the DNA polymerase, and a 5'-nonhybridizing propromoter sequence (P); (b) extending the second strand cDNA by DNA polymerase to produce a double stranded promoter, enabling RNA amplification through transcription via DNA-dependent RNA polymerase. Multiple copies of a sequence complementary to the target RNA are generated in this aspect.

Amplification via PCR

[00102] In some aspects of the second part of the invention, amplification methods such as linear or exponential PCR, are utilized to perform the amplification of marked cDNA.

[00103] In some aspects of the invention, exponential amplification of nucleic acids is used in the second step to amplify the marked DNA. These methods often depend on the product catalyzed formation of multiple copies of a nucleic acid 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) (see U.S. Pat. No. 4,683,195 and U.S. Pat. No. 4,683,202). The DNA amplification is based on the activity of the enzyme DNA polymerase, which elongates primer molecules that bind to the template DNA by adding dNTPs and thereby copies the template sequence (Saiki et. al, Science (1988), 239: 487-491). The forward primer binds to the 3'-end of the sense strand of the template, whereas the reverse primer binds to the 3'-end of the reverse strand, hereby defining the starting points of the polymerase reaction and eventually determining the length of the amplified product.

[00104] PCR is an amplification procedure is based on repeated cycles of denaturation, oligonucleotide primer annealing, and primer extension, usually by thermophilic 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.

[00105] 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

IDT Exhibit 1020.2256

amplification is that in the latter the product serves as a template for the formation of more products, whereas in the former process the starting sequence serves as a template for the formation of product but the product is not itself replicated. 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.

[00106] In one aspect, the invention provides methods of generating multiple copies of a polynucleotide sequence complementary to an RNA sequence of interest or an RNA sequence of interest in the presence of DNA, said method comprising the steps of: (a) extending a first primer hybridized to a target RNA with an RNA-dependent DNA polymerase in the presence of a DNA-dependent DNA polymerase inhibitor, wherein the first primer is a composite primer comprising a 5'-non-hybridizing region (tailed primer), whereby a complex comprising a first primer extension product and the target RNA is produced; (b) cleaving the target RNA in the complex of step (a) with an enzyme that cleaves RNA and from an RNA/DNA hybrid; (c) inactivating and/or removing the DNA-dependent DNA polymerase inhibitor as described herein; (d) extending a second primer hybridized to the first primer extension product with a DNA-dependent DNA polymerase, whereby a second primer extension product is produced to form a complex of first and second primer extension products; (e) amplifying each of the first and second strand cDNA using either or both the first and second primers via linear or exponential amplification using techniques described in sections above, such as PCR; whereby multiple copies of a polynucleotide sequence homologous or complementary to the RNA sequence of interest are generated. In some aspects, step (c) may be performed before or in combination with step (b).

[00107] In one embodiment, as illustrated in Figure 4a, the amplification methods of the invention results in generation of DNA products comprising sequences complementary to an RNA sequence(s) of interest is as follows: (a) A first primer having a 5'-sequence (A) that is non-complementary to the target RNA binds to an RNA species in a sample by hybridization of the 3'-sequence portion (B) to form a complex in the presence of a DNA-dependent DNA polymerase inhibitor such as actinomycin; (b) A reverse transcriptase (RT) extends the hybridized first primer along the target RNA strand to which it is hybridized, to form an RNA/DNA heteroduplex. The DNA-dependent DNA polymerase inhibitor prevents extension of product from DNA in the sample. The 5'-end of the first primer extension product has the first primer sequence (A); (c) The target RNA strand of the heteroduplex may be degraded, by methods described herein, to generate a single-stranded first strand cDNA; (d) The DNA-dependent DNA polymerase inhibitor is removed, inhibited, inactivated or sequestered; (d) The second primer with sequence (C) binds to first strand cDNA by hybridization to form a complex.

[00108] Linear or exponential PCR can then be used to amplify only the marked cDNA which is formed. PCR involves, for example, linear PCR (Figure 4b) or exponential PCR (Figure 4c), wherein the cDNA generated in the first step is amplified. The unique non-complementary sequence of the first primer is used in subsequent amplification steps to generate only the desired product.

IDT Exhibit 1020.2257

[00109] In Figure 4b, illustrating linear PCR, the second strand cDNA with the 3'-sequence (A') is hybridized to the primer with sequence (A), which is extended by DNA polymerase to generate multiple copies of DNA that are complementary to the target RNA.

[00110] Exponential PCR is illustrated in Figure 4c, wherein both the first and second strand cDNA are exponentially amplified using exponential PCR. The two cDNA strands are denatured and the two primers with sequence (A) and sequence (C) are used simultaneously to amplify both cDNA strands. A primer with sequence (C) is annealed to the first strand cDNA comprising sequences (C') and (A) and is extended to produce strands containing sequence (C). A primer with sequence (A) is annealed to the second strand cDNA comprising sequences (C) and (A') and is extended to produce strands containing sequence (C'). Exponential PCR can also be carried out with one primer comprising a sequence (A) and the other primer comprising a sequence (Y'), complementary to sequence (Y). It will be understood that the PCR primers need not comprise all of sequence (C) or sequence (Y'), but may only comprise a portion of such sequences. In some cases, multiple PCR primer pairs can be employed in order to obtain multiplex PCR and amplification of multiple sequences of interest.

Other methods of amplification

[00111] The second part of the invention may also amplify the marked DNA sequence using other amplification methods known in the art. The following are descriptions of some examples of amplification methods. It is understood that the invention is not limited to the amplification methods described herein. Any amplification method that can amplify the marked polynucleotides produced in the first part of the invention may be used. In some cases, amplification methods can comprise clonal amplification. Clonal amplification is the generation of one or more unique sets of amplified DNA such that each set is separated physically, or chemically from other sets. Methods of clonal amplification include amplifying DNA on a solid surface such as a bead or beads, an array, or a set of wells. In some cases, clonal amplification may be performed by isolating template polynucleotides in individual microreactors such as for example individual aqueous droplets of a water in oil emulsion. Each amplified template will then be isolated from the other templates to generate clonal populations of amplified product in each droplet. In some cases, the droplets further contain a solid surface such as a bead. In some cases, initially, each bead will carry a single DNA molecule. After amplification (e.g. PCR, or SPIA), each bead will contain thousands to millions of identical copies of the DNA molecule. Other methods of clonal amplification are provided herein.

[00112] One example 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, RNaseH, T7 RNA polymerase, NTPs and dNTPs.

[00113] 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

IDT Exhibit 1020.2258

conducting such an amplification include "midi-variant RNA" (amplifiable hybridization probe), NTPs, and Q-beta-replicase.

[00114] 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.

[00115] Another method for amplifying nucleic acids is 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 Fahy, et al., Patent Application WO 91-US8488 911113).

[00116] Another method for amplification of nucleic acids is Strand Displacement Amplification (SDA) (Westin et al., Nature Biotechnology (2000), 18: 199-202; Walker et al., Nucleic Acids Research (1992), 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.

[00117] Another method for amplification of nucleic acids is Rolling Circle Amplification (RCA) (Lizardi et al., Nature Genetics (1998), 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. The method of generating polonies is one method of clonal amplification.

[00118] 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/18957 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 (1998), 19:225-232), bridge PCR (U.S. Pat. No. 5,641,658), picotiter PCR (Leamon et al., Electrophoresis (2003), 24: 3769), and emulsion PCR (Dressman et al., PNAS (2003), 100: 8817). The methods of the invention provide new methods for generating and using polonies.

IDT Exhibit 1020.2259

[00119] A polony is a contraction of "polymerase colony," a small colony of DNA. Polonies are discrete clonal amplifications of a single DNA molecule, grown in a gel matrix or on a surface. This approach greatly improves the signal-to-noise ratio. Polonies can be generated using several techniques that include for example solid-phase PCR in polyacrylamide gels, bridge PCR on a solid phase, or SPIA amplification on a solid surface. In some cases, growth of clonal copies of DNA on bead surfaces may be considered a "polony" method. The method of generating polonies is one method of clonal amplification.

General techniques

[00120] The practice of the invention will 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).

[00121] Primers, oligonucleotides, and polynucleotides employed in the invention can be generated using standard techniques known in the art.

Components and Reaction Conditions used in the Methods of the Invention

Template Nucleic Acid

[00122] The RNA target is generally a polymeric nucleotide, which in the intact natural state can have about 30 to 5,000,000 or more nucleotides and in an isolated state can have about 20 to 50,000 or more nucleotides, usually about 100 to 20,000 nucleotides, more frequently 500 to 10,000 nucleotides. The RNA target to be amplified includes RNAs from any source and or/species, including human, animals, plants, and microorganisms such as bacteria, yeasts, viruses, viroids, molds, fungi, plants, and fragments thereof. Exemplary RNA targets can be obtained and purified using standard techniques in the art and includes 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, DNA-RNA heteroduplexes and fragments thereof. Amplification of a DNA target (including genomic DNA target) would require initial transcription of the DNA target into RNA form, which can be achieved using methods disclosed in Kurn, U.S. Pat. No. 6,251,639 B1, and by other techniques (such as expression systems) known in the art. Amplification of a RNA/DNA heteroduplex would require denaturation of the duplex to obtain a ssRNA, or denaturation followed by transcription of the DNA strand to obtain an RNA. The target RNA may be only a minor fraction of a complex mixture such as a biological sample and may be obtained from various biological materials by procedures well known in the art.

IDT Exhibit 1020.2260

[00123] The target RNA can be known or unknown and may contain more than one desired or suspected specific nucleic acid sequence of interest, each of which may be the same or different from each other. Therefore, the amplification process is useful not only for producing large amounts of one specific nucleic acid sequence, but also for amplifying simultaneously more than one different specific nucleic acid sequence located on the same or different nucleic acid molecules. The initial step of the amplification of a target RNA sequence is rendering the target single stranded. If the target nucleic acid is double stranded (e.g., RNA/DNA heteroduplex) the initial step could be target denaturation. Denaturation may also be carried out to remove secondary structure present in a RNA target molecule. The denaturation step may be thermal denaturation or any other method known in the art.

First Primer

[00124] The first primer is marked in a manner that when the first primer is used to form a first primer extension product, the first primer extension product (and in some cases subsequent extension products such as the second primer extension product) are marked. Various embodiments of the first primer are used in the methods of the invention are described herein. For the methods described herein, one or more first primers may be used.

[00125] Generally, the first primer comprises a 5'-unique sequence and is generally a tailed primer comprising a non-hybridizing 5'-sequence. The 5'-unique sequence allows the first strand cDNA to be uniquely marked in a sample containing other nucleic acids. In one embodiment the first primer is a primer comprising only DNA. In another embodiment, the first primer is a primer comprising only RNA. In another embodiment, the first primer is a composite primer comprising a combination of DNA and RNA. Extension of the first primer results in a complex of first and second strand cDNAs that comprise a portion with a particular characteristic (e.g., cleavable by an enzyme) or sequence content (for example comprising a promoter for DNA dependent RNA polymerase). In another embodiment, the first primer may comprise a 5'-unique end that can partially hybridize to the target. In another embodiment, the first primer may comprise all hybridizable portions but the 5'-end is uniquely marked with groups that distinguish it from the other nucleic acid samples.

[00126] Generation of primers suitable for extension by polymerization is well known in the art, such as described in PCT Pub. No. WO99/42618 (and references cited therein). The first primer may comprise a combination of RNA and DNA, with the 3'-end nucleotide being a nucleotide suitable for nucleic acid extension. The 3'-end nucleotide can be any nucleotide or analog that when present in a primer, is extendable by a DNA polymerase when hybridized to a polynucleotide template. Generally, the 3'-end nucleotide has a 3'-OH. Suitable primers include those that comprise at least one portion of RNA and at least one portion of DNA. For example, first primers may comprise a 5'-RNA portion and a 3'-DNA portion (in which the RNA portion is adjacent to the 3'-DNA portion); or 5'- and 3'-DNA portions with an intervening RNA portion. Accordingly, in one embodiment, the first primer comprises a 5'-RNA portion and a 3'-DNA portion, wherein the RNA portion is adjacent to the 3'-DNA portion. In another embodiment, the first primer comprises 5'- and 3'-DNA portions with at least one intervening RNA portion

IDT Exhibit 1020.2261

(i.e., an RNA portion between the two DNA portions). In yet another embodiment, the first primer of the invention comprises a 3'-DNA portion and at least one intervening RNA portion (i.e., an RNA portion between DNA portions).

[00127] The first primer extension product comprises a specific sequence complementary to an RNA sequence and a 5′ portion, preferably a 5′ end, which comprises RNA, DNA, or combinations thereof. For example, in some embodiments, the RNA portion of a first primer is 5′ with respect to the 3′ DNA portion. In still other embodiments, the 5′ RNA portion is adjacent to the 3′ DNA portion. In some embodiments, a first primer comprises a 5′ portion (for example, a 5′ RNA portion) that is either defined or universal and non-hybridizable to a target RNA under conditions which the composite primer hybridizes to target RNA. The 3'-DNA portion may be sequence specific and hybridizable to the target sequence or contain multiple sequences, random sequences, partially random sequences, and/or a poly-dT sequences for hybridizing to the Poly-A sequence on the mRNA target.

[00128] The 5'-region of the first primer of the present invention may include a propromoter sequence recognized by an RNA polymerase such as for example T7 RNA polymerase, SP6 RNA polymerase and the like.

[00129] In other embodiments, the first primers comprises a sequence entirely of DNA with a 3'-DNA portion that generally is designed to be hybridizable to a target RNA(s). This 3'-DNA portion may be sequence specific and hybridizable to a specific sequence of the target sequence, random sequences, partially random sequences, and/or a poly-dT sequences for hybridizing to the Poly-A sequence on the mRNA target. One or multiple first primers may be used for the amplification methods of the invention. The remaining portion(s) (such as the 5'-DNA portion) of the first primer generally, but not necessarily, comprises a defined or universal sequence that is not hybridizable to a target RNA (which would constitute a tail when the primer is bound to a target RNA). Thus, and as the description herein makes clear, reference to a primer that hybridizes to a sequence (or hybridization template) encompasses embodiments in which at least a portion of the primer is hybridized, as well as those embodiments in which an entire primer is hybridized.

[00130] In still other embodiments, a plurality of first primers are used for hybridizing to the target RNA. In some embodiments, the first primer that hybridizes to target RNA and the composite amplification primer that hybridizes to second primer extension product are the same. In other embodiments, the first primer that hybridizes to target RNA and the composite amplification primer that hybridizes to second primer extension product are different.

[00131] Insofar as many mRNAs have a unique poly-A 3'-end, amplification initiated from the 3'-end sequence of mRNAs is commonly performed for preparation of cDNA libraries and subsequent sequence analysis for determination of gene expression profiling or other applications. The methods of the invention are similarly suited for preparation of libraries of amplified 3'-portions of mRNAs. Similarly, the methods of the invention are suited for preparation of libraries of whole transcriptome amplification. It is understood that the methods of the invention are useful not only for producing large amounts of one specific nucleic acid sequence, but also for amplifying simultaneously more than one different specific

IDT Exhibit 1020.2262

nucleic acid sequence located on the same or different target nucleic acid molecules. For example, the methods of the invention are useful for amplifying all mRNA in a sample, or for amplifying a multiplicity of specific RNA species (in which case a multiplicity of the first primer each comprising a 3'-portion hybridizable to specific sequences of specific RNA species could be used) or family of RNA species in a sample.

[00132] In some embodiments, the 5'-DNA portion of a first primer includes the 5'-most nucleotide of the primer. In some embodiments, the 5'-RNA portion of a first primer includes the 5'-most nucleotide of the primer. In other embodiments, the 3'-DNA portion of a first primer includes the 3'-most nucleotide of the primer. In other embodiments, the 3'-RNA portion of a first primer includes the 3'-most nucleotide of the primer. In other embodiments, the 3'-DNA portion is adjacent to the 5'-RNA portion and includes the 3'-most nucleotide of the primer (and the 5'-RNA portion includes the 5'-most nucleotide of the primer).

Second Primer

[00133] The first primer extension product is used to generate additional DNA or RNA strands containing sequence related, either through homology or complementarity, to the original target RNA sequence. The primer extension product may be used as a single strand template. Alternatively, the primer extension product may be used as a double stranded template by creation of a second strand via use of a DNA-dependent DNA polymerase. In order to amplify the primer extension product, secondary primers may be used that bind to the primer extension product for use in the synthesis of the second cDNA strand, complementary to the cDNA. The second primer is complementary to the cDNA and comprises DNA, RNA fragments, tailed RNA/DNA, tailed DNA/DNA, or mixtures thereof. The second primer further comprises specific sequences, multiple sequences, or random sequences. Various embodiments of the second primer used in the methods of the invention are described herein. For the methods described herein, one or more second primers can be used.

[00134] In one embodiment, the second primer is an oligonucleotide that is supplied externally. The second primer in the methods of the invention (which primes generation of the second strand cDNA) comprises a sequence (which may or may not be the whole of the primer) that is hybridizable (under a given set of conditions) to a first strand cDNA at a site on the first strand cDNA such that the second strand cDNA would be homologous to the RNA sequence of interest. In another embodiment, the second primer is a primer comprising DNA (in some embodiments, consisting of DNA).

[00135] To achieve hybridization to a first strand cDNA (which, as is well known and understood in the art, depends on other factors such as, for example, ionic strength and temperature), the sequence of the second primer that is hybridizable to the first strand cDNA is preferably of at least about 60%, more preferably at least about 75%, even more preferably at least about 90%, and most preferably at least about 95% complementarity to the first strand cDNA. Generation of primers suitable for extension by polymerization is well known in the art.

[00136] In other embodiments, the second primer comprises one or more fragments of target RNA hybridized to the first primer extension product, said one or more fragments generated by cleaving RNA in

IDT Exhibit 1020.2263

the complex of target RNA and first primer extension product with an enzyme that cleaves RNA from an RNA/DNA heteroduplex. In some embodiments, this second primer comprises one or more fragments of the target RNA from cleavage due to heat, chemicals, and enzymes such as RNase. For example, the target RNA in the initial complex comprising the target RNA and first primer extension product is cleaved with an enzyme (such as RNaseH), or by heating the reaction mixture to cleave the RNA target, such that at least one fragment of the template RNA remains hybridized to the composite primer extension product. In this embodiment of the invention, one (or more) template RNA fragment(s) serves as a second primer in the manner described above, to generate a fragment extension product which has the same function as the second primer extension product in the amplification methods described above.

[00137] Suitable second primers and second primer RNA fragments in the methods of the invention are long enough such that it does not dissociate from the first strand cDNA, preferably from about 3 to about 30, more preferably from about 5 or 6 to about 25, even more preferably from about 10 to about 20, and most preferably from about 12 to about 20, nucleotides in length.

[00138] In yet another embodiment, the second primer is a random primer. In other embodiments, the second primer comprises a portion (for example, a 3′ portion) that comprises a random sequence (i.e., a sequence designed to be hybridizable (under a given set of conditions) to one or more sequences in the sample). In some embodiments, the second primer comprises a 5′ portion that is non-hybridizable to a first primer extension product (tailed primer). In some embodiments, the 5'-region of the second primer further comprises defined or universal sequences. In some embodiments, the second primer comprises only a sequence that is hybridizable to a first primer extension product (primer with no tail).

[00139] In embodiments involving transcription, the second primer may comprise propromoter sequence (also referred to herein as "propromoter polynucleotide") that is capable of hybridizing to displaced first strand cDNA. Hybridization of the propromoter polynucleotide to a displaced primer extension product and extension of the 3'-end of the displaced first primer extension product (if there is an overhang) results in a double stranded promoter region that drives transcription (via DNA-dependent RNA polymerase) to produce sense RNA products. This transcription approach is described herein and in Kurn et al., U.S. Pat. No. 6,251,639.

[00140] Promoter oligonucleotides of the present invention can also be modified to prevent the synthesis of DNA there from. In one embodiment, the promoter oligonucleotides comprise a blocking moiety attached at their 3'-termini to prevent primer extension in the presence of a polymerase. Indeed, according to the present invention, at least about 80% of the oligonucleotides present in the amplification reaction which comprise a promoter further comprise a 3'-blocking moiety. In further embodiments, at least about 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% of the oligonucleotides provided to the amplification reaction which comprise a promoter are further modified to comprise a 3'-blocking moiety. In a specific embodiment, any oligonucleotide used in an amplification reaction of the present invention which comprises a promoter sequence must further comprise a 3'-terminus blocking moiety.

[00141] In a preferred embodiment, the propromoter sequence is located in the 5'-portion of the oligonucleotide and the hybridizing sequence is located in the 3'-portion of the oligonucleotide. In one

IDT Exhibit 1020.2264

embodiment, and most typically, the promoter and hybridizing sequences are different sequences. In another embodiment, the promoter and hybridizing sequences overlap in sequence identity. In yet another embodiment, the promoter and hybridizing sequences are the same sequence, and thus are in the same location on the PTO. In the embodiments wherein hybridization of the PTO to the primer extension product results in a duplex comprising an overhang (the 5'-end of the PTO that does not hybridize to the displaced primer extension product, typically comprising all or part of the propromoter sequence), DNA polymerase fills in the overhang to create a double stranded promoter capable of effecting transcription by a suitable RNA polymerase.

[00142] In another embodiment, the PTO comprises a sequence that is 5' to the propromoter sequence, i.e., the PTO comprises additional nucleotides (which may or may not be transcriptional regulatory sequences) located 5' to the propromoter sequence. Generally, but not necessarily, the sequence is not hybridizable (under a given set of conditions) to the primer extension product.

Composite Amplification Primers

[00143] Composite amplification primers (or amplification composite 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.

[00144] 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 a ribonuclease when hybridized to the DNA 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.

[00145] 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 a preferred embodiment, hybridization of the 3'-DNA portion of the composite primer to its complementary sequence in the second strand cDNA is favored over the hybridization of the homologous sequence in the 5'-end of the displaced strand to the second strand cDNA.

IDT Exhibit 1020.2265

DNA-dependent DNA polymerase inhibitors

[00146] The methods of the current invention inhibit formation of unwanted amplification products through the use of DNA-dependent DNA polymerase inhibitors during generation of the first primer extension product. DNA-dependent DNA polymerase inhibitors prevent enzymatic activity that generates primer extension products along DNA present in the sample with the target RNA. This method is a novel approach to amplification of polynucleotides of or DNA complementary to the target RNA in samples comprising, or which may comprise, DNA targets. Traditionally, target RNA would be purified from the DNA in the starting material (e.g. cell lysates), resulting in decreased quantity and quality of the RNA due to artifacts and losses introduced during the purification steps. Failure to isolate the RNA would result in random amplification of nucleic acid segments found in DNA as well as the RNA target or targets. The methods of the present invention allow for amplification of select RNA species, or the total RNA species, in the sample by combining the process of specifically reverse transcribing an RNA sequence in the presence of a DNA-dependent DNA polymerase inhibitor, removing the inhibitor, and allowing amplification of the uniquely marked cDNA, resulting in specific amplification of the target RNA sequences or of the sequences complementary to the target RNA sequence or sequences.

[00147] Furthermore, certain DNA-dependent DNA polymerase inhibitors are known to be effective in preventing other replication artifacts such as premature synthesis of the second strand cDNA and therefore allow amplification of only the first primer extension product. Actinomycin is an example of this type of DNA-dependent DNA polymerase inhibitor and is useful in the synthesis of single stranded first strand cDNA while inhibiting second strand synthesis.

[00148] There are several known DNA-dependent DNA polymerase inhibitors known in the art. For example, Actinomycin-D acts as a DNA-dependent DNA polymerase inhibitor by binding to DNA and preventing initiation of replication (Guy and Taylor (1978) PNAS 75:6088-92.) Other examples of DNA-dependent DNA polymerase inhibitors include, but are not limited to, actinomycin (dactinomycin), alpha-amanitin, aphidicolin (Cozad and Warner (1982) Gamete Research 6:155-60; Gonzcol and Plotkin (1985) Arch Virology 84:129-34; Haraguchi et al. (1983) Nucleic Acids Research 11:1197-1209), BP5, novobiocin (Schneck and Staudenbauer (1977) Nuc Acids Res 4:2057-64), rifampicin, rifamycin (Frolova et al. (1977) Nuc Acids Res 4:523-8), sulfoquinovosylmonoacylglycerol, sulfoquinovosyldiacylglycerol (Ohta et al. (2000) Mutat Res 467:139-52; Ohta et al. (1999) Biol Pharm Bull 22:111-16), ursane, oleanane triterpenoids, ursolic acid, oleanolic acid (Deng et al. (1999) J Nat Prod 62:1624-6), mikanolide, dihydromikanolide (U.S. Pat. No. 6,767,561), dehydroaltenusin (Mizushina et al. (2000) J Biol Chem 275:33957-61), catapol (Pungitore et al. (2004) J Nat Prod 67:357-61), taxinine, cephalomanninine (Oshige et al. (2004) Bioorganic and Medicinal Chem 12:2597-601), dipeptide alcohols (Kato et al. (2005) Int J Mol Med 16:653-9), corylifolin; bakuchiol; resveratrol; Neobavaisoflavone; daidzein; bakuchicin (Sun et al. (1998) J Nat Prod 61:362-6), levodopa, dopamine (Wick (1980) Cancer Research 40:1414-8), anacardic acid and oleic acid (Chen et al. (1998) Chem Comm 24:2769-70).

[00149] The DNA-dependent DNA polymerase inhibitors of the present invention are added in an amount effective to inhibit DNA-dependent DNA polymerase activity. As a result, replication of DNA, when

IDT Exhibit 1020.2266

present in the sample in combination with RNA target or targets, is inhibited. The amount of a DNA-dependent DNA polymerase inhibitor that should be added to achieve the desired inhibition is well known and understood in the art. For example, the DNA-dependent DNA polymerase inhibitor Actinomycin D may be added in a concentration from about 0.1 μg/ml to about 100 μg/ml. In a preferred embodiment, 50% of DNA replication is inhibited by addition of a DNA-dependent DNA polymerase inhibitor. In another preferred embodiment, 60% of DNA replication is inhibited by addition of a DNA-dependent DNA polymerase inhibitor. In yet other preferred embodiments, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% of DNA replication is inhibited by addition of a DNA-dependent DNA polymerase inhibitor.

[00150] Combinations of DNA-dependent DNA polymerase inhibitors may be used to achieve the desired inhibition. Use of combinations of DNA-dependent DNA polymerase inhibitors may be desirable to minimize unwanted effects on the reaction conditions by decreasing the concentration of each inhibitor in the reaction mixture. Alternatively, the combination of inhibitors may be able to achieve a greater amount of inhibition, such as by inhibiting DNA-dependent DNA polymerase activity at various stages.

[00151] It is known that several of these DNA-dependent DNA polymerase inhibitors also act to inhibit DNA-dependent RNA polymerase. Thus, by selecting the appropriate inhibitors, the methods of the present invention may also be useful for inhibiting transcription. This is especially useful in reaction mixtures that contain DNA-dependent RNA polymerases, such as extracts from whole cell lysates.

[00152] Following generation of a marked DNA sequence complementary to the target RNA, the DNA-dependent DNA polymerase inhibitor is removed from the reaction conditions or inhibited, inactivated, or sequestered from functioning to allow for the synthesis of a second strand cDNA and amplification of the primer extension product. Techniques for removal of a DNA-dependent DNA polymerase inhibitor from a reaction condition are well known in the art. In some cases, the RNA-dependent DNA polymerase is also removed. These methods include binding of the nucleic acid components in the reaction mixture to a solid surface such as beads, separating the solid surfaces and removing the soluble portion of the reaction mixture which comprises the DNA-dependent DNA polymerase inhibitor, resuspending the beads to recover the bound nucleic acid components, and proceeding with the cleaned reaction mixture as described herein. Bead separation may be carried out by, for example, centrifugation, magnetic separation of magnetic beads, and the like. Other solid surfaces which bind nucleic acid are also useful for carrying out the methods of the invention. DNA-dependent DNA polymerase inhibitors may be inhibited from functioning through binding of an inhibitor (e.g. an antibody directed to the DNA-dependent DNA polymerase inhibitor). In addition, inactivation of the inhibitor may be carried out by an enzyme. For example, enzymatic inactivation of actinomycin D has been described previously (see Hou and Perlman, JBC (1970), 245: 1289-1295).

IDT Exhibit 1020.2267

Polynucleotide Comprising a Propromoter and a Region which Hybridizes to a Primer Extension Product

[00153] Some embodiments involving amplification of an RNA sequence of interest, methods of the invention employ a propromoter polynucleotide comprising a propromoter and a region which hybridizes to a primer extension product. In some embodiments, the propromoter polynucleotide is provided as a PTO, as described in greater detail below.

Propromoter Template Oligonucleotide

[00154] In some embodiments, RNA amplification methods employ a promoter sequence, provided by a propromoter template oligonucleotide (PTO), for in-vitro transcription by DNA-dependent RNA polymerase (see US Pat Nos. 5545522, 5716785, and PNAS (1990), 87: 1663-1667, each hereby incorporated in its entirety by reference).

[00155] A PTO for use in the methods and compositions of the invention is a single-stranded polynucleotide, generally DNA, comprising a propromoter sequence that is designed for formation of a double-stranded promoter of an RNA polymerase, and a portion capable of hybridizing to the 3'-end of a primer extension product. In an embodiment of the invention, the portion capable of hybridizing to the 3'-end of a primer extension product comprises a sequence which is hybridizable to a defined end sequence of the second primer extension product (and thus, subsequently the 3'-end of the single stranded DNA products). In another embodiment, the portion capable of hybridizing to the 3'-end of a primer extension product comprises a random sequence. In another embodiment, the portion capable of hybridizing to the 3'-end of a primer extension product comprises a sequence, which is hybridizable to sequences found at the 3'-end of a multiplicity of first strand cDNAs.

[00156] Promoter sequences that allow transcription of a template DNA are known in the art and have been discussed above. Preferably, the promoter sequence is selected to provide optimal transcriptional activity of the particular RNA polymerase used. Criteria for such selection, i.e., a particular promoter sequence particularly favored by a particular RNA polymerase, is also known in the art. For example, the sequences of the promoters for transcription by T7 DNA dependent RNA polymerase and SP6 are known in the art. The promoter sequence can be from a prokaryotic or eukaryotic source.

[00157] In some embodiments, the PTO comprises an intervening sequence between a propromoter sequence and a portion capable of hybridizing to the 3'-end of the primer extension product. Suitable length of the intervening sequence can be empirically determined, and can be at least about 1, 2, 4, 6, 8, 10, 12, 15 nucleotides. Suitable sequence identity of the intervening sequence can also be empirically determined, and the sequence is designed to preferably, but not necessarily, enhance degree of amplification as compared to omission of the sequence. In one embodiment, the intervening sequence is a sequence that is designed to provide for enhanced, or more optimal, transcription by the RNA polymerase used. Generally, the sequence is not related (i.e., it does not substantially hybridize) to the target nucleic acid. More optimal transcription occurs when transcriptional activity of the polymerase from a promoter that is operatively linked to said sequence is greater than from a promoter that is not so linked. The

IDT Exhibit 1020.2268

sequence requirements for optimal transcription are generally known in the art as previously described for various DNA dependent RNA polymerases, such as in U.S. Pat. Nos. 5,766,849 and 5,654,142, and can also be empirically determined.

[00158] In one embodiment, the PTO cannot function efficiently as a primer for nucleic acid extension. Techniques for blocking the primer function of the PTO include any that prevent addition of nucleotides to the 3'-end of the PTO by a DNA polymerase. Such techniques are known in the art, including, for example, substitution or modification of the 3'-hydroxyl group, or incorporation of a modified nucleotide, such as a dideoxynucleotide, in the 3'-most position of the PTO that is not capable of anchoring addition of nucleotides by a DNA polymerase. It is possible to block the 3'-end using a label, or a small molecule which is a member of a specific binding pair, such as biotin. It is also possible to render the 3'-end non-extendable by addition of nucleotides which cannot hybridize to a primer extension product, either due to non-complementarity or due to structural modifications which do not support hydrogen bonding. In other embodiments, the PTO is not blocked.

[00159] The length of the portion of the PTO that hybridizes to a primer extension product of interest is preferably from about 5 to about 50 nucleotides, more preferably from about 10 to about 40 nucleotides, even more preferably from about 15 to about 35 nucleotides, and most preferably from about 20 to 30 nucleotides. In some embodiments, the hybridizing portion is at least about any of the following: 3, 5, 10, 15, 20; and less than about any of the following: 30, 40, 50, 60. The complementarity of the hybridizing portion is preferably at least about 25%, more preferably at least about 50%, even more preferably at least about 75%, and most preferably at least about 90%, to its intended binding sequence on the primer extension product of interest.

DNA Polymerase, an Agent Capable of Cleaving an RNA-DNA heteroduplex, and RNA Polymerase for SPIA

[00160] The isothermal amplification methods of the invention employ the following enzymes: an RNA-dependent DNA polymerase, a DNA-dependent DNA polymerase, and an agent capable of cleaving a RNA strand of a RNA/DNA heteroduplex (for example, a ribonuclease such as RNase H). Some examples of the enzymes which may be used in the methods and compositions are described herein. One or more of these activities may be found and used in a single enzyme. For example, RNase H activity may be supplied by an RNA-dependent DNA polymerase (such as reverse transcriptase) or may be provided in a separate enzyme. Similarly, a DNA polymerase may comprise both RNA-dependent and DNA-dependent DNA polymerase enzyme activities, or separate enzymes may be used. A DNA-dependent DNA polymerase and an enzyme that cleaves RNA may also be the same enzyme. A DNA-dependent DNA polymerase, an RNA-dependent DNA polymerase, and the enzyme that cleaves RNA can also be the same enzyme. Another RNase useful for carrying out the methods of the invention is RNase I, which cleaves single-stranded RNA.

[00161] Reverse transcriptases useful for this method may or may not have RNase H activity. The selective RNase may be provided as an RNase H activity of a reverse transcriptase, or may be provided as

IDT Exhibit 1020.2269

a separate enzyme, e.g., as an E. coli RNase H or a T. thermophilus RNase H, or derivatives thereof. Other enzymes which selectively degrade RNA present in an RNA:DNA duplex may also be used. Many reverse transcriptases, such as those from avian myeloblastosis 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 reverse transcriptase which lacks or has reduced levels RNase H activity. Reverse transcriptases devoid of RNase H or with reduced levels 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. In these cases, the addition of an RNase H from other sources, such as that isolated from E. coli, can be employed for the formation of the double stranded cDNA. The RNA-dependent DNA polymerase activity and DNA-dependent DNA polymerase activity may be provided by the same enzyme (for example, Bst polymerase), or these activities may be provided in separate enzymes. DNA polymerases with strand displacement activity are also useful.

[00162] Enzymes for use in the compositions, methods and kits of the present invention include any enzyme having reverse transcriptase activity. Such enzymes include, but are not limited to, retroviral reverse transcriptase, retrotransposon reverse transcriptase, hepatitis B reverse transcriptase, cauliflower mosaic virus reverse transcriptase, bacterial reverse transcriptase, E. coli DNA polymerase and klenow fragment, Tth DNA polymerase, Taq DNA polymerase (Saiki, R. K., et al., Science 239:487-491 (1988); U.S. Pat. Nos. 4,889,818 and 4,965,188), Tne DNA polymerase (WO 96/10640), Tma DNA polymerase (U.S. Pat. No. 5,374,553), C. Therm DNA polymerase from Carboxydothermus hydrogenoformans (EP0921196A1, Roche, Pleasanton, Calif., Cat. No. 2016338), ThermoScript (Invitrogen, Carsbad, Calif. Cat. No. 11731-015) and mutants, fragments, variants or derivatives thereof. As will be understood by one of ordinary skill in the art, modified reverse transcriptases may be obtained by recombinant or genetic engineering techniques that are routine and well-known in the art. Mutant reverse transcriptases can, for example, be obtained by mutating the gene or genes encoding the reverse transcriptase of interest by site-directed or random mutagenesis. Such mutations may include point mutations, deletion mutations and insertional mutations. Preferably, one or more point mutations (e.g., substitution of one or more amino acids with one or more different amino acids) are used to construct mutant reverse transcriptases of the invention. Fragments of reverse transcriptases may be obtained by deletion mutation by recombinant techniques that are routine and well-known in the art, or by enzymatic digestion of the reverse transcriptase(s) of interest using any of a number of well-known proteolytic enzymes. Mutant DNA polymerase containing reverse transcriptase activity can also be used as described in U.S. patent application Ser. No. 10/435,766, incorporated by reference in its entirety.

[00163] One aspect of the invention is the formation of a complex comprising an RNA/DNA partial heteroduplex. This process generally utilizes the enzymatic activities of an RNA-dependent DNA polymerase, and a DNA-dependent DNA polymerase. Generally, RNA in the RNA/DNA partial heteroduplex is cleaved by an agent (such as an enzyme, such as a ribonuclease) capable of cleaving RNA from an RNA/DNA hybrid, generating a 3'-single stranded portion with sequences that are complementary

IDT Exhibit 1020.2270

to RNA in a composite primer (and thus forming a binding site for a composite primer). The DNA primer extension product may also be separated (at least partially) from the RNA template using an enzyme which degrades the RNA template. Suitable enzymes, i.e., "selective RNAses," are those which act on the RNA strand of an RNA:DNA complex, and include enzymes which comprise an RNase H activity.

[00164] RNA-dependent DNA polymerases for use in the methods and compositions of the invention are capable of effecting extension of a primer according to the methods of the invention. Accordingly, a preferred RNA-dependent DNA polymerase is 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 and compositions of the invention include reverse transcriptase and, for example, a DNA polymerase that possesses both DNA-dependent and RNA-dependent DNA polymerase activity, such as Bst DNA polymerase. The RNA-dependent DNA polymerase may comprise a RNase H enzyme activity, or separate enzymes may be used.

[00165] DNA-dependent DNA polymerases for use in the methods and compositions of the invention are capable of effecting extension of the composite primer according to the methods of the invention. Accordingly, a preferred polymerase is one that is capable of extending a nucleic acid primer along a nucleic acid template that is comprised at least predominantly of deoxynucleotides. The formation of the complex comprising the RNA/DNA partial heteroduplex can be carried out by a DNA polymerase which comprises both RNA-dependent DNA polymerase and DNA-dependent DNA polymerase activities (such as Bst DNA polymerase, or a reverse transcriptase). In one embodiment ,amplification of an RNA sequence according to methods of the invention involves the use of a DNA polymerase that is able to displace a nucleic acid strand from the polynucleotide to which the displaced strand is bound, and, generally, the more strand displacement capability the polymerase exhibits (i.e., compared to other polymerases which do not have as much strand displacement capability) is preferable. Preferably, the DNA polymerase has high affinity for binding at the 3'-end of an oligonucleotide hybridized to a nucleic acid strand. Preferably, the DNA polymerase does not possess substantial nicking activity. Generally, the DNA polymerase preferably has little or no 5' to 3' exonuclease activity so as to minimize degradation of primer, or primer extension products. Generally, this exonuclease activity is dependent on factors such as pH, salt concentration, whether the template is double stranded or single stranded, and so forth, all of which are familiar to one skilled in the art. Mutant DNA polymerases in which the 5' to 3' exonuclease activity has been deleted, are known in the art and are suitable for the amplification methods described herein. Mutant DNA polymerases which lack both 5' to 3' nuclease and 3' to 5' nuclease activities have also been described, for example, exo $^{-/-}$-Klenow DNA polymerase. It is preferred that the DNA polymerase displaces primer extension products from the template nucleic acid in at least about 25%, more preferably at least about 50%, even more preferably at least about 75%, and most preferably at least about 90%, of the incidence of contact between the polymerase and the 5' end of the primer extension product. In some embodiments, the use of thermostable DNA polymerases with strand displacement activity is preferred. Such polymerases are known in the art, such as described in U.S. Pat. No. 5,744,312 (and references cited therein). Generally, the DNA polymerase has little to no proofreading activity

IDT Exhibit 1020.2271

[00166] Suitable DNA polymerases for use in the methods and compositions of the invention include those disclosed in U.S. Pat. Nos. 5,648,211 and 5,744,312, which include exo - Vent (New England Biolabs), exo - Deep Vent (New England Biolabs), Bst (BioRad), exo - Pfu (Stratagene), Bca (Panvera), sequencing grade Taq (Promega), exo -/- Klenow DNA polymerase, and thermostable DNA polymerases from thermoanaerobacter thermohydrosulfuricus.

[00167] One of the ribonuclease for use in the methods and compositions of the invention is capable of cleaving ribonucleotides in an RNA/DNA hybrid or heteroduplex. Preferably, the ribonuclease cleaves ribonucleotides in an RNA/DNA heteroduplex regardless of the identity and type of nucleotides adjacent to the ribonucleotide to be cleaved. It is preferred that the ribonuclease cleaves independent of sequence identity. Examples of suitable ribonucleases for the methods and compositions of the invention are well known in the art, including ribonuclease H (RNase H), e.g., Hybridase.

[00168] As is well known in the art, DNA-dependent DNA polymerase activity, RNA-dependent DNA polymerase activity, and the ability to cleave RNA from a RNA/DNA heteroduplex may be present in different enzymes, or two or more activities may be present in the same enzyme. Accordingly, in some embodiments, the same enzyme comprises RNA-dependent DNA polymerase activity and cleaves RNA from an RNA/DNA heteroduplex. In some embodiments, the same enzyme comprises DNA-dependent DNA polymerase activity and cleaves RNA from an RNA/DNA heteroduplex. In some embodiments, the same enzyme comprises DNA-dependent DNA polymerase activity, RNA-dependent DNA polymerase activity and cleaves RNA from an RNA/DNA heteroduplex. In some embodiments, different enzymes comprise RNA-dependent DNA polymerase activity and DNA-dependent DNA polymerase activity. In some embodiments, different enzymes comprise RNA -dependent DNA polymerase activity and cleave RNA from an RNA/DNA heteroduplex. In some embodiments, different enzymes comprise DNA-dependent DNA polymerase activity and cleave RNA from an RNA/DNA heteroduplex. In other embodiments the RNA targets are degraded by the use of other RNases such as RNase 1, for example, for removing the target RNA following extension of all DNA first primer along the RNA targets by transcription, or a combination of RNase H and other RNases.

[00169] The DNA-dependent RNA polymerases for use in the methods and compositions of the invention are known in the art. Either eukaryotic or prokaryotic polymerases may be used. Examples include T7, T3 and SP6 RNA polymerases. Generally, the RNA polymerase selected is capable of transcribing from the promoter sequence provided by the propromoter polynucleotides as described herein. Generally, the RNA polymerase is a DNA-dependent polymerase, which is preferably capable of transcribing from a single stranded DNA template so long as the promoter region is double stranded.

[00170] In general, the enzymes used in the methods and compositions of the invention should not produce substantial degradation of the nucleic acid components of said methods and compositions.

Reaction Conditions and Detection

[00171] Appropriate reaction media and conditions for carrying out the methods of the invention are those that permit nucleic acid amplification according to the methods of the invention. Such media and

IDT Exhibit 1020.2272

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. WO 99/42618. 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^{2+} or Mn^{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 strand binding proteins (for e.g., 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. Preferably, the amplification reactions (particularly, primer extension other than the first and second strand cDNA synthesis steps, and strand displacement) are performed isothermally, which avoids the cumbersome thermocycling process. The amplification reaction is carried out at a temperature that permits hybridization of the oligonucleotides (primer and/or PTO) of the invention to the template polynucleotide and primer extension products, 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. In some embodiments that include RNA transcription, the temperature for the transcription steps is lower than the temperature(s) for the preceding steps. In these embodiments, 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.

[00172] Nucleotide and/or nucleotide analogs, such as deoxyribonucleoside triphosphates, that can be employed for synthesis of the primer extension products in the methods of the invention are provided in the amount of from preferably about 50 to about 2500 μM, more preferably about 100 to about 2000 μM, even more preferably about 200 to about 1700 μM, and most preferably about 250 to about 1500 μ M. In some embodiments, a nucleotide or nucleotide analog whose presence in the primer extension strand enhances displacement of the strand (for example, by causing base pairing that is weaker than conventional AT, CG base pairing) is included. Such nucleotide or nucleotide analogs include deoxyinosine and other modified bases, all of which are known in the art. Nucleotides and/or analogs, such as ribonucleoside triphosphates, that can be employed for synthesis of the RNA transcripts in the methods of the invention

IDT Exhibit 1020.2273

are provided in the amount of from preferably about 0.25 to about 6 mM, more preferably about 0.5 to about 5 mM, even more preferably about 0.75 to about 4 mM, and most preferably about 1 to about 3 mM.

[00173] The oligonucleotide components of the amplification reactions of the invention are generally in excess of the number of target nucleic acid sequence to be amplified. They can be provided at about or at least about any of the following: 10, 10^2, 10^4, 10^6, 10^8, 10^{10}, 10^{12} times the amount of target nucleic acid. Composite primers and PTO can each be provided at about or at least about any of the following concentrations: 50 nM, 100 nM, 500 nM, 1000 nM, 2500 nM, 5000 nM.

[00174] In the methods of the invention, the steps may be carried out in the order listed or, in some cases, may be carried out in a different order. In some methods a later step depends on the formation of a product from an earlier step, in which case such steps must be carried out in the order listed. One of ordinary skill in the art will understand which steps should be carried out in the order listed, and which steps can be carried out in a different order.

[00175] In some embodiment, the foregoing components are added simultaneously at the initiation of the amplification process. In another embodiment, components are added in any order prior to or after appropriate time points during the amplification process, as required and/or permitted by the amplification reaction. Such time points, some of which are noted below, can be readily identified by a person of skill in the art. The enzymes used for nucleic acid amplification according to the methods of the invention can be added to the reaction mixture either prior to the target nucleic acid denaturation step, following the denaturation step, or following hybridization of the primer to the target RNA, as determined by their thermal stability and/or other considerations known to the person of skill in the art. The first strand cDNA (first primer extension product) and the second strand cDNA (second primer extension product) synthesis reactions can be performed consecutively, followed by the amplification steps (binding by another primer, primer extension and strand displacement). In these embodiments, the reaction conditions and components may be varied between the different reactions.

[00176] In some embodiments, the amplification process can be stopped at various time points, and resumed at a later time. Said time points can be readily identified by a person of skill in the art. One time point is at the end of first strand cDNA synthesis. Another time point is at the end of second strand cDNA synthesis. Methods for stopping the reactions are known in the art, including, for example, cooling the reaction mixture to a temperature that inhibits enzyme activity or heating the reaction mixture to a temperature that destroys an enzyme. Methods for resuming the reactions are also known in the art, including, for example, raising the temperature of the reaction mixture to a temperature that permits enzyme activity or replenishing a destroyed (depleted) enzyme. In some embodiments, one or more of the components of the reactions is replenished prior to, at, or following the resumption of the reactions. For example, it may be necessary to replenish the composite primer prior to beginning the linear amplification reaction if the same composite primer is being used. Alternatively, the reaction can be allowed to proceed (i.e., from start to finish) without interruption.

[00177] In some embodiments the reaction can be allowed to proceed without purification of intermediate complexes, for example, to remove primer. Products can be purified at various time points, which can be

IDT Exhibit 1020.2274

readily identified by a person of skill in the art. One time point is at the end of first strand cDNA synthesis. Another time point is at the end of second strand cDNA synthesis. In some embodiments, the removal of primers and/or target at the end of a defined step by enzymes with appropriate nuclease activities are also useful, for example, cleavage of the RNA portion of free composite tailed primer prior to the isothermal amplification step by treatment with RNase 1. In some embodiment, routine purification of the complex of first and second cDNA results in slightly higher amplification efficiency in subsequent linear amplification steps.

[00178] The detection of the amplification product is indicative of the presence of the target sequence. Quantitative analysis is also feasible. Direct and indirect detection methods (including quantitation) are well known in the art. For example, by comparing the amount of product amplified from a test sample containing an unknown amount of a polynucleotide containing a target sequence to the product of amplification of a reference sample that has a known quantity of a polynucleotide that contains the target sequence, the amount of target sequence in the test sample can be determined. The amplification methods of the invention can also be extended to analysis of sequence alterations and sequencing of the target nucleic acid. Further, detection could be effected by, for example, examination of translation products from RNA amplification products. The global amplification by the methods of the invention and/or the amplification of selected targets, when present in the sample, are useful for various methods which enable highly parallel nucleic acid interrogations.

Uses of the invention

[00179] The invention provides methods and compositions to selectively amplify RNA in the presence of DNA. The amplification products from this invention can be used for a variety of purposes including, but not limited to, those described herein. For example, purposes using the amplification products of the invention include methods of sequencing, quantitating, or detecting a specific transcript or group of transcripts, genotyping (such as nucleic acid mutation detection, determining the presence or absence of a sequence of interest, analyzing splice variants), preparing an immobilized nucleic acid (which can be a nucleic acid immobilized on a microarray, a surface, a bead, or one or more wells in a plate), and characterizing nucleic acids using methods such as probe hybridization (Southern and Northern blotting, hybridization to probe arrays). The analysis of cellular RNA can be used in methods including gene expression profiling, Q-PCR, digital PCR, SAGE, methods of subtractive hybridization and the preparation of probes for subtractive hybridization, and methods of preparing libraries (which can be cDNA and/or differential hybridization libraries). Amplification products generated may also be further manipulated to generate labeled targets (with or without fragmentation) for analysis using microarrays and the like. Amplification products generated may also be used for analysis by sequencing. In some cases, sequencing may be performed by massively parallel sequencing methods including but not limited to nanopore sequencing, and massively parallel sequencing by various commercial platforms such as a Genome Sequencer from 454/Roche, the SOLiD system from Applied Biosystems, the Polonator by Dover

IDT Exhibit 1020.2275

Systems, the Helicos Genetic Analysis Platform by Helicos Biosciences, the Genome Analyzer by Illumina, and Single-Molecule Realtime (SMRT) technology from Pacific Biosystems.

[00180] The invention allows selective RNA amplification from a sample comprising total nucleic acids (i.e. DNA and RNA), thereby providing a means for amplification and analysis, detection, and quantification of target RNA sequences directly from cell or sample lysate without purification of the RNA targets. The sample may be from the same cells or tissues or from a combination of cells and other nucleic acid components, or other biological samples such as environmental samples. For example, samples may include viral RNA in a host sample, viral RNA in an environmental sample, or cell-free nucleic acids such as in blood or other bodily fluids. The samples may be single cell samples, samples of a few cells, or samples with small numbers of cells, for example 10s to hundreds of cells. The methods could be used, for example for performing gene expression analysis on single to small numbers of cells using the cell lysate directly as the sample with little or no further purification. In some cases the cell lysate could be treated to remove or denature all enzymes and/or proteins in the sample resulting in a total nucleic acid sample that can be used directly for the methods described herein.

[00181] The amplification methods of the invention are useful, for example, for pre-amplification and target preparation for analysis such as for example sequencing, Q-PCR, digital PCR, SAGE, or massively parallel analysis such as massively parallel sequencing, high density PCR arrays, microarray hybridization, and massively parallel Q-PCR of an RNA sequence or set of RNA sequences of interest. The sequencing process is carried out by amplifying a target RNA containing the sequence of interest by any of the methods described herein. Addition of nucleotides during primer extension is analyzed using methods known in the art, for example, incorporation of a terminator nucleotide or sequencing by synthesis (e.g. pyrosequencing). For sequencing methods based on methods described herein wherein the amplified product is DNA, the appropriate dNTPs, or analogs thereof, which may be labeled or unlabeled, are used. For sequencing methods based on methods described herein wherein the amplified product is RNA, the appropriate rNTPs, or analogs thereof, which may be labeled or unlabeled, may be used. Analysis of the reaction products for elucidation of sequence information can be carried out using any of various methods known in the art for sequencing.

[00182] In some embodiments, the methods are useful for pre-amplification and/or preparing target polynucleotide(s) for sequencing by synthesis using the methods commercialized by 454/Roche Lifesciences including but not limited to the methods and apparatus described in *Margulies et al., Nature (2005) 437:376-380 (2005)*; and US Patent Nos. 7,244,559; 7,335,762; 7,211,390; 7,244,567; 7,264,929; 7,323,305. In general, target polynucleotides may be amplified by the methods of the present invention, immobilized onto beads, and compartmentalized in a water-in-oil PCR emulsion. In some cases, alternative amplification methods may be employed in the water-in-oil emulsion such as any of the methods provided herein. When the emulsion is broken, amplified fragments remain bound to the beads. The beads may be enriched and loaded into wells of a fiber optic slide so that there is approximately 1 bead in each well. Nucleotides are flowed across and into the wells in a fixed order in the presence of polymerase, sulfhydrolase, and luciferase. Addition of nucleotides complementary to the target strand

IDT Exhibit 1020.2276

results in a chemiluminescent signal that is recorded such as by a camera. The combination of signal intensity and positional information generated across the plate allows software to determine the DNA sequence.

[00183] In some embodiments, the methods are useful for pre-amplification and/or preparing target polynucleotide(s) for sequencing by the methods commercialized by Helicos BioSciences Corporation (Cambridge, MA) as described in U.S. application Ser. No. 11/167046, and US Patent Nos. 7501245; 7491498; 7,276,720; and in U.S. Patent Application Publication Nos. US20090061439; US20080087826; US20060286566; US20060024711; US20060024678; US20080213770; and US20080103058. In general, target polynucleotides may be amplified by the methods of the present invention, and then immobilized onto a flow-cell surface. Polymerase and labeled nucleotides are then flowed over the immobilized DNA. After fluorescently labeled nucleotides are incorporated into the DNA strands by a DNA polymerase, the surface is illuminated with a laser, and an image is captured and processed to record single molecule incorporation events to produce sequence data.

[00184] In some embodiments, the methods are useful for pre-amplification and/or preparing target polynucleotide(s) for sequencing by the sequencing by ligation methods commercialized by Applied Biosystems (e.g. SOLiD sequencing). In general, target polynucleotides may be amplified by the methods of the present invention, and then incorporated into a water in oil emulsion along with polystyrene beads and amplified by for example PCR. In some cases, alternative amplification methods may 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. 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.

[00185] In some embodiments, the methods are useful for pre-amplification and/or preparing target polynucleotide(s) for sequencing by the sequencing by ligation methods commercialized by Dover Systems. Generally, target polynucleotides may be pre-amplified and/or prepared by the methods of the present invention. The target polynucleotides may then amplified in an emulsion in the presence of magnetic beads. Any amplification methods may be employed in the water-in-oil emulsion such as any of the methods provided herein. The resulting beads with immobilized clonal polynucleotide polonies are

IDT Exhibit 1020.2277

then purified by magnetic separation, capped, amine functionalized, and covalently immobilized in a series of flow cells. Then, a series of anchor primers are flowed through the cells, where they hybridize to the synthetic oligonucleotide sequences at the 3' or 5' end of proximal or distal genomic DNA tags. Once an anchor primer is hybridized, a mixture of fully degenerate nonanucleotides ('nonamers') and T4 DNA ligase is flowed into the cell; each of the nonamer mixture's four components being labeled with one of four fluorophores, which correspond to the base type at the query position. The fluorophore-tagged nonamers selectively ligate onto the anchor primer, providing a fluorescent signal that identifies the corresponding base on the genomic DNA tag. Once the probes are ligated, fluorescently labeling the beads, the array is imaged in four colors. Each bead on the array will fluoresce in only one of the four images, indicating whether there is an A, C, G, or T at the position being queried. After imaging, the array of annealed primer-fluorescent probe complex, as well as residual enzyme, are chemically striped using guanidine HCl and sodium hydroxide. After each cycle of base reads at a given position have been completed, and the primer-fluorescent probe complex has been stripped, the anchor primer is replaced, and a new mixture of fluorescently tagged nonamers is introduced, for which the query position is shifted one base further into the genomic DNA tag. Seven bases are queried in this fashion, with the sequence performed from the 5' end of the proximal tag, followed by six base reads with a different anchor primer from the 3' end of the proximal tag, for a total of 13 base pair reads for this tag. This sequence is then repeated for the 5' and 3' ends of the distal tag, resulting in another 13 base pair reads. The ultimate result is a read length of 26 bases (thirteen from each of the paired tags).

[00186] In some embodiments, the methods are useful for pre-amplification and/or preparing target polynucleotide(s) for sequencing by the methods commercialized by Illumina as described US Patent Nos. 5,750,341; 6,306,597; and 5,969,119. In general, target polynucleotides may be pre-amplified and/or prepared by the methods of the present invention to produce amplified nucleic acid sequences tagged at one (e.g. (A)/(A') or both ends (e.g. (A)/(A') and (C)/(C')). In some cases, single stranded nucleic acid tagged at one or both ends is amplified by the methods of the present invention (e.g. by SPIA or linear PCR). The resulting nucleic acid is then denatured and the single stranded amplified polynucleotides are randomly attached to the inside surface of flow-cell channels. Unlabeled nucleotides are added to initiate solid-phase bridge amplification to produce dense clusters of double-stranded DNA. The initiate the first base sequencing cycle, four labeled reversible terminators, primers, and DNA polymerase are added. After laser excitation, florescence form each cluster on the flow cell is imaged. The identity of the first base for each cluster is then recorded. Cycles of sequencing are performed to determine the fragment sequence one base at a time. For paired-end sequencing, such as for example, when the polynucleotides are labeled at both ends by the methods of the present invention, sequencing templates can be regenerated in-situ so that the opposite end of the fragment can also be sequenced.

[00187] In some embodiments, the methods are useful for pre-amplification and/or preparing target polynucleotide(s) for sequencing by the methods commercialized by Pacific Biosciences as described in US Patent Nos. 7462452; 7476504; 7405281; 7170050; 7462468; 7476503; 7315019; 7302146; 7313308; and US Application Publication Nos. US20090029385; US20090068655; US20090024331; and

IDT Exhibit 1020.2278

US20080206764. In general, target polynucleotides may be pre-amplified and/or prepared by the methods of the present invention. The amplified polynucleotides may then be immobilized in zero mode waveguide arrays. Polymerase and labeled nucleotides are added in a reaction mixture, and nucleotide incorporations are visualized via fluorescent labels attached to the terminal phosphate groups of the nucleotides. The fluorescent labels are clipped off as part of the nucleotide incorporation. In some cases, circular templates are utilized to enable multiple reads on a single molecule.

[00188] The DNA or RNA amplification products generated according to the methods of the invention are also suitable for analysis for the detection of any alteration (interchangeably called "mutations") in the target nucleic acid sequence, as compared to a reference nucleic acid sequence which is identical to the target nucleic acid sequence other than the sequence alteration. The sequence alterations may be sequence alterations present in the genomic sequence or may be sequence alterations which are not reflected in the genomic DNA sequences, for example, alterations due to post transcriptional alterations, and/or mRNA processing, including splice variants. Sequence alterations include deletion, substitution, insertion and/or transversion of one or more nucleotide.

[00189] The invention provides methods to characterize (for example, detect presence or absence of and/or quantify) an RNA sequence of interest by generating DNA or RNA products using amplification methods of the invention, and analyzing the products by detection/quantification methods such as those based on array technologies or solution phase technologies or combinations thereof. These amplified products may be labeled or unlabeled. The detection and/or quantitation of specific nucleic acid sequences is an important technique for identifying and classifying microorganisms, diagnosing infectious diseases, detecting and characterizing genetic abnormalities, identifying genetic changes associated with cancer, studying genetic susceptibility to disease, and measuring response to various types of treatment. Such procedures are also useful in detecting and quantitating microorganisms in foodstuffs, water, industrial and environmental samples, seed stocks, and other types of material where the presence of specific microorganisms may need to be monitored. Other applications are found in the forensic sciences, anthropology, archaeology, and biology where measurement of the relatedness of nucleic acid sequences has been used to identify criminal suspects, resolve paternity disputes, construct genealogical and phylogenetic trees, and aid in classifying a variety of life forms.

[00190] A number of methods to detect and/or quantitate nucleic acid sequences are well known in the art. These include hybridization to a labeled probe, and various permutations of the polymerase chain reaction (PCR), coupled with hybridization to a labeled probe (For example, see U.S. Pat. Nos. 4,683,195; 4,683,202; 4,800,159; Mullis et al., Meth. Enzymol. (1987) 155: 335-350; and Murakawa et al., DNA (1988) 7:287-295).

[00191] Various methods for the detection and quantification of gene expression levels have been developed in recent years. For example, microarrays, in which either defined cDNAs or oligonucleotides are immobilized at discrete locations on, for example, solid or semi-solid substrates, or on defined particles, enable the detection and/or quantification of the expression of a multitude of genes in a given specimen.

IDT Exhibit 1020.2279

[00192] The amplification methods of the invention are also suitable for use in determining the levels of gene expression in a sample since the methods described herein are capable of amplifying one or a plurality of 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 mRNAs, in a sample is indicative of the relative expression levels of the various genes (gene expression profile). Thus, determination of the amount of target RNA 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.

[00193] The single-stranded DNA and RNA products of the methods of the invention are useful in preparing libraries, including cDNA libraries and subtractive hybridization libraries. Using the methods of the invention, libraries may be prepared from limited amount of starting material, for example, mRNA extracted from limited amount of tissue or even single cells. Accordingly, in one aspect, the methods of the invention provide preparing a library from the single stranded DNA or RNA products of the invention. In another aspect, the invention provides methods of preparing a library from the double stranded cDNA produced by the methods of the invention comprising two composite primers. Methods for preparing libraries from double stranded cDNA are well known in the art. In still another aspect, the invention provides methods for making a library, said method comprising the preparation of subtractive hybridization probes using any of the methods described herein.

[00194] The methods of the invention offer the ability to efficiently amplify mRNA in the presence of DNA under conditions that provide for high specificity of target amplification and which is generally reflective of the distribution in the input RNA. Thus, the utility of the detection/quantification methods which can be used with the amplification products of the invention, such as those based on the powerful array technology, real time PCR, quantitative TaqMan, quantitative PCR using molecular beacons and the like, should be greatly enhanced.

Compositions and Kits of the Invention

[00195] The invention also provides compositions and kits used in the methods described herein and may be incorporated as parts of kits, e.g., diagnostic kits for clinical or criminal laboratories, or nucleic amplification kits for general laboratory use. The present invention thus includes kits which include some or all of the reagents necessary to carry out the methods of the present invention, e.g., sample preparation reagents, oligonucleotides, binding molecules, stock solutions, nucleotides, polymerase inhibitors, enzymes, positive and negative control target sequences, test tubes or plates, labeling reagents, fragmentation reagents, detection reagents, purification matrices, and an instruction manual.

[00196] The compositions of the invention may be any or any combination of component(s), reaction mixture, and/or intermediate described herein. For example, the invention provides a composition comprising a first primer, reverse transcriptase, and one or more types of DNA-dependent DNA polymerase inhibitors, and combinations thereof. The invention also provides compositions comprising the amplification products described herein. Accordingly, the invention provides a population of DNA or

IDT Exhibit 1020.2280

RNA molecules which are copies or the complement of a target sequence, which are produced by any of the methods described herein. The compositions are generally in a suitable medium, although they can be in lyophilized form. Suitable media include, but are not limited to, aqueous media, such as pure water or buffers.

[00197] In another aspect, the invention provides a population of sense polynucleotide (e.g., DNA) molecules and antisense polynucleotide (e.g., DNA) molecules which are copies and complements of a target sequence, which are produced by any of the methods described herein. The invention also includes compositions and various configurations (such as arrays) of these populations, which may be homogeneous (same sequence) or heterogeneous (different sequence). These populations may be any assembly of sequences obtained from the methods described herein, including those based on mRNA, as well as certain species or classes of mRNA, other RNA species, or viral RNA.

[00198] The invention provides kits for carrying out the methods of the invention. Accordingly, a variety of kits are provided in suitable packaging. The kits may be used for any one or more of the uses described herein, and, accordingly, may contain instructions for any one or more of the following uses for samples comprising total nucleic acids: amplifying an RNA sequence; sequencing of an RNA sequence of interest; detection of sequence mutation based on amplifying an RNA sequence (e.g., genotyping or nucleic acid mutation detection); detection of an RNA sequence alteration including splice variants or fused transcripts; determining presence or absence of a sequence of interest; methods of gene expression analysis and profiling; methods of detection and quantification of RNA targets in samples comprising total nucleic acids; methods of whole transcriptome analysis for the detection, discovery, and/or quantification of sense or antisense transcripts, or coding and non-coding transcripts; methods of preparing amplification products for massively parallel analysis including massively parallel sequencing (e.g. pyrosequencing, sequencing by synthesis, sequencing by hybridization, single molecule sequencing, nanopore sequencing, and sequencing by ligation), high density PCR arrays, microarray hybridization, SAGE, digital PCR, and massively parallel Q-PCR; methods of subtractive hybridization; methods of preparing a subtractive hybridization probe; methods of differential amplification; methods of preparation of libraries (including cDNA and differential expression libraries); methods of preparation of an immobilized nucleic acid (which can be a nucleic acid immobilized on a microarray), and methods of characterizing amplified nucleic acid products generated by the methods of the invention.

[00199] The kits of the invention comprise one or more containers comprising any combination of the components described herein. Kits may include, but are not limited, to the following examples.

[00200] A kit may comprise any of the composite primers described herein with any of the listed DNA-dependent DNA polymerase inhibitors herein. In some embodiments, a kit comprises two or more composite primers, reverse transcriptase, and DNA-dependent DNA polymerase inhibitors, which may or may not be separately packaged. A kit may comprise a composite primer and a polynucleotide comprising a propromoter sequence, which may be a PTO, with DNA-dependent DNA polymerase inhibitors. A kit may further comprise a second primer, which can comprise a random primer. The composite primer may be labeled or unlabeled.

IDT Exhibit 1020.2281

[00201] Kits may also optionally further include any of one or more of the enzymes described herein (for example, RNA-dependent DNA polymerase such as reverse transcriptase, and ribonuclease such as RNase H), as well as deoxynucleoside triphosphates (labeled or unlabeled) and/or ribonucleoside triphosphates (labeled or unlabeled) and/or nucleotide analogs and/or non-canonical nucleotides. Kits may also include one or more suitable buffers as described herein. Kits useful for nucleic acid sequencing may optionally include labeled or unlabeled nucleotide analogs that upon incorporation into a primer extension product or RNA transcript effect termination of nucleotide polymerization. The kits may comprise tailed deoxyoligonucleotides, wherein the tailed 5'-sequence is non-hybridizable to the target RNA and may further comprise a single-stranded DNA sequence of a promoter for DNA-dependent RNA polymerase.

[00202] The kits of the invention may optionally include a set of instructions, generally written instructions, although electronic storage media (e.g., magnetic diskette or optical disk) containing instructions are also acceptable, relating to the use of components of the methods of the invention for the intended nucleic acid amplification, and/or, as appropriate, for using the amplification products for purposes such as nucleic acid sequencing and detection of sequence mutation. The instructions included with the kit generally include information as to reagents (whether included or not in the kit) necessary for practicing the methods of the invention, instructions on how to use the kit, and/or appropriate reaction conditions. For example, kits of the invention can comprise: a composite primer (which can comprise a poly-dT portion and/or can be a random primer), a DNA-dependent DNA polymerase inhibitor(s), and instructions for using the primer with a DNA-dependent DNA polymerase inhibitor to amplify RNA according to methods of the invention. In another example, any of these kits further comprises one or more controls (which can be, for example, RNA template, composite primers, and/or double stranded cDNA complex (comprising first and second strand cDNA) comprising a 3'-single-stranded DNA portion).

[00203] One or more reagents in the kit can be provided as a dry powder, usually lyophilized, including excipients, which on dissolution will provide for a reagent solution having the appropriate concentrations for performing any of the methods described herein. Each component can be packaged in separate containers or some components can be combined in one container where cross-reactivity and shelf life permit. The component(s) of the kit may be packaged in any convenient, appropriate packaging. The components may be packaged separately, or in one or multiple combinations. Where kits are provided for practicing amplification methods of the invention that involve transcription, the RNA polymerase (if included) is preferably provided separately from the components used in the steps prior to the transcription steps.

[00204] The relative amounts of the various components in the kits can be varied widely to provide for concentrations of the reagents that substantially optimize the reactions that need to occur to practice the methods disclosed herein and/or to further optimize the sensitivity of any assay.

[00205] The following Examples are provided to illustrate, but not limit, the invention.

IDT Exhibit 1020.2282

EXAMPLES

Example 1: Amplification of sequences from a target RNA in the presence of DNA using actinomycin to inhibit DNA-dependent DNA polymerase activity

[00206] A RNA target is amplified using a method involving a first primer, comprising composite RNA/DNA portions, wherein the 3'-DNA end of the primers comprises either poly-dT sequence or a random sequence. First strand cDNA synthesis is performed with this first primer in the presence of Actinomycin D, an inhibitor of DNA-dependent DNA polymerase. Removal of Actinomycin D and RNA target from the reaction mixture is performed by binding the nucleic acids to magnetic beads. Second strand cDNA synthesis is then carried out and is followed by amplification, which is performed using a composite amplification primer, DNA polymerase with strand displacement activity, and RNase H. The global RNA amplification was carried out using the WT-Ovation Pico amplification system (NuGEN Technologies Inc. http://www.nugeninc.com/html/03_products5.html) with additional reagents and protocol modifications described below. As described herein, the results of the example demonstrate selective amplification of RNA target from samples comprising mixtures of RNA and DNA.

Materials

[00207] All reagents for the WT-Ovation Pico RNA Amplification System (NuGEN Technologies Inc.) were used in this example. HeLa (S3) total RNA was purchased from Ambion. Human genomic DNA was purchased from Clontech. Random hexamer (N6 DNA oligonucleotide) was purchased from Operon. Actinomycin D was purchased from USB.

General comments

[00208] The WT-Ovation Pico amplification system was used to perform global RNA amplification. For selective amplification of RNA targets from samples comprising genomic DNA, the procedure was modified to include Actinomycin D in the step involving first strand cDNA synthesis. Additional clean up following first strand cDNA synthesis was added to the manufacturer protocol, which result in the removal of Actinomycin D prior to second strand cDNA synthesis. In addition, second strand cDNA synthesis was carried out in the presence of random hexamers (DNA oligonucleotides). All other steps were the same as for the WT-Ovation Pico RNA Amplification System.

First strand cDNA synthesis

[00209] First stand cDNA synthesis is initiated randomly across the length of RNA transcripts using composite first primers containing a 3'-random sequence DNA portion (for example, a random hexamer sequence) and a 5'-non-complementary and non-hybridizable RNA portion. In addition to the composite first primer, another composite first primer, comprising a 3'-poly-DT sequence DNA portion and the same 5'-non-hybridizing RNA portion as the previous primer, is used. In both the WT-Ovation Pico RNA amplification protocol and the modified procedure of the current invention, 1 to 2 µl of total RNA sample

IDT Exhibit 1020.2283

comprising 0.5 to 50 ng total RNA purified from a sample to be analyzed, with or without DNA, are mixed with 1 to 2 µl of a mixture containing the two first primers. The resulting mixture was transferred to a heating block or a thermal cycler and incubated at 65°C for 2 minutes. The mixture is then removed from the heating block or thermocycler and placed on ice.

[00210] A second mixture containing reverse transcriptase and a buffer were added to the first mixture above to a total volume of 10 µl. In samples subjected to the selective RNA amplification, Actinomycin D was added to the mixture to a final concentration of 50 ng/µl. The reactions were carried out by incubation at the following temperatures and durations: 4°C for 1min, 25°C for 10 min, 42°C for 10 min, 70°C for 15 min., 4°C and hold. The reaction mixture is then placed on ice.

Removal of Actinomycin D

[00211] Prior to second strand cDNA synthesis, removal of the DNA-dependent DNA polymerase and Actinomycin D is performed using magnetic beads under conditions that allow binding of the nucleic acids in the mixture to the beads. This clean up step allows separation of the reaction medium containing Actinomycin D from the nucleic acids and was conducted following the manufacturer instructions. To the reaction mixture above, 16 µl of bead suspensions (RNAClean, Agencourt) were added and the mixture was incubated at room temperature for 10 min. The beads were separated using a magnetic plate and the supernatant was discarded. The beads with the bound nucleic acids from the above reaction mixture were washed three times with 70% ethanol and air dried for 10 min at room temperature. This clean up step was performed for samples subjected to the selective RNA amplification according to the method of the invention and is a modified protocol of the WT-Ovation Pico RNA amplification.

Second strand cDNA synthesis

[00212] A mixture of DNA polymerase and buffer was added to the first strand synthesis reaction mixture for synthesis of the second strand cDNA. In the case of the modified protocol for selective RNA amplification according the method of the current invention, the reaction mixture also contained random hexamers (N6 DNA oligonucleotides) at 100 ng/µl. Second strand cDNA synthesis was carried out in a total volume of 20µl. For selective RNA amplification according to the method of the current invention, the beads used for the removal of Actinomycin D, as described above, were left in the second strand cDNA synthesis reaction mixture. Second strand cDNA synthesis reactions were carried out by incubation at the following temperatures for the following durations: 4°C for 1 min, 25°C for 10 min, 50°C for 30 min, 70°C for 5 min, and cooled to 4°C.

Purification of the double stranded cDNA

[00213] To each second strand cDNA reaction mixture, 32 µl of bead suspension (RNAClean® Agencourt) were added and the mixture was incubated for 10 min at room temperature. The beads were separate on a magnetic plate and further processed as described herein.

IDT Exhibit 1020.2284

Amplification

[00214] Amplification was carried out according to the manufacturer's instruction with modification of the total reaction volume from 160 µl to 80 µl (www.nugeninc.com).

[00215] Beads were suspended in the amplification mixture containing 20 µl of composite amplification primer, 40 µl of reaction buffer, and 20 µl enzyme mixture. The amplification reactions were carried out at following temperatures and durations: 40°C for 1 min., 47°C for 60 min., and 95°C for 5 min. The reaction were cooled to 4°C and held at this temperature until purification of the amplification products.

Results

[00216] The selective amplification of RNA from samples comprising low levels of input total HeLa RNA, with or without excess genomic human DNA, employed a modified protocol of the WT-Ovation Pico RNA amplification system, which included Actinomycin D in the first strand cDNA synthesis (to inhibit DNA-dependent DNA polymerase) and removal of Actinomycin D prior to second strand cDNA synthesis. The results from selective RNA amplification from a sample comprising DNA, are shown in Table 1, and includes the yield of amplified cDNA products generated by the non-modified protocol and the modified protocol, which is designed according to the method of the current invention. In control reactions, the selective amplification of RNA is demonstrated by the marked reduction in the yield of amplification products from samples comprising only genomic DNA and no RNA.

Table 1

Input		Yield µg SPIA™ amplified cDNA	
RNA	DNA	WT-Ovation Pico	Selective RNA amplification
1 ng		5.92	5.86
		5.56	5.34
1ng	1ng	5.80	5.02
		4.46	5.65
1 ng	9 ng	5.83	5.76
		5.53	5.68
No RNA	9 ng	4.82	2.66
		5.24	2.31
No RNA	No DNA	1.00	0.74
		0.85	0.76

IDT Exhibit 1020.2285