9. A method according to any preceding claim, wherein the modified nucleic acid precursor is dUTP.

10. A method according to any one of Claims 4-9, wherein the DNA glycosylase is uracil DNA-glycosylase (UDG).

11. A method according to any one of Claims 4-10, wherein the extended nucleic acid is cleaved at the abasic site by means of an enzyme which cleaves at a nucleic acid abasic site.

12. A method according to Claim 11, wherein the enzyme is an AP endonuclease.

13. A method according to any preceding claim, wherein the modified precursor nucleotide partially replaces one of the normal precursor nucleotides.

14. A method according to any one of Claims 1-3, 5-9 and 13, wherein steps i) and ii) continue in a cyclical manner until one of the reagents becomes limiting.

15. A method according to any one of Claims 4-13, wherein steps i)-iii) continue in a cyclical manner until one of the reagents becomes limiting.

16. A method according to any preceding claim, which is carried out under isothermal conditions.

17. A method according to any preceding claim, which results in the accumulation of displaced single stranded downstream fragments of nucleic acid specified by the locations of modified bases in a complementary nucleic acid strand.

18. A method according to any preceding claim for generating multiple copies of discrete single stranded primers downstream of an initiating nucleic acid primer.

19. A method according to Claim 17 or 18, wherein the displaced downstream fragments are extended in a secondary reaction.

20. A method according to any one of Claims 17-19, wherein the displaced downstream fragments are extended on a secondary template nucleic acid.

21. A method according to any one of Claims 17-20, wherein multiple secondary templates are immobilised on a DNA chip.

22. A method according to any preceding claim for use in detection diagnostics.

23. A method according to Claim 22, which is used in the detection of pathogens.

24. A method according to Claim 22, which is used in the detection of the presence or absence of mutations.

25. A method according to Claim 22, which is used in the detection of polymorphisms.

26. A method according to any preceding claim for the quantification of the level of a nucleic acid in a sample.

27. A method according to any preceding claim for use in signal amplification from any nucleic acid that can function as a primer or template.

28. A method according to any preceding claim, for use in DNA computing.

29. A method according to Claim 1, substantially as hereinbefore described and exemplified.

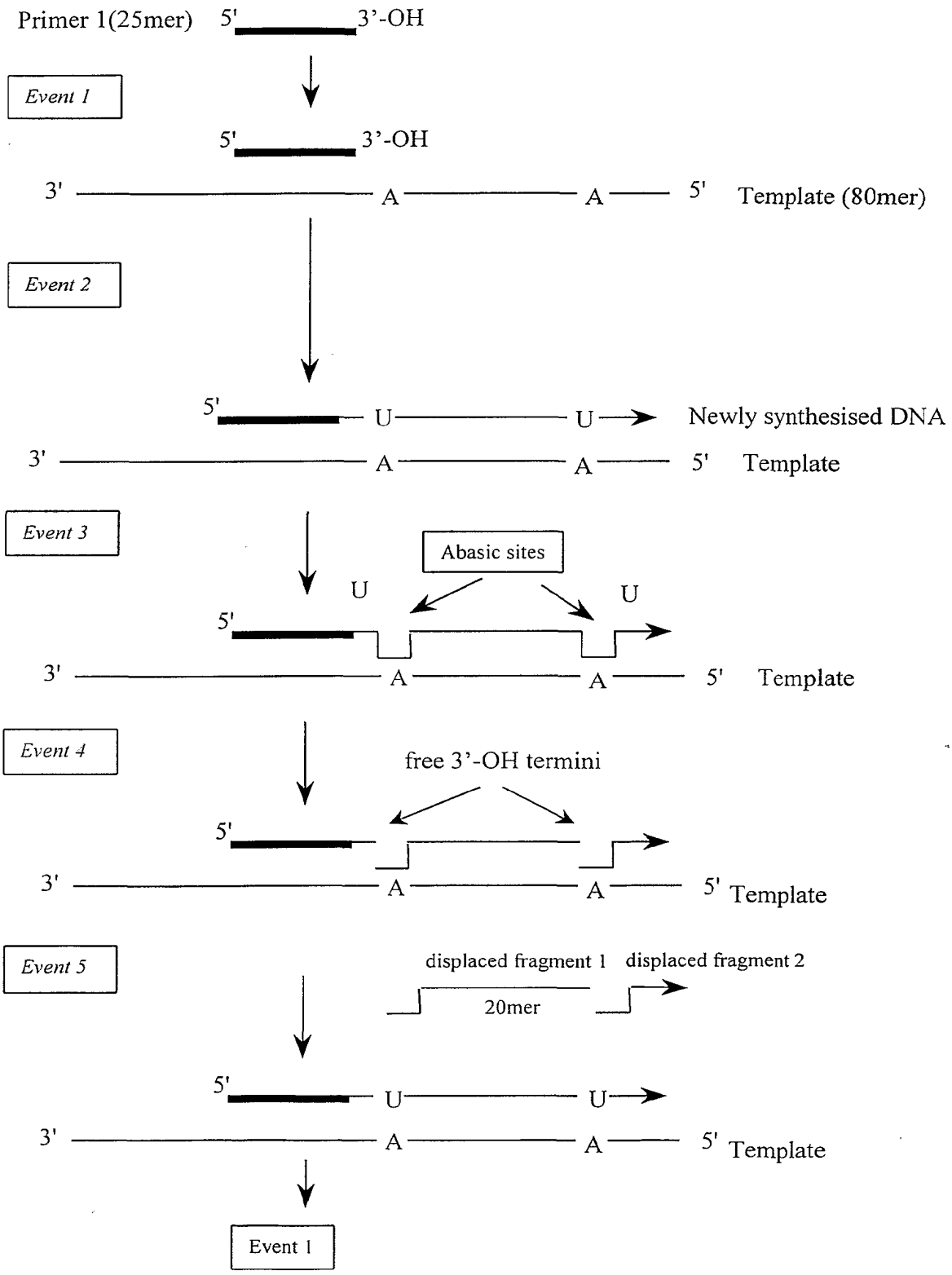

Primer 1(25mer) 5' 3'-OH
Event 1
5' 3'-OH
3' A A 5' Template (80mer)
Event 2
5' U U Newly synthesised DNA
3' A A 5' Template
Event 3
Abasic sites
U U
5'
3' A A 5' Template
Event 4
free 3'-OH termini
5'
3' A A 5' Template
Event 5
displaced fragment 1
displaced fragment 2
20mer
5' U U
3' A A 5' Template
Event 1

FIG. 1

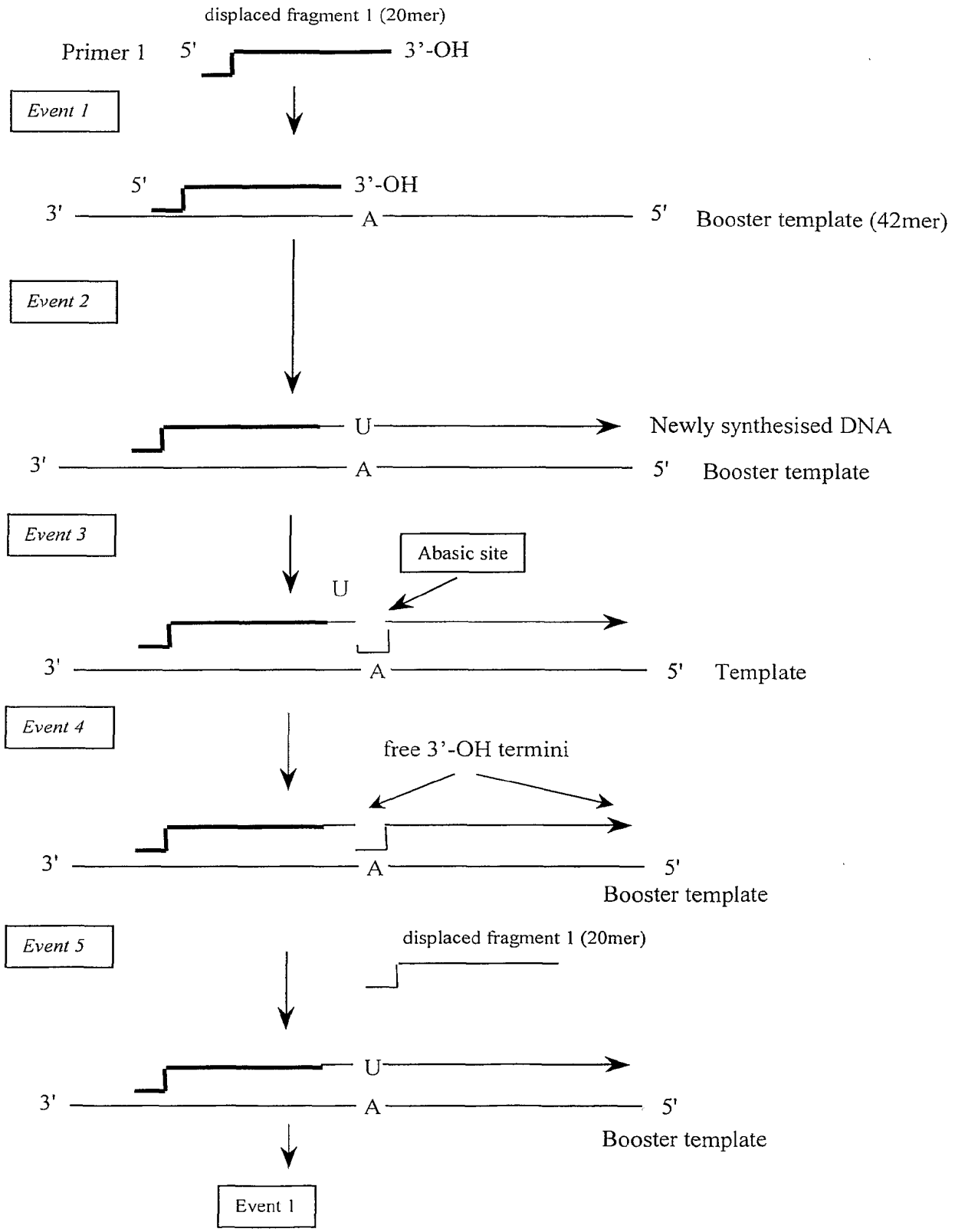

displaced fragment 1 (20mer)
Primer 1
5'
3'-OH
Event 1
5'
3'-OH
3'
A
5'
Booster template (42mer)
Event 2
U
Newly synthesised DNA
3'
A
5'
Booster template
Event 3
Abasic site
U
3'
A
5'
Template
Event 4
free 3'-OH termini
3'
A
5'
Booster template
Event 5
displaced fragment 1 (20mer)
U
3'
A
5'
Booster template
Event 1

FIG. 2

Espacenet

Bibliographic data: WO0009756 (A1) — 2000-02-24

ADAPTER DIRECTED EXPRESSION ANALYSIS

Inventor(s): HUNKAPILLER MICHAEL W; RICHARDS JOHN H + (HUNKAPILLER, MICHAEL, W, ; RICHARDS, JOHN, H)

Applicant(s): PERKIN ELMER CORP [US] + (THE PERKIN-ELMER CORPORATION)

Classification: - international: *C12Q1/68;* (IPC1-7): C12Q1/68
- cooperative: **C12Q1/6809**; **C12Q1/6837**; **C12Q1/6874** more

Application number: WO1999US18422 19990812 Global Dossier

Priority number (s): US19980135381 19980817

Also published as: AU5560199 (A) US6232067 (B1) US6258539 (B1)

Abstract of WO0009756 (A1)

The present invention relates to methods and compositions for simultaneously analyzing multiple different polynucleotides of a polynucleotide composition comprising multiple diverse polynucleotide sequences. The subject methods and compositions may also be applied to analyze or identify single polynucleotides, however, the subject methods and compositions are particularly useful for analyzing large diverse populations of polynucleotides, e.g., cDNA libraries. Most embodiments of the invention involve hybridizing terminus probes (of known base sequence) and internal fragment probes (of known base sequence) at adjacent positions on an adapter-modified restriction fragment generated from polynucleotide for analysis, and subsequently joining the terminus probes and internal fragment probes to each other. The terminus probe hybridizes to bases of restriction endonuclease recognition site present at the terminus of a restriction fragment generated from the polynucleotide for analysis. Internal fragment probes hybridizes to the same strand of the restriction fragment that the terminus probe hybridizes to and hybridizes to the restriction fragment portion of adapter-modified representative restriction fragments. The terminus probes and internal fragment probes may be marked so as to facilitate the simultaneous testing of multiple polynucleotides for the presence of many possible nucleotide base sequences. The identity or expression of a particular polynucleotide of interest may be ascertained (or

IDT Exhibit 1020.8658

at least partially determined) by producing a short identifier sequence derived from the nucleotide base sequence information obtained from (1) the hybridization of a terminus probe and an internal fragment probe, each of known base sequence, at adjacent positions on a polynucleotide of interest, and (2) the recognition site of a restriction endonuclease used to generate the polynucleotide molecule of interest. Multiple identification sequences may be obtained in parallel, thereby permitting the rapid characterization of a large number of diverse polynucleotides. Parallel processing may be achieved by differentially marking terminus probes or internal fragment probes. Parallel processing may be achieved by using ordered arrays of oligonucleotides that are terminus probes.

IDT Exhibit 1020.8659

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12Q 1/68

A1

(11) International Publication Number: **WO 00/09756**

(43) International Publication Date: 24 February 2000 (24.02.00)

(21) International Application Number: PCT/US99/18422

(22) International Filing Date: 12 August 1999 (12.08.99)

(30) Priority Data:
09/135,381 17 August 1998 (17.08.98) US

(71) Applicant: THE PERKIN-ELMER CORPORATION [US/US]; 850 Lincoln Centre Drive, Foster City, CA 94404 (US).

(72) Inventors: HUNKAPILLER, Michael, W.; 1333 Pebble Drive, San Carlos, CA 94070 (US). RICHARDS, John, H.; 667 Deodar Lane, Bradbury, CA 94010 (US).

(74) Agent: BORTNER, Scott, R.; The Perkin-Elmer Corporation, 850 Lincoln Centre Drive, Foster City, CA 94404 (US).

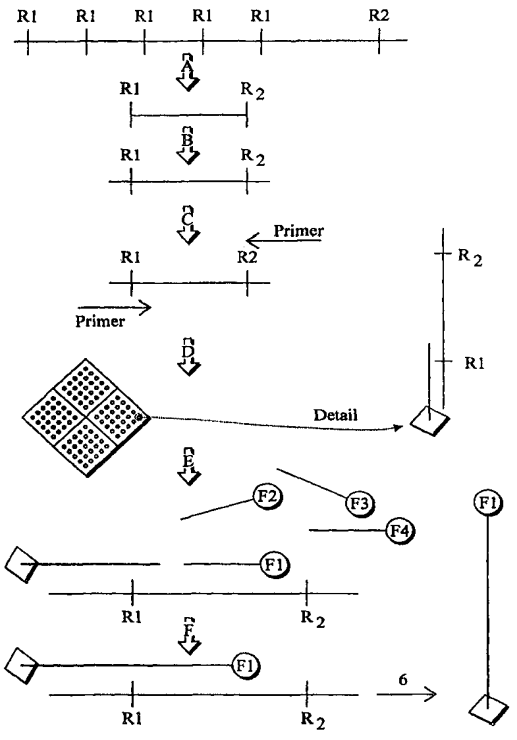
(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: ADAPTER DIRECTED EXPRESSION ANALYSIS

(57) Abstract

The present invention relates to methods and compositions for simultaneously analyzing multiple different polynucleotides of a polynucleotide composition comprising multiple diverse polynucleotide sequences. The subject methods and compositions may also be applied to analyze or identify single polynucleotides, however, the subject methods and compositions are particularly useful for analyzing large diverse populations of polynucleotides, e.g., cDNA libraries. Most embodiments of the invention involve hybridizing terminus probes (of known base sequence) and internal fragment probes (of known base sequence) at adjacent positions on an adapter-modified restriction fragment generated from polynucleotide for analysis, and subsequently joining the terminus probes and internal fragment probes to each other. The terminus probe hybridizes to bases of restriction endonuclease recognition site present at the terminus of a restriction fragment generated from the polynucleotide for analysis. Internal fragment probes hybridizes to the same strand of the restriction fragment that the terminus probe hybridizes to and hybridizes to the restriction fragment portion of adapter-modified representative restriction fragments. The terminus probes and internal fragment probes may be marked so as to facilitate the simultaneous testing of multiple polynucleotides for the presence of many possible nucleotide base sequences. The identity or expression of a particular polynucleotide of interest may be ascertained (or at least partially determined) by producing a short identifier sequence derived from the nucleotide base sequence information obtained from (1) the hybridization of a terminus probe and an internal fragment probe, each of known base sequence, at adjacent positions on a polynucleotide of interest, and (2) the recognition site of a restriction endonuclease used to generate the polynucleotide molecule of interest. Multiple identification sequences may be obtained in parallel, thereby permitting the rapid characterization of a large number of diverse polynucleotides. Parallel processing may be achieved by differentially marking terminus probes or internal fragment probes. Parallel processing may be achieved by using ordered arrays of oligonucleotides that are terminus probes.



IDT Exhibit 1020.8660

FOR THE PURPOSES OF INFORMATION ONLY

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

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

IDT Exhibit 1020.8661

ADAPTER DIRECTED EXPRESSION ANALYSIS

Field of the Invention

Embodiments of the invention are in the field of polynucleotide analysis throughout the use of multiple hybridization probes.

Background

The cells that constitute different tissues in an organism, although having the same genomic DNA, differ significantly from one another with respect to the specific genes that are expressed and the levels of expression. Similar differences in gene expression can be observed when comparing cells that are obtained from a healthy organism and corresponding cells from an organism manifesting a disease state. Other examples of variations in gene expression include changes induced by exposing a cell to a pharmaceutical compound or toxin. It is of interest to provide methods for analyzing changes in gene expression. Methods of analyzing gene expression find wide use in both research and diagnostics.

Summary

Embodiments of the invention described herein relate to methods of analyzing an individual polynucleotide or a polynucleotide mixture comprising multiple diverse polynucleotides, typically a cDNA mixture is formed from an RNA population of interest. Not only is the analysis of RNA populations of major interest in research, such analysis may be used to predict, diagnose, or treat a variety of diseases. Various embodiments of the invention permit the simultaneous analysis of a large number of different mRNA molecules that form a given mRNA population. Various embodiments of the invention also permit the convenient isolation of polynucleotides of interest identified through the subject analytical techniques.

In accordance with the description of the invention provided herein, the identity of a particular polynucleotide of interest may be ascertained by producing a short identifier sequence based on the nucleotide sequence information obtained from (1) the recognition site of a restriction endonuclease used to generate a restriction fragment from the polynucleotide of interest, and (2) the hybridization of: (i) a terminus probe of known sequence, and (ii) an internal fragment probe of known sequence, at adjacent positions on the polynucleotide of interest.

IDT Exhibit 1020.8662

Polynucleotide sequence databases may be conveniently searched for previously identified polynucleotide sequences that match or partially match the identifier sequence. Alternatively, the subject methods may be used to "fingerprint" complex polynucleotide populations without the need to generate identifier sequences. The identification sequences may also be used to develop oligonucleotide primers (or probes) to isolate the polynucleotides from which the identifier sequence is derived.

In preferred embodiments of the invention, representative restriction fragments for analysis are joined to adapters prior to contacting either terminus probes or internal fragments probes. Multiple identifier sequences may be obtained in parallel, thereby permitting the rapid characterization of a large number of polynucleotides. The terminus probe and internal fragment probe as used in the subject methods may each be identified by a "marker" that is correlated with the known base sequence of the probe oligonucleotide so as to facilitate the rapid characterization of a large number of diverse polynucleotides in parallel. Parallel analysis of multiple diverse polynucleotides may be carried by using ordered arrays of oligonucleotides (terminus probes) such that the position of the oligonucleotides in the array serve as markers to identify the base sequence of the oligonucleotide in the array.

In one embodiment of the invention, methods are provided for analyzing diverse polynucleotide mixtures such as a cDNA mixture generated from an RNA population. Restriction fragments are formed by digesting the polynucleotide population for analysis with a restriction endonuclease. Preferably, representative restriction fragments are generated from the different cDNA molecules in the mixture in such a way that only a single restriction fragment is recovered for each polynucleotide analyzed. By employing representative restriction fragments, quantitative (or semi-quantitative) measurements of the relative amounts of different polynucleotides in a polynucleotide mixture for analysis may be greatly facilitated. Adapters may be ligated to the termini of the representative restriction fragments so as to produce a set of adapter-modified representative restriction fragments. The representative restriction fragments may then be optionally amplified in a nucleic acid amplification reaction employing primers specific for the adapters, thereby producing an amplified set of adapter-modified representative restriction fragments. The amplified set of adapter-modified representative restriction fragments (or a corresponding non-amplified set) may then be contacted under nucleic acid hybridization conditions with marked terminus probes so that hybridization may take place between each of

IDT Exhibit 1020.8663

the different adapter-modified representative restriction fragments and each terminus probe present so as to permit hybridization of the probe to complementary strands of the matching adapter-modified representative restriction fragments. Terminus probes may be marked by virtue of their location on an oligonucleotide array. An oligonucleotide array comprising a plurality of oligonucleotide features, wherein each feature of the array is a terminus probe, may be used to analyze a plurality of polynucleotides in parallel. After hybridization with the terminus probe, the adapter-modified representative restriction fragments that have hybridized to the array are contacted (under nucleic acid hybridization conditions) with a labeled probe solution comprising at least one internal fragment probe. The internal fragment probes may be marked, preferably with a fluorescent dye, so as to identify array sites at which hybridization has occurred. Solutions containing multiple differentially labeled internal fragment probes may be used to simultaneously test different internal fragment probes for hybridization to multiple different representative restriction fragments in parallel. Labeled probes that have hybridized to adapter-modified representative restriction fragments at positions adjacent to the terminus of the array feature oligonucleotides are subsequently joined (e.g., by ligase catalyzed ligation), thereby covalently attaching the internal fragment probe to the array. The location of the joined terminus probe or probes on the array may subsequently be identified, thus serving to identify which of the oligonucleotide features (terminus probes) are complementary to a given representative restriction fragment. Sequence information from the internal fragment probe and the terminus probe ligated to the internal fragment probe may be used to obtain an identifier sequence corresponding to the restriction fragment that hybridized to the array at a given feature. Alternatively, arrays of terminus probes may be used to "fingerprint" complex polynucleotide populations with or without the generation of identifier sequences.

Other embodiments of the invention include oligonucleotide arrays comprising features that are complete sets of terminus probes. Embodiments of the subject arrays include arrays that comprise multiple subarrays, wherein at least two of the subarrays comprise the same set of terminus probes; preferably each set of terminus probes in each subarray is a complete set of terminus probes. Each subarray may be organized so as to provide for the addition of oligonucleotide probe solution to one array subunits without having the solution contacting the other subunits.

IDT Exhibit 1020.8664

In another embodiment of the invention, terminus probes and internal fragment probes are hybridized at adjacent positions on an adapter-modified representative restriction fragment (or adapter-modified restriction fragments) and subsequently joined to each other prior to the step of immobilization on a sorting array. The use of sorting arrays permits the step of joining terminus probes and internal fragment probes to take place in solution rather than on an array. In embodiments of the invention employing sorting arrays, either the terminus probes or the internal fragment probes may be marked with an array sorting signal. In embodiments of the invention employing a sorting array, sorting signal marked terminus probes are used in conjunction with the internal fragment probes that are marked with a detectable label. Conversely, internal fragment probes may be marked with sorting signals when used in conjunction with detectably labeled terminus probes. For example, a representative restriction fragment is formed from a polynucleotide of interest. Adaptors are subsequently joined to the termini of the representative restriction fragment so as to produce adapter-modified representative restriction fragments. Preferably, the adapter-modified representative restriction fragments are amplified in a nucleic acid amplification reaction (e.g., PCR) with primers that specifically hybridize to the two adapters so as to increase sensitivity. A terminus probe is then hybridized to a strand of the adapter-modified representative restriction fragment. An internal fragment probe of known sequence is then hybridized to the same strand of the adapter-modified representative restriction fragment at a position adjacent to an end of a terminus probe that has hybridized to the restriction fragment. The internal fragment probes and the terminus probes that have hybridized at adjacent positions on a strand of an adapter-modified restriction fragment are then joined together, e.g., ligated, while hybridized to the strand of the adapter-modified representative restriction fragment. The array sorting signals on the probes may be used to identify which probe has hybridized to a given adapter-modified restriction fragment by virtue of the ability of the array sorting signal to specifically bind to a receptor for the sorting signal at a predetermined location on an array. The detectable label on the detectable label marked probe joined to an array sorting signal marked probe may be used to identify the sequence of the detectable label marked probe.

Other embodiments of the invention include kits for carrying out the methods of the invention. The kits comprise two or more reagents necessary for carrying out an embodiment of the subject methods. Embodiments of the kits of the invention may include an oligonucleotide

IDT Exhibit 1020.8665

array and a set of internal fragment probes and/or terminus probes designed for use in conjunction with the arrays in the kit. The kits may further comprise adapters designed to be used in conjunction with the subject arrays and labeled oligonucleotide probe sets. The inventions also include kits for carrying out embodiments of the invention employing probes that are marked with sorting signals. Kits for use with sorting signals comprise a sorting signal array and probes marked with sorting signals for use in conjunction with the arrays. Optionally, the kits of the invention may comprise other reagents required for performing the subject methods, such reagents include, primers, buffers, DNA, polymerases, DNA ligases, and restriction endonucleases.

Brief Description of Drawings

Figures 1A and 1B are schematic diagrams showing how terminus probes and internal fragment probes may hybridize adapter-modified representative restriction fragments. The variable and constant regions of the terminus probes are indicated. Embodiments of both possible rrealtive orientations of the variable and constant regions on a terminus probe are shown. The vertical lines on the adapter-modiied restriction fragments show the actual junctions between the adapters and the restriction fragment.

Figure 2 is a schematic representation of a method of analyzing complex polynucleotides in accordance with the methods of the invention. R1 and R2 are used to indicate restriction sites. Step A indicates the generation of representative restriction fragments from cDNAs. Step B shows the addition of adapters to the representative restriction fragments to form adapter-modified representative restriction fragments. Step C shows the addition of amplification primers and performance of a PCR reaction to generate adapter-modified representative amplification products. Step D shows the hybridization of a strand of an adapter-modified representative restriction fragment to a terminus probes of an array. An overview of an array comprising subarrays is shown. The hybridization of the adapter-modified representative restriction fragment is shown as a detailed view of an individual array feature. Step E shows the hybridization of a fluorescently labeled internal fragment probe to an adapter-modified representative restriction fragment strand hybridized to a terminus probe. The different circled "F"s are used to indicate fluorescent dyes on different internal fragment probes. Only the F1

IDT Exhibit 1020.8666

labeked probe is hybridized. Step F shows the ligation of the terminus probe to internal fragment probe. The adapter-modified representative restriction fragment is washed away in step G.

Figure 3 is a schematic diagram of formation of a representative restriction fragment by sequential restriction endonuclease digestion. In Step A, cDNA from mRNA is formed using a biotinylated poly dT primer and is bound to avidin immobilized on a solid support. The circled B indicates biotin. In Step B, a first restriction endonuclease (R1) is added and the digestion products are removed by washing. In Step C, a second restriction endonuclease R2 is added and the representative restriction fragment is rcovered.

Figure 4 is a schematic diagram of an embodiment of a technique for generating representative restriction fragments by sequential restriction endonuclease digestion employing a type IIs restriction endonuclease. In Step A, cDNA formed by priming first strand synthesis with a biotinylated poly dT that is bound to avidin immobilized on a solid support is treated with a first restriction endonuclease (R1). In Step B, an adapter having a recognition site (indicated by the bolded rectangle region) for a type IIs restriction endonuclease (R2). In Step C, the type IIs restriction endonuclease is added and the representative restriction fragment isolated. In Step D, a second adapter is added.

Figure 5 is a schematic diagram of an embodiment of the invention employing sorting arrays. In Step A, a terminus probe is marked with an array sorting signal (indicated by the rectangle on the 3' end of the probe) is hybridized to a strand of the adapter-modified representative restriction fragment. In Step B, an internal fragment probe is labeled with a fluorescent dye (indicated by a circled "F1") is hybridized at a position adjacent to the hybridized terminus probe. The non-hybridizing internal fragment probes are indicated by the other circled Fs. In Step C, the terminus probe and the internal fragment probe are ligated together. In Step D, the ligated fluorescently labeled probes are contacted with a sorting array and bind to an array sorting signal receptor at the indicated location on a cirting array. The single rectangles on the array indicate un bound array sorting signal receptors.

Figure 6 is a schematic example showing of the joining of an adaptor (for EcoRI) to an end of a representative restriction fragment andshowing the orientation of terminus probes and internal fragment probes for use in conjunction with this adapter modified restriction fragment. The Xs represent bases of restriction fragment, excluding the restriction site bases. The Y's represent complementary bases on the other strand, excluding the restriction site bases. The

IDT Exhibit 1020.8667

variable region of the terminus probes is indicates by bases Y_1-Y_5. The terminus probe is indicated by bases Y_6-Y_{10}. In step A, an adapter is joined to a restriction fragment. In step B, a terminus probe is shown in alignment with the adapter-modified restriction fragment. In step C, a terminus probe and internal fragment probe are shown in alignment with the adapter-modified restriction fragment.

Definitions

1. The term "terminus probe" as used herein refers to an oligonucleotide that is complementary to a specific portion of a strand of an adapter-modified restriction fragment (including restriction fragments that are representative restriction fragments). Terminus probes serve to identify (through hybridization) one or more nucleotides of a restriction fragment that are adjacent to the junction formed between an adapter and a restriction fragment. Terminus probes are complementary to one or more bases adjacent to both sides of the junction. Terminus probes are complementary to bases at and near the junction formed between an adapter and a restriction fragment. Terminus probes may thus be conveniently referred to as complementary to junctions between the adapter and the restriction fragment, even though the actual junction is formed by the juxtaposition of only two bases. Because adapter-modified restriction fragments may have two junctions between adapters and the restriction fragment (one for each terminus of the restriction fragment), the terminus probe may be selected so as to be complementary nucleotide bases at either of the two junctions (but not both junctions). In many embodiments of the invention, multiple terminus probes are selected to be used in conjunction with one another, i.e., sets of terminus probes, thereby providing for the simultaneous analysis of multiple polynucleotides when the different probes are used in conjunction with one another. The different terminus probes that form a given set of terminus probes are selected so as to hybridize to the equivalent junction and some strands of the different adapter-modified restriction fragments produced in a given embodiment of the subject methods. By "equivalent junction," it is intended that the junction formed between the same adapter and the same restriction fragment terminus be used for all the adapter-modified restriction fragment analyzed by a given set of terminus probes.

Terminus probes are said have a "constant region" and a "variable region." A given nucleotide base in a terminus probe is either in the constant region or the variable region, but not

IDT Exhibit 1020.8668

both. The "constant region" is said to be constant because the constant regions of a set of terminus probes are functionally the same as each other with respect to their hybridization specificity as used in the methods of the invention. The constant region is complementary to (1) nucleotides of the restriction endonuclease recognition site used to generate a terminus of the restriction fragment for analysis, and (2) at least a portion (and preferably all) of the adapter adjacent to the restriction endonuclease recognition site at the adapter/restriction fragment junction. The length of the constant region will be, in part, determined by (1) the length of the recognition sequence of the relevant restriction endonuclease, and (2) the length of the adapter used in conjunction with restriction fragments. Typically, the length of the constant region is 5-20 nucleotide bases in length. The constant region may extend past the region of the terminus probe designed to hybridize to adapter-modified restriction fragments. This additional probe region may be used to reduce steric hindrance between solid supports and restriction fragment hybridization. The "variable region" is said to be variable because the variable region of a set of terminus probes have different nucleotide base sequences. The variable region of a terminus probe is complementary to a portion (typically 1-8 nucleotide bases) of the restriction fragment that are adjacent to the restriction endonuclease recognition site base or bases at the terminus of the restriction fragment used to form the adapter-modified restriction fragment.

The "variable region" of a terminus probe may be located either 5' or 3' with respect to the constant region. The selection of the relative orientation of the variable region with respect to the constant region in a given embodiment of the invention will vary in accordance with choice of which strand of the adapter-modified restriction fragment is selected for analysis and in accordance with the choice of which terminus of the adapter-modified restriction fragment is to be analyzed (see Figure 1).

(2) The term "a set of terminus probes" as used herein refers to a plurality of different terminus probes used in conjunction with each other, wherein each probe in the set has a functionally identical constant region, e.g., all of the constant regions are identical or have essentially the same sequence-specific hybridization properties, and the variable regions are different from one another. The term "A complete set of terminus probes" refers to a set of terminus probes that includes all possible nucleotide variable sequences (for the four major nucleotide bases, A, C, G, and T, or functional equivalents thereof). All variable regions of the oligonucleotides in a set of terminus probes are the same length. Thus the number of different

IDT Exhibit 1020.8669

terminus probes required to form a complete set of terminus probes increases with the length of the variable region. The number of different oligonucleotides required to form a complete set increases with the length of the variable regions in the terminus probes that form a set (4^N, wherein N is the length of the variable region).

(3) The term "internal fragment probe" refers to an oligonucleotide that may hybridize to a strand of an adapter-modified restriction fragment for analysis at a position immediately adjacent to the location on the strand to which a terminus probe may hybridize. Internal fragment probes hybridize to the nucleotide of the restriction fragment portion of adapter-modified restriction fragments, but do not hybridize to the adapter sequence.

(4) The term "complete set" as used in reference to internal fragment probes refers to a set of oligonucleotides having all possible nucleotide bases (of the four major nucleotide bases, A, C, G and T or equivalents thereof) for length of the oligonucleotide. All oligonucleotides in a complete set are of the same length. The number of different oligonucleotides required to form a complete set increases with the length of the oligonucleotides in a set (4^N, wherein N is the oligonucleotide length).

(5) The term "oligonucleotide array" as used herein refers to a solid support that has a plurality of different oligonucleotides attached at pre-determined spatial locations on the solid support. The support may be in any of a variety of shapes. In preferred embodiments of the invention, the solid support is substantially planar and the different oligonucleotides are arranged in a two-dimensional matrix. Oligonucleotide arrays may be prepared by a variety of methods well known to those skilled in the art. Such methods include the spotting of oligonucleotides on to the solid support, photolithographically controlled in situ synthesis on the solid support using phosphoramidites with photolabile protecting groups, and in situ synthesis on a solid support using conventional phosphoramidites. Various methods of synthesizing oligonucleotide arrays have been described and can be found in, among other places, U.S. Patent Nos. 5,510,270; 5,405,783; 5,143,854; 5,489,678; 5,733,509; 5,412,087; 5,436,327; and PCT Publication WO 95/25116. There is a pre-determined correlation between a given spatial location on the solid support of the array and the particular oligonucleotide located at that given spatial location that is created during the synthesis of the array. Thus by referring to a particular spatial location on the array, the nucleotide base sequence of the oligonucleotide present at that spatial location may be known.

IDT Exhibit 1020.8670

(6) The term "feature" as used herein with respect to oligonucleotide arrays and sorting arrays refers a group of essentially functionally identical macromolecules, e.g., oligonucleotides, at a pre-determined location on an array.

(7) The term "oligonucleotide" as used herein is used broadly to refer to any naturally occurring nucleic acid, or any synthetic analogs thereof, that have the chemical properties required for use in the subject methods, e.g., the ability to sequence specifically hybridize different polynucleotides. Thus, examples of oligonucleotides include DNA, RNA, phosphorthioates PNAs (peptide nucleic acids), phosphoramidates and the like. Methods for synthesizing oligonucleotides are well known to those skilled in the art, examples of such synthesis can be found for example in U.S. 4,419,732; 4,458,066; 4,500,707; 4,668,777; 4,973,679; 5,278,302; 5,153,319; 5,786,461; 5,773,571; 5,539,082; 5,476,925; and 5,646,260.

(8) The terms "sorting array" and "sorting subarray" refer to an array formed by an array sorting signal receptors that are attached a fixed pre-determined spatial location on a solid support. There is a pre-determined correlation between a given spatial location on the solid support of the array and the sorting signal receptor located at a given spatial location that is created during the manufacture of the array. Thus by referring to a particular spatial location on a sorting signal array, the identity of the sorting signal receptor at that particular spatial location may be known. An oligonucleotide array may be an embodiment of a sorting signal array. Preferred sorting arrays for use in the invention are oligonucleotide arrays.

(9) The term "array sorting signal" as used herein refers to a member of a specific binding pair. The second member of the specific binding pair is referred to as an array sorting signal receptor. In a preferred embodiment of the invention, the array sorting signals are oligonucleotides.

(10) The term "array sorting signal receptor" as used herein refers to a member of a specific binding pair. The second member of the specific binding pair is an array sorting signal. In a preferred embodiment of the invention, the array sorting signals are oligonucleotides.

(11) The term "specific binding pair" as used herein refers to a pair of molecules, typically macromolecules, that specifically bind to each other. Each member of the specific binding pair may be referred to as a "specific binding pair member." Examples of specific binding pairs include, complementary oligonucleotides, antibody-antigen pairs, lectin-sugar

IDT Exhibit 1020.8671

pairs, receptor-ligand pairs, and the like. Preferably, specific-binding pair members bind to each other with high affinity as well as high specificity.

(12) The term "representative restriction fragment" refers to a polynucleotide restriction endonuclease digestion product that is derived from a larger polynucleotide in such a way as to produce a predetermined constant number, preferably one, restriction fragment for the polynucleotide from which the representative restriction fragment was derived. For example, a representative restriction fragment of a 4.5 Kb cDNA may be a 0.8 Kb subfragment having EcoRI and Hind III generated termini and produced by treating the cDNA with the indicated restriction endonuclease. Although the representative restriction fragments are preferably formed through the use of restriction endonuclease digestions of polynucleotides, it will be appreciated by those skilled in the art that the functional equivalents of representative restriction fragments can be produced by the sequence specific cleavage mechanisms other than the use of restriction endonucleases, e.g., oligonucleotides joined to metallic cations. The term "representative restriction fragment," as used herein may refer collectively to representative restriction fragments, adapter-modified representative restriction fragments, and the amplification products of adapter-modified representative restriction fragments. Those instances in which the term "representative restriction fragment" refers only to representative restriction fragment, will be apparent by virtue of the context in which the term is used, e.g., the subsequent manipulations to be performed on the representative restriction fragment. Similarly, the term "restriction fragment" may be used herein to include "representative restriction fragments" as well as conventional restriction fragments. By virtue of context, it will be apparent to those skilled in the art when the term should be construed to exclude representative restriction fragments.

(13) The terms "restriction endonuclease recognition site" and "recognition site" as used herein refers not only to the nucleotides that form a restriction endonuclease recognition site, but also includes: (1) residual nucleotides that were part of a recognition site and remain after restriction endonuclease digestion, (2) nucleotides that are complementary to the residual nucleotides that were part of the recognition site and remain after restriction endonuclease digestion (and subsequent manipulations). By virtue of the context in which the term is used, it will be readily understood by a person skilled in the art that when term "restriction fragment endonuclease recognition site" refers to a complete recognition site and when the term refers to the portion of the recognition site that remains after digestion and subsequent manipulations.

IDT Exhibit 1020.8672

(14) The term "fingerprint" as used herein refers to a set of data relating to a complex polynucleotide population in which the relative concentrations of the different polynucleotide that formed the population are measured.

(15) The term "identifier sequence" as used herein refers to a small (e.g., 10-30 base pairs in length) polynucleotide sequence that is present in a larger polynucleotide. The identifier sequence is of sufficient length to permit the identification of a larger polynucleotide comprising the identifier sequence. The identifier sequence may be of contiguous or non-contiguous base sequence information. Preferably, the identifier is a contiguous sequence of nucleotide bases.

(16) The term "marker" as used herein, refers to a compound or method for tracking the identity of an oligonucleotides of known base sequence. A marker may be specifically associated with a given oligonucleotide, the base sequence of the nucleotide may be determined because of the predetermined correlation between the base sequence of the oligonucleotide and the marker. The specific oligonucleotide associated with a marker is said to be "marked."

(17) The term "adapter" as used herein refers to a double-stranded oligonucleotide having a terminus that is capable of being joined to the terminus of a restriction fragment. The terminus of the adapter may have a 3' overhand, a 5' overhang, or may be blunt-ended. As the terminus of the adapter is designed to be complementary to the terminus of a restriction fragment produced by a given restriction endonuclease, the nature of the terminus of the adapter will vary in accordance with the nature of the termini of the restriction fragments for ligation. The terminus nucleotide moieties of the adapter are selected so as to be compatible with the particular joining method used to join the adapter to the restriction fragment of interest. For example, when the joining is catalyzed by a DNA ligase, the 5' terminal nucleotide is phosphorylated and the 3' terminal nucleotide has a hydroxy group. Adapters for use in the subject methods comprise sufficient additional nucleotide (additional with respect to those nucleotides that are required to form an overhanging terminus) to permit the adapter to be used as a site for primer binding in a nucleic acid amplification reaction, e.g., PCR. In preferred embodiments of the invention, the nucleotide bases of the adapter terminus for joining a restriction fragment are selected so as to not recreate the restriction endonuclease site used to generate the terminus of the restriction fragment. Guidance on the joining of adapters to restriction fragments can be found in, among other places, U.S. Patent Nos. 5,093,245 and 5,366,877.

IDT Exhibit 1020.8673

(18) The term "joining" as used herein, with respect to oligonucleotides or polynucleotides refers to the covalent attachment of two separate nucleic acids to produce a single larger nucleic acid with a contiguous backbone. Preferred methods of polynucleotide joining are ligase (e.g., T-4 ligase) catalyzed reactions. However, non-enzymatic ligation methods may also be employed. Examples of ligation reactions that are non-enzymatic include the non-enzymatic ligation techniques described in U.S. Patent Nos. 5,780,613; 5,476,930, which are herein incorporated by reference.

Descriptions of Specific Embodiments

The present invention relate to methods and compositions for simultaneously analyzing multiple different polynucleotides of a polynucleotide composition comprising multiple diverse polynucleotide sequences. The subject methods and compositions may also be applied to analyze or identify single polynucleotides; however, the subject methods and compositions are particularly useful for analyzing large diverse populations of polynucleotides, e.g., cDNA libraries. Most embodiments of the invention involve hybridizing terminus probes (of known base sequence) and internal fragment probes (of known base sequence) at adjacent positions on an adapter-modified restriction fragment generated from polynucleotide for analysis, and subsequently joining the terminus probes and internal fragment probes to each other. The terminus probe hybridizes to bases of a restriction endonuclease recognition site present at the terminus of a restriction fragment generated from the polynucleotide for analysis. Internal fragment probes hybridize to the same strand of the restriction fragment that the terminus probe hybridizes to and hybridizes to the restriction fragment portion of adapter-modified representative restriction fragments. The terminus probes and internal fragment probes may be marked so as to facilitate the simultaneous testing of multiple polynucleotides for the presence of many possible nucleotide base sequences.

Analysis of polynucleotide populations in accordance with methods of the invention may be used to provide one or more of the following types of information: (1) the nucleotide sequence of one or more polynucleotides in a complex polynucleotide composition, (2) partial nucleotide sequences of one or more polynucleotides in a complex polynucleotide composition, or (3) the relative concentrations of one or more different polynucleotides in a complex polynucleotide composition. Analysis of large complex populations of polynucleotides by the subject methods

IDT Exhibit 1020.8674

may be used to produce sufficient information about a polynucleotide population that differences between polynucleotide populations may be ascertained. Thus in some embodiments of the invention, "fingerprints" of a given polynucleotide population may be compared with "fingerprints" of other complex polynucleotide populations so as to determine differences in gene expression between the two populations. In addition to providing fingerprints of complex polynucleotide populations, some nucleotide base sequence information may be obtained for one or more polynucleotides in the population. An important example of a polynucleotide composition that may be analyzed by the invention is a cDNA preparation derived from an RNA population. The analysis of polynucleotide mixtures, particularly cDNA preparations, has numerous practical uses such as measuring gene expression for diagnostic or research purposes. Of particular interest are embodiments of the present invention that permit the majority of different polynucleotides in an RNA population may be detected.

The identity or expression level of a particular polynucleotide of interest may be ascertained (or at least partially determined) by producing a short identifier sequence derived by combining from the nucleotide base sequence information obtained from (1) the hybridization of a terminus probe and an internal fragment probe, each of known base sequence, at adjacent positions on a polynucleotide of interest, and (2) the recognition site of a restriction endonuclease used to generate the polynucleotide molecule of interest. The combining of the different base sequence information inputs to produce an identifier sequence may be carried out by a programmable calculating device, e.g., an electronic computer, so as to conveniently automate the process when applied to complex polynucleotide populations. An identifier sequence may consist of contiguous or non-contiguous base sequence information. In a typical embodiment, (1) a terminus probe hybridizes to a strand of an adapter-modified representative restriction fragment at a position that includes the junction between the restriction fragment and the adapter, (2) an internal fragment probe hybridizes to the restriction fragment at a position immediately adjacent to a terminus of the hybridized terminus probe, and (3) the terminus probe and the internal fragment probe are subsequently joined together, e.g., by ligation. An identifier sequence may be used to produce oligonucleotide primers or probes to isolate the polynucleotides from which the identifier sequence was derived. Multiple identifier sequences may be obtained in parallel, thereby permitting the rapid characterization of a large number of diverse polynucleotides. Parallel processing may be achieved by differentially marking terminus probes

IDT Exhibit 1020.8675

or internal fragment probes. Parallel processing may be achieved by using ordered arrays of oligonucleotides that are terminus probes.

Preferred embodiments of the invention employ sets of first and second oligonucleotides of known sequence that are used in combination to hybridize to restriction fragments for analysis. These first and second oligonucleotides are the terminus probes and internal probes used in the subject methods. By using two sets of oligonucleotide probes in combination with each other, the total number of different oligonucleotides used to analyze most of the different polynucleotides in a complex nucleotide population is sharply reduced. For example, by using a first set of all possible 5-mers ($1024 = 4^5$) and a second set of all possible 5-mers (1024), in combination with each other, 1,048,576 ($4^{10} = 4^5 x 4^5$) different 10 base combinations may be interrogated with only 2048 different probes. As there are approximately 100,000 genes expressible in a human cell, two sets of 5-mer oligonucleotides may be used to conveniently divide a cDNA population derived from a human cell (or comparably complex eukaryotic cell) into several separate groups containing sufficiently small numbers of different restriction fragments so as to provide for the convenient analysis of the cDNA libraries and other complex polynucleotide compositions. Additional sequence information about the polynucleotides for analysis may be obtained from restriction endonuclease recognition sites in the polynucleotides for analysis. For example, the restriction endonuclease EcoRI has a 6 base pair recognition sequence. Thus, for example, by combining the sequence information obtained from identifying which two 5-mer oligonucleotides hybridize adjacent to an EcoRI recognition site, 16 bases of sequence information may be obtained so as to provide a unique identifying sequence that may be used to search sequence databases, prepare hybridization probes, or prepare amplification primers.

Pools of internal fragment probes, each of known sequence and labeled with a distinctive detectable label, may be used to decrease the number of steps necessary to obtain an identifier sequence for a given representative restriction fragment in the subject methods. For example, four different internal fragment probes may be simultaneously tested for the ability to hybridize adjacent to the terminus of a terminus probe hybridized to a representative restriction fragment, if each of the four different oligonucleotides is labeled with a different detectable label, i.e. a fluorescent dye, the presence of a particular label being indicative of the hybridization of the oligonucleotide modified by the label. If four different labels were not used, the alternative

IDT Exhibit 1020.8676

would have been to employ four separate hybridization reactions. Thus, the degree of savings in manipulation steps is in part directly proportional to the number of different differentially labeled oligonucleotide probes used in the hybridization steps.

In a preferred embodiment of the invention, the restriction fragments for analysis are representative restriction fragments. The use of representative restriction fragments rather than more complex mixtures of restriction fragments minimizes quantitation problems associated with attempting to correlate the analysis of multiple restriction fragments derived from each polynucleotide for analysis. By measuring the quantity of a representative restriction fragment produced from a polynucleotide composition, the quantity of the larger polynucleotide from which the representative restriction fragment was derived may be conveniently measured. Similarly, the relative quantities of different polynucleotides may be compared by comparing the relative quantities of different representative restriction fragments.

Embodiments of the invention include many different methods of analyzing polynucleotides. One embodiment of the subject methods includes the steps of: (1) forming a restriction fragment (preferably a representative restriction fragment) from a polynucleotide for analysis, (2) hybridizing a terminus probe to a single strand of the restriction fragment at a position on the restriction fragment that includes the restriction endonuclease-generated terminus, (3) hybridizing an internal fragment probe to the same strand of the restriction fragment such that the terminus probe is hybridized at a position immediately adjacent to the terminus probe, and (4) joining the terminus probe to the internal fragment probe. Terminus probes and internal fragment probes may be marked so as to facilitate the parallel analysis of multiple constituents of a complex polynucleotide population. Terminus probes are preferably marked by their spatial location on an oligonucleotide array and internal fragment probes are preferably marked by a fluorescent dye. The markers on the terminus probes and internal fragment probes that have been joined together may then be ascertained in accordance with the specific marking techniques used.

In a preferred embodiment of the invention, at least one adapter is joined to a restriction endonuclease-generated terminus of a representative restriction fragment, more preferably, two adapters are ligated to the two termini of the representative restriction fragments. The adapters are joined to the restriction fragment termini prior to hybridization of the restriction fragments to terminus probes and internal fragment probes. The amplification of the fragment by PCR (or a similar nucleic acid amplification technique) through the use of primers that can anneal to

IDT Exhibit 1020.8677

strands of the adapters. Additionally, the adapters may be used to recover of the restriction fragment from which an identifier sequence is derived, e.g., by performing PCR with a primer derived from the identifier sequence and a second primer specific for an adapter joined to the terminus that did not hybridize to the relevant terminus probe. Furthermore, an adapter may be used as a template for facilitating the hybridization of the terminus probe to the restriction fragments for analysis.

Junctions are formed at the site of joining between the termini of the representative restriction fragment and the adapters. Terminus probes are designed to hybridize to the nucleotides forming one of these junctions, i.e., to hybridize to regions of the restriction fragment and adapter that are adjacent to the junctions. Because the adapters are the same or contain regions with a common or substantially similar polynucleotide sequence, the terminus probes that are used to hybridize to the adapter modified representative restriction fragments have a constant region for hybridizing to the adapter and to terminal restriction endonuclease recognition site. Additionally, the terminus probes have a variable region for hybridizing to the restriction fragment nucleotides adjacent to the restriction endonuclease recognition site at the terminus. The variable regions of the terminus probes serve to divide the adapter-modified restriction fragments into subgroups based on their base sequences adjacent to a terminal restriction endonuclease recognition site.

In a preferred embodiment of the invention, adapter modified representative restriction fragments are amplified in a nucleic acid amplification reaction, e.g., PCR. The product of the amplification product are referred to herein as "adapter-modified representative restriction fragment amplification fragments." The terms "adapter-modified representative restriction fragment amplification products" and "adapter-modified representative restriction fragments" may be used interchangeably in the methods described herein unless indicated otherwise by the context of usage of the term. A nucleic acid amplification step may be used when the signal from the adapter-modified restriction fragments is not sufficient for the production of useful levels of detectable signal in the given polynucleotide sample for analysis.

Terminus probes and internal fragment probes as used in the subject methods are identified with a "marker" that is correlated with to the base sequence of the oligonucleotide. Thus by identifying the marker specifically associated with a given oligonucleotide, the base sequence of the nucleotide may be determined because of the predetermined correlation between

IDT Exhibit 1020.8678

the base sequence of the oligonucleotide and the marker. The marker may take on any of a variety of different forms. Such marker forms include: predetermined locations on arrays, fluorescent tag molecules, chromophore tag molecules, chemiluminescent tag molecules, specific binding pair members, temporally distinct addition of oligonucleotides, and the spatially distinct addition of oligonucleotides. Different forms of markers may be used alone or in combination with one another.

In one embodiment of the invention, the terminus probes may be marked by virtue of predetermined locations on a solid support, e.g., an oligonucleotide array. Thus by knowing the base sequence of a polynucleotide feature at a particular spatial location of an array, the complementary base sequence of a strand of the adapter modified restriction fragment that hybridizes to the array at the specific location may be ascertained. The array serves to "spatially" mark terminus probes.

In those embodiments of the invention in which the terminus probe is a feature of an oligonucleotide array, the internal fragment probe may be marked so as to distinguish themselves from one another by virtue of the spatial location on the array to which the internal fragment probe is added. The use of oligonucleotide arrays in various embodiments of the invention may serve one or more functions. First, by employing arrays of different oligonucleotides, a given restriction fragment may be simultaneously tested for the ability to hybridize to a large number of different terminus probes of a known sequence. Second, multiple restriction fragments may be tested for the ability to hybridize to a terminus probe on the array.

In preferred embodiments of the invention, spatial marking may be achieved by providing an oligonucleotide array divided into various subarrays, wherein each subarray comprises essentially the same set of oligonucleotides features (in the same or different spatial arrangement from each other). For example, an array of terminus probes having variable regions 5 nucleotides in length may comprise 1024 identical subarrays, wherein, each of the subarrays is identical to each other and is formed by 1024 different terminus probes. The subarrays may be organized in such a way as to permit an oligonucleotide solution to be added to one subarray without contacting the other subarrays. In this example, 1024 internal fragment probes (all possible 5-mer sequences) are separately added to each of the different subarrays. The identity of the internal fragment probe may thus be tracked by virtue of the specific subarray to which it is

IDT Exhibit 1020.8679

added. The spatial marking effect of using multiple subarrays may be achieved using an equivalent system multiple distinct arrays on separate solid supports.

In other embodiments of the invention, the internal fragment probes may be marked by virtue of temporal properties assigned to oligonucleotides of known sequence. The term "temporal properties" or "temporal marking" refer to methods employing the sequential addition of oligonucleotides. For example, (1) adapter-modified representative restriction fragment amplification products may be hybridized to terminus probes that are features of an oligonucleotide array, (2) then the hybridized adapter-modified restriction fragments are contacted with a internal fragment probe, (3) the internal fragment probe is then hybridized to the terminus probe, (4) the array is then analyzed, i.e., interrogated, so as to determine which features have been joined to internal fragment probes steps (2)-(4) are then repeated several times with a different internal fragment probe. Thus, by using successive rounds of hybridization with different internal fragment probes and correlating the hybridization round with the probes used during that particular round of hybridization, the internal fragment probes may be distinguished from one another. The different internal fragment probes used in temporal marking may be the same or different from one another.

Another method of using a "marker" to identify an oligonucleotide probe is by means of different detectable labels, e.g., fluorescent dye molecules. Detectable labels are chemical compounds that may be detected in relatively small amounts by virtue of their chemical or physical properties. Detectable labels may take on any of a variety of forms such as fluorescent dyes, chromophores, chemiluminescent molecules, radioactive isotopes, spin labels, enzymes, and the like. A wide variety of labels and labeling techniques may be used to mark probes, with the identifying criteria that the label/labeling method should not significantly interfere with. The detectable labels may be detected either directly or indirectly. Indirect detection employs a chemical intermediate that produces the signal detected by the actual detection device employed. Examples of indirect labeling include enzymes used with a fluorescent substrate, biotin labels used in conjunction with enzymatically labeled avidin, used in conjunction with fluorescently labeled antibodies, and the like. Chemical tags such as fluorescent dyes and other detectable labels permit multiple oligonucleotides (e.g., internal fragment probes to be used together in the same solution. Detectable labels are also of interest because, depending upon the particular detection scheme employed, they permit the quantitative (or semi-quantitative) detection of the

IDT Exhibit 1020.8680

label and hence the quantitative (or semi-quantitative) detection of the probe joined to the detectable label. In a preferred embodiment of the invention, the detectable labels employed are fluorescent dyes. An example of how to use probes marked with detectable labels is as follows. For example, (1) adapter modified representative restriction fragment amplification products may be hybridized to terminus probes that are features in an array, (2) then the hybridization adapter-modified restriction fragments are contacted with a pool of four different internal fragment probes, wherein each of the internal fragment probes are labeled with a distinct fluorescent dye that is correlated with the base sequence of the internal fragment probes, (3) the internal fragment probes are joined to the terminus probes that have hybridized to the adapter-modified restriction fragment amplification products, (4) the array is then interrogated so as to identify the specific fluorescent labels that have been attached to specific oligonucleotide features. Methods and compounds for attaching detectable labels to oligonucleotides are well known to those skilled in the art. Examples of such methods can be found in Hermanson, Bioconjugate Techniques, Academic Press, San Diego (1996); U.S. Patent Nos. 5,366,860; 5,231,141; 5,188,937; 4,605,735; 4,667,025; 4,789,737; and 4,820,812. The specific methods and instrumentation suitable for detecting a specific detectable label will vary in accordance with the physical and chemical properties of the label. Such methods and instrumentation are well known by those skilled in the art. Examples of such methods and instrumentation can be found in U.S. Patent No. 5,324,633 and PCT Patent Application No. WO 95/22068.

A detailed example of using detectable chemical labels and spatial marking is as follows. A cDNA library is prepared. Representative restriction fragments are formed from the cDNA library. The representative restriction fragments have a Hind III derived terminus and a EcoRI derived terminus. Adapters specific for the Hind III terminus and the EcoRI terminus are ligated to the representative restriction fragments, thereby forming adapter-modified representative restriction fragments. The adapter-modified representative restriction fragments are mixed with a pair of amplification primers specific for the adapters used to generate the adapter-modified representative restriction fragments and subjected to a PCR amplification reaction so as to generate a set of adapter-modified representative restriction fragment amplification products. The amplification products are then added to an array of terminus probes. The array comprise 256 identical subarrays, each subarray comprising a complete set of 1024 terminus probes. Each of the terminus probes has a constant region designed to be complementary to a strand of the

IDT Exhibit 1020.8681

adapter joined to the Hind III terminus of the representative restriction fragments. The terminus probes have a 5 nucleotide variable region on their 3' end. To each subarray is added a different solution comprising four different internal fragment probes, each probe being labeled with different fluorescent dye capable of being distinguished from one another based on their spectroscopic properties. Each internal fragment probe is 5 nucleotides in length. Thus a complete set of 5 nucleotide internal fragment probes is prepared and distributed among 256 separate solutions. The internal fragment probe containing solutions are added to each subarray in such a manner that the solution added to a given subarray does not contact other subarrays. A DNA ligase is added in order to catalyze the joining of internal fragment probes that have hybridized at adjacent positions on an adapter-modified representative restriction fragment. The array is subsequently washed to remove terminus probes that have not been joined to the array as well as removing the adapter-modified representative restriction fragments. The array is the scanned in fluorescence detection imaging device capable of (1) spectrally resolving the four fluorescent dyes originally attached to the internal fragment probes and (2) determining the location of the dye on the array. The base sequence of the terminus probes that have been joined and the base sequence of the internal fragment probes may then be combined with each other and the Hind III recognition sequence to produce identifier sequences corresponding to the different cDNAs in the original library. Moreover, the intensity of the fluorescence is indicative of the relative concentrations of the cDNAs in the library.

In another embodiment of the invention, specific binding pair members may be used to mark the terminus or internal fragment probes of known sequence. Of particular interest are the use of specific binding pair members, the pair of members being (i) array-sorting signals and (ii)array-sorting signal receptors. In preferred embodiments of the invention, the array sorting signals are single-stranded oligonucleotides and the array-sorting signal receptor are complementary oligonucleotides. By attaching array-sorting signals to the terminus probes or internal fragment probe, the methods of the invention may be used in such way that the joining of the terminus probe and internal fragment probes takes place in solution rather than on a solid phase (e.g., an array). In those embodiments of the invention employing array-sorting signals on the terminus probes or internal fragment probes marked with array-sorting signals, a sorting array, i.e., an array of array-sorting signal receptors, is employed to be used in conjunction with the array sorting signals. By attaching an array-sorting signal receptor at a specific location on

IDT Exhibit 1020.8682

an array, the presence of an oligonucleotide having the cognate array-sorting signal may be detected. Array-sorting signals may be present on the terminus probes or internal fragment probes. In those embodiments of the invention in which terminus probes are marked with an array sorting signal, the internal fragment probes to be used in conjunction with such probes are preferably marked with a detectable label, e.g., a fluorescent dye. Conversely, in those embodiments in which internal fragment probes are labeled with an array sorting signal, the terminus probes to be used in conjunction with such probes are preferably marked with a detectable label, e.g., a fluorescent dye. The array sorting signals are preferably selected so as to minimize cross-hybridization between the different sorting signal and sorting signal receptors. In those embodiments of the invention employing oligonucleotides as sorting signals and sorting signal receptors, cross-hybridization may be avoided (or minimized) by selecting oligonucleotide sequences with minimal amounts of sequence homology between the different oligonucleotides.

In one embodiment of the invention employing array-sorting signals, (1) terminus probes are prepared, wherein each of the different terminus probes is labeled with a different array-sorting signal, (2) the terminus probes are hybridized to adapter-modified representative restriction fragments (or adapter-modified representative restriction fragment amplification products) prepared from the polynucleotide population of interest, (3) detectably labeled (e.g., fluorescently labeled) internal fragment probes are contacted with the terminus probes hybridized to the adapter-modified representative restriction fragments, (4) internal fragment probes that have hybridized adjacent to the terminus probes are joined to the terminus probes, (5) the polynucleotide products of the joined terminus probe and internal fragment probes pairs are then contacted with a sorting array, i.e., an array of array-sorting signal receptors, and (6) the array is then interrogated so as to detect those array-sorting signal receptors that have bound the detectable label of the internal fragment probes. Internal fragment probes alone cannot bind to the array because the internal fragment probes do not have array-sorting signals. Terminus probes that have bound alone to the array cannot be detected because the oligonucleotides lack detectable labels (e.g., fluorescent dyes). Thus, only the product of terminus probes joined to internal fragment probes are detected by arrays of receptor-sorting signals. In order to test large numbers of different terminus and internal fragment probes in parallel, the adapter-modified representative restriction fragments may be divided into several aliquots, to each aliquot may be added a complete set of terminus probes (each probe labeled with distinctive array sorting

IDT Exhibit 1020.8683

signals) and a subset of internal fragment probes. Each of the internal fragment probes in the subset may be labeled with a distinctive fluorescent label. After joining terminus probes and internal fragment probes to each other, each aliquot may be contacted with a separate subarray of array-sorting signal receptors. An example of an embodiment of the invention employing array sorting signals is as follows: (1) a mixture of adapter-modified representative restriction fragments may be divided into 256 separate aliquots, (2) to each aliquot is added a set of 1024 terminus probes, each oligonucleotide having a 6 base constant region and a five base variable region, wherein all possible variable regions are represented and each oligonucleotide is marked with a distinctive array-sorting signal, (3) to each aliquot is added a set of four different internal fragment probe 5-mers, each of the four different oligonucleotides labeled with a different fluorescent label; a total of 1024 different 5-mers covering all possible 5-base sequence variations is distributed among the 256 different aliquots, (4) after hybridization and the joining of the terminus and internal fragment probes is completed, each aliquot is distributed onto a separate array of array-sorting signals receptors, wherein each array or subarray has 1024 signal-sorting receptor specific for the array sorting signal on the terminus probes.

Formation of Representative Restriction Fragments

Many different methods of producing representative restriction fragments may be used to practice the methods of the invention. These methods of producing representative restriction fragments may be used interchangeably with the different embodiments of the invention. In embodiments of the invention that include the step of amplifying adapter-modified representative restriction fragments, it is preferred that the method of generating representative restriction fragments be a method that generates representative restriction fragments that are substantially the same length. Methods for generating representative restriction fragments of substantially the same length may employ type IIS restriction endonucleases as described below in this section.

Methods of preparing representative restriction fragments include the technique of sequential restriction endonuclease digestion with two restriction endonuclease (having different recognition sites) as applied to immobilized DNA fragments. For the sake of convenience, this technique may be referred to as "sequential restriction digestion." Examples of sequential restriction digestion include the following method and variations thereof. cDNA is prepared from an RNA preparation of interest. The cDNA is immobilized on a solid phase. A large

IDT Exhibit 1020.8684

representative sample of the cDNA components of the cDNA preparation are immobilized at the same end of the cDNA molecule (typically the end of the cDNA corresponding to the mRNA polyA tail is immobilized). The immobilized cDNA is then digested with a first restriction endonuclease. The released, i.e., not bound, restriction fragments may then be washed away. Thus, the restriction endonuclease digestion process results in the production of an immobilized cDNA fragment having a terminus generated by a restriction endonuclease. A first adapter may optionally be ligated to the cDNA termini produced by the first restriction enzyme at this time. The immobilized cDNA is then digested with a second restriction enzyme having a different recognition site than the first restriction endonuclease. A second adapter is then ligated to the cDNA termini produced by the second restriction endonuclease. The released cDNA restriction fragments, each having termini produced by different restriction enzymes, are then collected and may subsequently be joined to adapters specific for each of the two termini. If an adapter had not been previously ligated to the termini produced by the first restriction endonuclease digestion, then first adapters are ligated to these termini. After the second adapter has been ligated, formation of adapter-modified representative restriction fragments has been completed.

The cDNA fragments may be immobilized to a solid phase at a predetermined end of the cDNA by a variety of methods, provided such methods do not substantially interfere with subsequent restriction endonuclease or ligation reactions. In a preferred embodiment of the invention, cDNA is synthesized using a biotinylated polyT primer to initiate first strand synthesis (priming from the RNA polyA tail). The biotinylated portion of the cDNA is then complexed with immobilized streptavidin or avidin. Of course, many specific binding pairs, e.g., antibody-hapten pairs, may be substituted for avidin-biotin to achieve the same immobilization effect. The solid support may be of any of a variety of forms such as beads, sheets, membranes, chips, fiber, and the like. Similarly, the solid support may be formed of any of a number of materials compatible with immobilization of nucleic acids, including, but not limited to, glass or polymers, e.g., polystyrene, polyacrylamide, polycarbonate, polyethylene, polypropylene, agarose, and the like.

Most restriction endonucleases are suitable for use in the restriction endonuclease digestion steps of the subject method. Restriction endonucleases are widely available commercially, and procedures for using them are well known to persons of ordinary skill in the

IDT Exhibit 1020.8685

art of molecular biology. Suitable restriction endonucleases may produce either blunt ends or overhanging ends.

Type IIs restriction endonucleases may also be used as a restriction endonuclease in sequential restriction endonuclease digestion. Type IIs restriction endonuclease have recognition sites that are different than the cleavage site. Because Type IIs restriction endonucleases are of particular interest because they may be used to produce small representative restriction fragments of a uniform size because the property of type IIs enzymes to cleave at a fixed distance from the recognition site, irrespective of the cleavage sequence. To use a type IIs restriction endonuclease for sequential restriction endonuclease digestion, a first adapter having type II restriction endonuclease recognition site may be employed. For example, after digestion of immobilized cDNA fragments with a first restriction endonuclease, an adapter having a type IIs restriction endonuclease recognition site is ligated to the immobilized restriction fragments, the type IIs enzyme is then added, the dually digested restriction fragments are collected, and a second adapter is added (a polymerase catalyzed "filling-in" step may be used depending on the particular type IIs enzyme employed).

In order to maximize the degree of representation of an mRNA population among the representative restriction fragments. The initial cDNA preparation obtained from the cells of interest may be split into two portions and digested with the first and second restriction endonucleases in both possible temporal orders, i.e., representative restriction fragments are formed employing both possible orientations. For example, a cDNA population could be split into two fractions and immobilized at the 5' end on separate sets of streptavidin derivatized magnetic beads (sets A and B). Set A is first digested with Hind III, washed, and then digested with EcoRI. Set B is first digested with EcoRI, washed, and then digested with Hind III. Thus, the representative fragments from set A contains fragments from RNA transcripts that have an EcoRI site closer to the 5' end than the Hind II site. Similarly, the representative fragments from set B contains fragments from RNA transcripts that have a Hind III site closer to the 5' end than the EcoRI site. Thus, by splitting the cDNA preparation into two portions and digesting with restriction enzymes in both temporal order, the degree of representation may be greatly increased. Subsequent analysis by the subject methods employing internal fragment probes and terminus probes may be performed separately on the dual preparations of representative restriction fragments.

IDT Exhibit 1020.8686

Another method of deriving unique polynucleotide fragments from larger polynucleotides is described in Sherman in U.S. Patent No. 5,712,126 (Weisman and Prashar), which is incorporated by reference herein. An example of the method of producing representative restriction fragments as described in U.S. Patent No. 5,712,126, involves selectively amplifying the 3' portion of an mRNA by (a) priming first strand cDNA synthesis with a primer that hybridizes to a polyA tail and has a 5' non-hybridizing sequence region that can be used for priming amplification, (b) synthesizing the second strand, (c) cleaving the double-stranded cDNA with a restriction endonuclease, (d) ligating an adapter to the cleavage fragments, wherein the adapter consists of two partially hybridized nucleic acid strands, wherein portions of the two strands are non-complementary to each other and portions of the two strands are complementary to each other, (d) amplifying the ligated cleavage fragments using a first primer whose sequence comprises at least a portion of the 5' sequence of the oligonucleotide primer of step (a) and a second primer with a sequence that comprises at least a portion of the sequence of one strand of the adaptor in the non-complementary portion, thereby selectively amplifying a DNA fragment comprising sequence complementary to a 3' end of the mRNA.

Addition of Adapters

Adapters are joined to the termini of representative restriction fragments in most embodiments of the invention. The adapters may be joined to the termini in the same joining reaction or in two joining reactions (e.g., ligation) performed sequentially. Two adapters may be joined to a representative restriction, such that a single adapter is joined to each terminus. The two adapters are different from one another. However, the two adapters may be sufficiently similar to one another so as to permit annealing and amplification with a single primer oligonucleotide rather than a pair of two different primers. Methods for joining adapters to restriction fragments are well known to those of ordinary skill in the art. Guidance in using adapters can be found in, among other places, U. S. Patent No. 5,693,245.

Nucleic Acid Amplification

Adapter-modified representative restriction fragments may be amplified by a variety of primer-dependent polynucleotide amplification techniques. A variety of primer-dependent polynucleotide amplification techniques may be used for amplification. Such techniques include

IDT Exhibit 1020.8687

strand displacement amplification, 3SR amplification, and the like. The polymerase chain reaction (PCR) is particularly preferred for amplifying the adapter-modified representative restriction fragments. The polymerase chain reaction is described in, among other places, Diffenbach and Dveksler, PCR Primer Cold Spring Harbor Press, Cold Spring Harbor, NY (1995) and U.S. Patent No. 4,683,202; 4,683,195; 4,800,159; 4,965,188; and 5,333,675.

In embodiments of the invention employing polynucleotide amplification, the primers for use in the polynucleotide amplification primers are selected so a to work in conjunction with the adapters used in the given embodiment. One or more different primers may be used for a given adapter. The primers are selected so as to specifically anneal to portions of the adapters that have been joined to the representative restriction fragments.

Hybridization of Terminus probes to Representative Restriction Fragments

A terminus probe of a known sequence may be hybridized to a representative restriction fragment (including adapter-modified representative restriction fragments and amplification products thereof) using conventional nucleotide acid hybridization techniques. Examples of nucleic acid hybridization techniques can be found, among other places in Sambrook et al., Molecular Cloning Cold Spring Harbor Press, Cold Spring Harbor, NY (1989). Nucleic acid hybridization parameters, e.g., time, temperature, salt concentrations, etc., may be varied to optimize the desired result.

Arrays

Other embodiments of the invention include oligonucleotide arrays having a set of oligonucleotide features, wherein the oligonucleotides of each feature have a constant region and variable region. The variable region is joined directly to the constant region at the 3' end of the constant region. The set of oligonucleotide features on an array of the invention includes all possible variations (i.e., a complete set of variations) of the nucleotide sequence of the variable region. The constant region is functionally constant between the different features of the subject arrays. The term "functionally constant" as used herein refers the property of either (1) being identical, or (2) being identical with respect to the portion of the constant region that is designed to hybridize to a representative restriction fragment or an adapter-modified representative restriction fragment. In the second case, additional nucleotide may be present on the 5' end of

IDT Exhibit 1020.8688

the oligonucleotides that form a feature. Both the constant region and the variable may vary in length in accordance with the particular embodiment of the invention. The length of the variable region is preferably between 3 and 12 nucleotides in length, 3-6 being more preferred. In a most preferred embodiment of the invention, the variable region of the features is five bases in length. Because a complete set of features comprise all possible nucleotide base sequence variations of the variable region, the number of features that constitute a complete set will vary in accordance with the (the number of possible bases raised to the power of the length in a nucleotide). Thus, in embodiments of the invention employing a five base variable region, a set of array features comprises $4^5 = 1024$ features. The length of the constant region is sufficient to hybridize to the selected restriction endonuclease recognition site of representative restriction fragments and to either all or part of the strand of the adapter joined to the selected restriction endonuclease recognition site of the representative restriction fragments and to either all or part of the strand of the adapter joined to the selected restriction endonuclease recognition site of a representative restriction fragment. The nucleotide base sequence of the constant region may or may not be preferably complementary to the relevant portion of the adapter modified representative restriction fragments. The constant regions are of sufficient length and of the proper nucleotide base composition to permit them to hybridize to a selected end of the adapter-modified representative restriction fragments. Upon hybridization with an adapter-modified representative restriction fragment, the double-stranded region formed between the adapter-modified representative restriction fragment and the constant region of the feature terminates at the nucleotide of restriction recognition site of the hybridized representative restriction fragment.

The arrays of the invention may comprise one or more subarrays. The subarrays that constitute a large array may each comprise essentially the same set of oligonucleotide features. The spatial arrangement of the oligonucleotide features of the different subarrays that constitute the larger array may be the same or different than each other. the subarrays may be physically isolated from one another so as to permit the application of one labeled oligonucleotide probe solution at one array without carrying over to other subarrays.

Kits

The invention also includes kits for performing one or more of the different methods for analyzing polynucleotide population described herein. Kits generally contain two or more

IDT Exhibit 1020.8689

reagents necessary to perform the subject methods. The reagents may be supplied in pre-measured amount for individual assays so as to increase reproducibility.

In one embodiment, the subject kits comprise adapters and primers for use with adapters to amplify adapter-modified restriction fragments. Kits may further comprise arrays of terminus probes or sets of internal fragment probes for use in conjunction with the adapters. Other embodiments of the subject kits include arrays of terminus probes and sets of internal fragment probes for use in conjunction with the terminus probes of the array. Still other embodiments of the subject kits include kits that comprise (1) sorting-signal receptor arrays and (2) terminus probes or internal fragment probes, wherein the probes are appropriately marked with sorting signals for use in conjunction with the sorting signal receptor array. Probes of the subject kits may be marked with detectable signals suitable for use in the subject methods. The kits of the invention may also include one or more additional reagents required for various embodiments of the subject methods. Such additional reagents include, but are not limited to: restriction enzymes, DNA polymerases, buffers, nucleotides, and the like.

INCORPORATION BY REFERENCE

All publications, patent applications, and patents referenced in the 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.

EQUIVALENTS

All publications, patent applications, and patents mentioned in this specification are indicative of the level of skill of those skilled in the art to which this invention pertains. Although only a few embodiments have been described in detail above, those having ordinary skill in the molecular biology art will clearly understand that many modifications are possible in the preferred embodiment without departing from the teachings thereof. All such modifications are intended to be encompassed within the following claims. The foregoing written specification is considered to be sufficient to enable skilled in the art to which this invention pertains to practice the invention. Indeed, various modifications of the above-described modes for carrying out the invention which are apparent to those skilled in the field of molecular biology or related fields are intended to be within the scope of the following claims.

IDT Exhibit 1020.8690

CLAIMS

What is claimed is:

1. A method of analyzing a polynucleotide, said method comprising the steps.

forming a representative restriction fragment corresponding to the polynucleotide, wherein the representative restriction fragment has a first and second terminus and at least one of the termini is generated by a restriction endonuclease,

hybridizing a terminus probe to a single strand of the restriction fragment at a position on the restriction fragment including the terminus generated by the restriction endonuclease,

hybridizing an internal fragment probe to the single strand of restriction fragment at a position adjacent to the terminus probe, and

joining the terminus probe to the internal fragment probe.

2. The method according to **claim 1**, said method further comprising the step of,

combining (i) the nucleotide sequence information from the terminus probe with (ii) the nucleotide sequence information from the internal fragment probe and (iii) the nucleotide sequence of the recognition site of the restriction endonuclease used to produce the terminus, so as to produce an identifier sequence.

3. The method according to **claim 1** wherein the polynucleotide is a cDNA.

4. The method according to **claim 2** further comprising the step of comparing the identifier sequence with a DNA sequence database.

IDT Exhibit 1020.8691

5. The method of **claim 2** further comprising the steps of preparing a oligonucleotide hybridization probe comprising a base sequence encoding the identifier sequence.

6. A method according to **claim 1**, wherein the terminus probe is a feature of an oligonucleotide array.

7. A method according to **claim 1** further comprising the steps of joining first and second adapters to the first and second termini of the representative restriction fragment, respectively, whereby an adapter-modified representative restriction fragment is formed, wherein the joining of the adapters occurs before the hybridizing of the terminus probe and the internal probe.

8. A method according to **claim 7**, further comprising the step of amplifying the adapter-modified representative restriction fragment, whereby amplification products of the adapter-modified representative restriction fragments are formed.

9. A method according to **claim 8**, wherein the step of hybridizing the terminus probe comprises

contacting the amplification products with an oligonucleotide array under nucleic acid hybridization conditions, wherein the array comprises features that are terminus probes, said terminus probes having constant and variable regions, whereby a strand of the amplification product is hybridized to a terminus probe.

10. A method according to **claim 9**, wherein the step of hybridizing the internal fragment probe comprises,

contacting the amplification product strand hybridized to a terminus probe with a solution comprising an internal fragment probe, wherein the contact occurs under nucleic acid hybridization conditions.

IDT Exhibit 1020.8692

11. The method according to **claim 10**, wherein the solution comprises a plurality of internal fragment probes.

12. The method of **claim 11**, wherein each of the internal fragment probes is labeled with a different fluorescent label.

13. A method according to **claim 12**, further comprising the step of of contacting the hybridized amplification products with a second solution comprising at least one internal fragment probe that has a different nucleotide sequence than an internal fragment probe in the first solution.

14. A method according to **claim 13**, wherein the second solution comprises a plurality of different terminus probes , wherein each terminus probe in the second solution is labeled with a distinctive label.

15. A method according to **claim 11**, wherein the array comprises at least one subarray of 1024 distinct features, the variable region of each feature is five nucleotides in length, and each of the variable regions of the subarray has a different nucleotide sequence from the other variable regions of the terminus probes in the same subarray.

16. A method according to **claim 12**, wherein the constant region of the feature are at least 4 nucleotides in length.

17. A method according to **claim 12**, wherein the constant regions of the features are identical to each other.

18. A method according to **claim 12**, wherein the array comprises a plurality of subarrays wherein at least two of the subarrays comprise the same set of features.

IDT Exhibit 1020.8693

19. A method according to **claim 18**, wherein the oligonucleotide probe solution comprises a plurality of internal probes and each different probe is labeled with a distinct identification label.

20. A method according to **claim 19**, wherein the oligonucleotide probes are labeled with different fluorescent labels.

21. A method according to **claim 19**, further comprising the step of of contacting the hybridized amplification products with a second solution comprising at least one internal fragment probe that has a different nucleotide sequence than an internal fragment probe in the first solution.

22. A method according to **claim 21**, wherein the second solution comprises a plurality of different terminus probes , wherein each terminus probe in the second solution is labeled with a distinctive label.

23. A method according to **claim 7**, wherein the internal fragment probe or the terminus probe comprises an array sorting signal.

24. A method according to **claim 23**, comprising the step of contacting the adapter-modified representative restriction fragment with an array comprising a plurality sorting signal receptors at a predetermined locations on the array.

25. A method according to **claim 23**, wherein the array sorting signal and the sorting signal receptors are polynucleotides.

26. A method according to **claim 1**, wherein the representative restriction fragment is generated by a method comprising the steps,

immobilizing the polynucleotide on a solid support,

contacting the polynucleotide with a first restriction endonuclease, whereby an immobilized restriction fragment is produced, and

IDT Exhibit 1020.8694

purifying the immobilized restriction fragment.

27. The method of **claim 26**, further comprising the steps of

contacting the immobilized restriction fragment with a second restriction endonuclease, whereby the representative restriction fragment is produced, and

purifying the representative restriction fragment.

28. The method according to **claim 2**, further comprising the steps,

joining a linker to the terminus produced by the first restriction enzyme on the immobilized restriction fragment, whereby an adapter-modified immobilized restriction fragment was produced, and

contacting the adapter-modified immobilized restriction fragment with a second restriction endonuclease, whereby the representative restriction fragment is produced.

29. The method of **claim 28**, wherein the adapter comprises a type IIS restriction site and the second restriction endonuclease is a type IIS restriction endonuclease recognizes the type IIS restriction site in the adapter and cleaves within the immobilized restriction fragment.

30. A polynucleotide population analysis kit, said kit comprising,

an oligonucleotide array comprising a plurality of terminus probes,

a plurality of internal probes.

31. The kit according to **claim 30**, wherein the internal probes are in a first solution

32. The kit according to **claim 30**, said kit further comprising a second solution, wherein the second solution comprises a plurality of internal probes, each of the internal probes in the second solution is labeled with a distinctive label, and at least one of the

IDT Exhibit 1020.8695

internal probes in the second solution has a different nucleotide base sequence than the internal probes in the first solution.

33. The kit according to **claim 30**, wherein the array comprises a plurality of subarrays.

34. A polynucleotide population analysis kit, said kit comprising,

a sorting array comprising a plurality of sorting signal receptors,

a plurality of terminus probes marked with sorting signal, wherein the sorting signal are specific for the sorting signal receptors on the sorting array,

a plurality of internal fragment probes labeled with a detectable labeled, wherein at least two of the internal fragment probes are labeled with different detectable labels.

35. A polynucleotide population analysis kit, said kit comprising,

a sorting array comprising a plurality of sorting signal receptors,

a plurality of internal fragment probes marked with sorting signal, wherein the sorting signal are specific for the sorting signal receptors on the sorting array,

a plurality of terminus probes labeled with a detectable labeled, wherein at least two of the internal fragment probes are labeled with different detectable labels.

IDT Exhibit 1020.8696

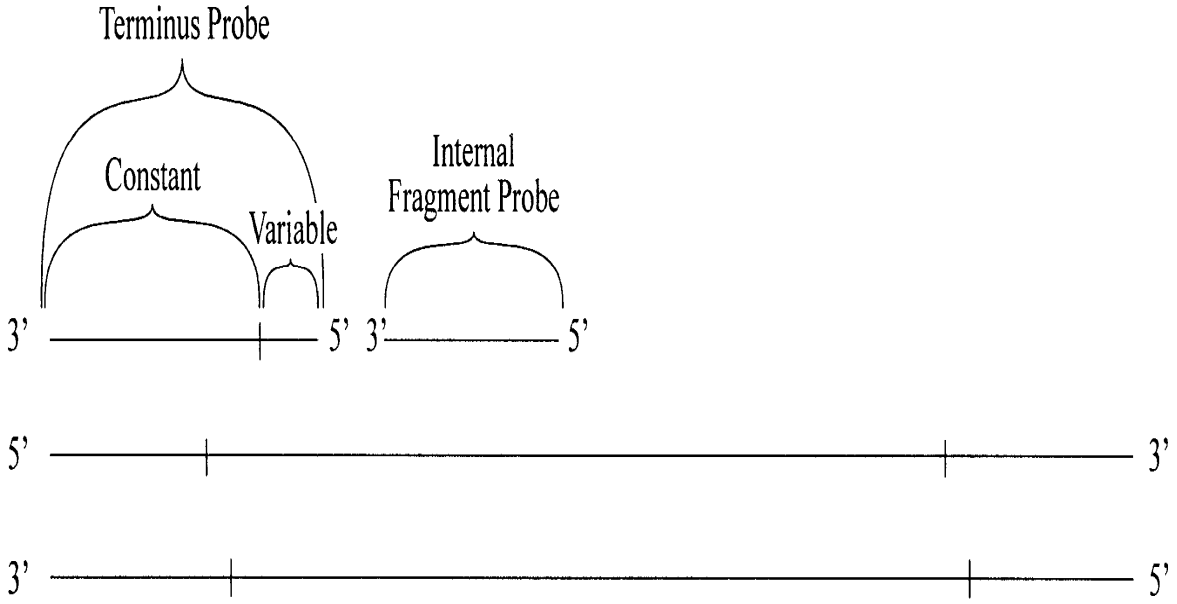

Terminus Probe
Constant
Variable
Internal
Fragment Probe
3'
5'
3'
5'
5'
3'
3'
5'

Fig. 1A

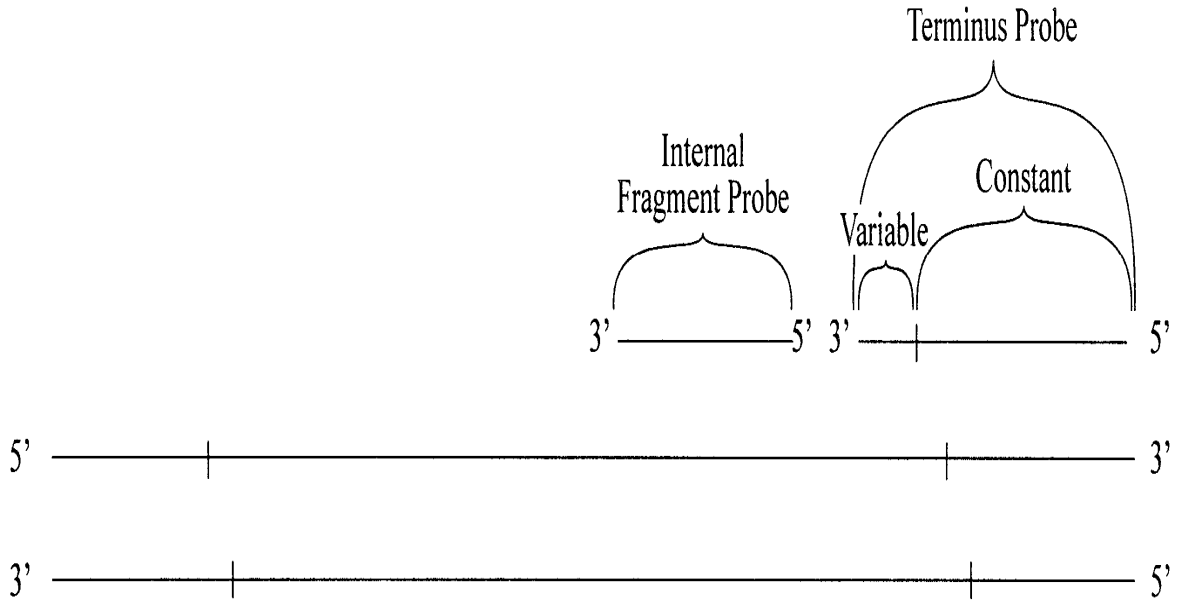

Terminus Probe
Internal
Fragment Probe
Constant
Variable
3'
5'
3'
5'
5'
3'
3'
5'

Fig. 1B

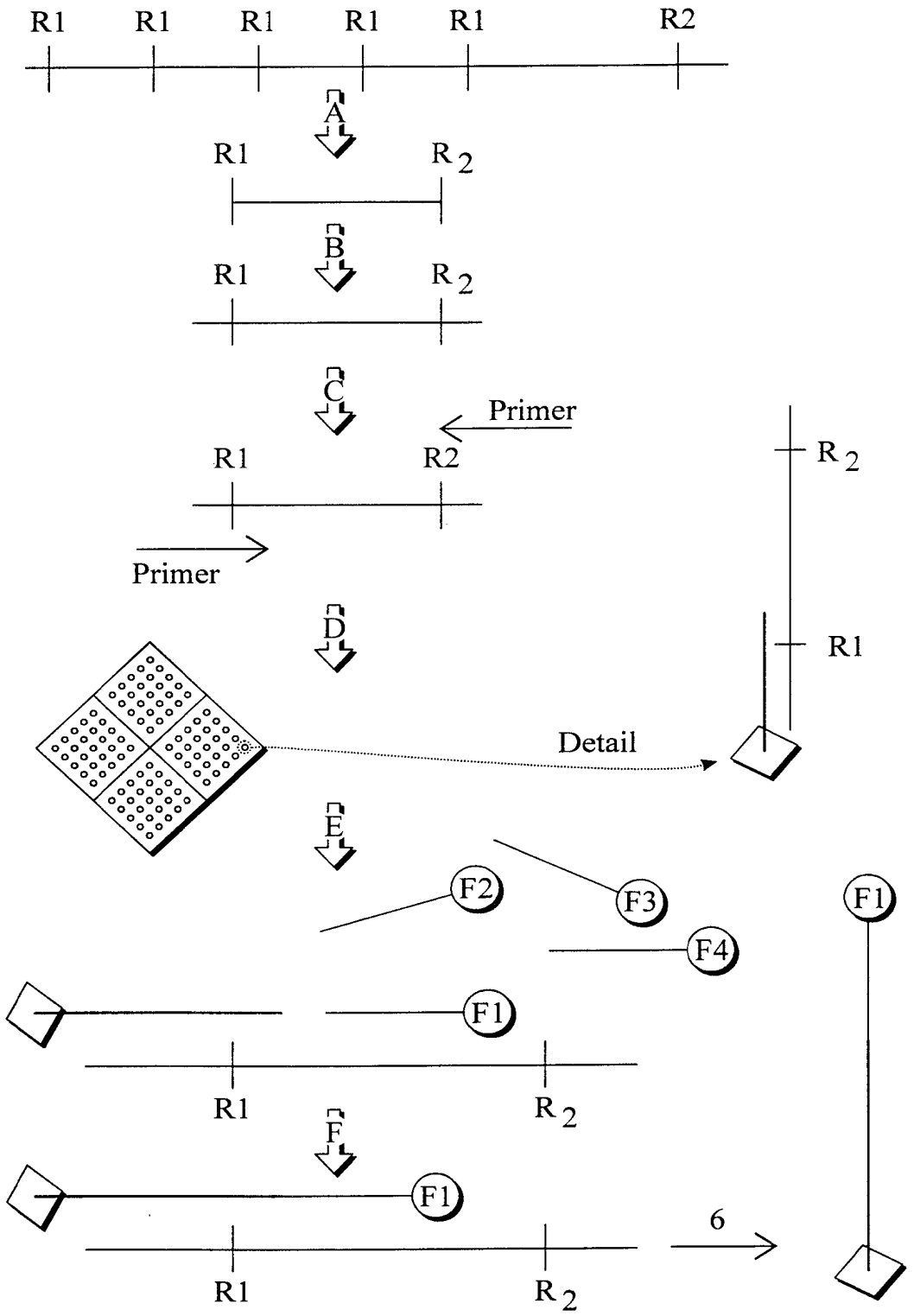


Fig. 2

IDT Exhibit 1020.8698

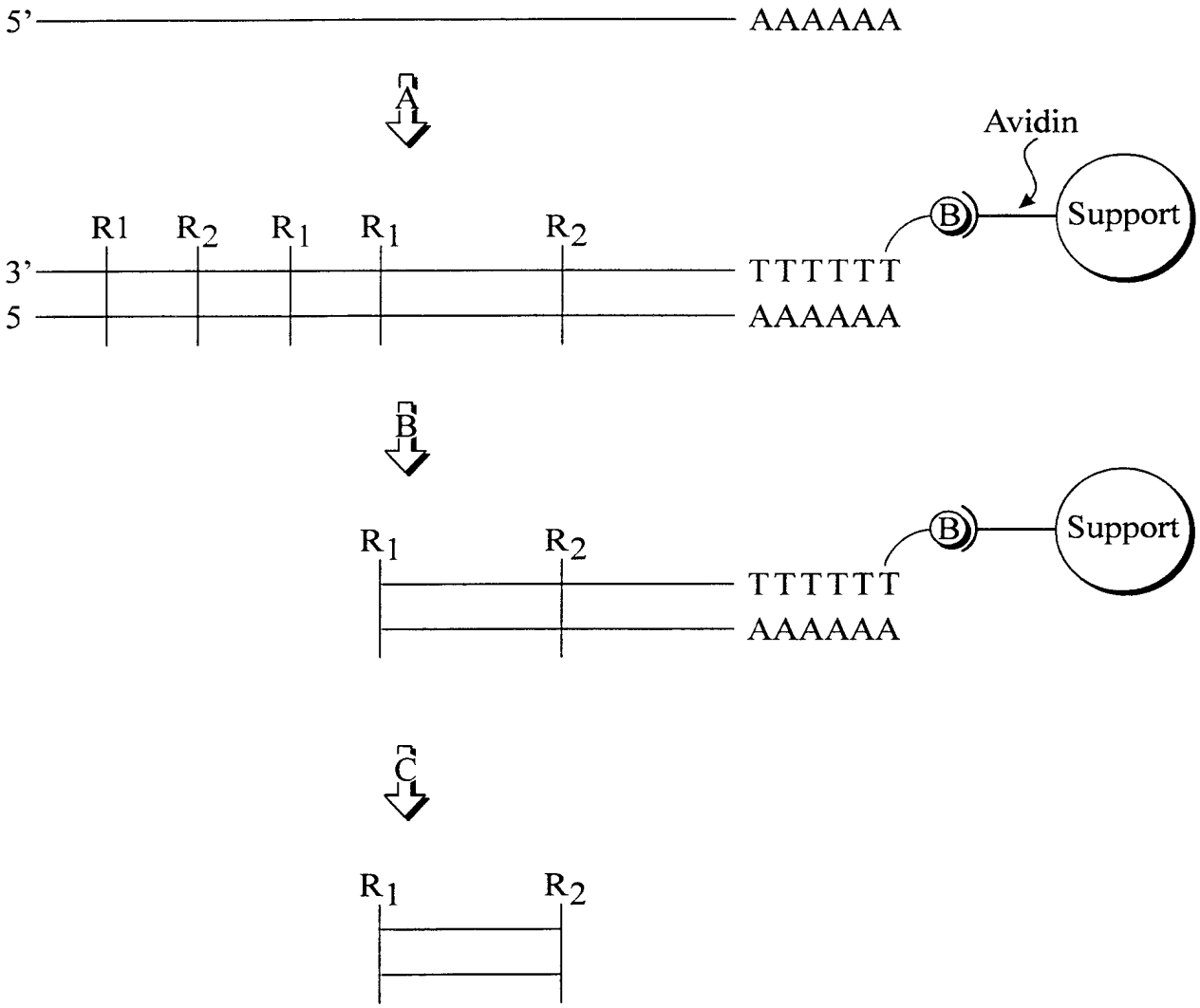


Fig. 3

IDT Exhibit 1020.8699

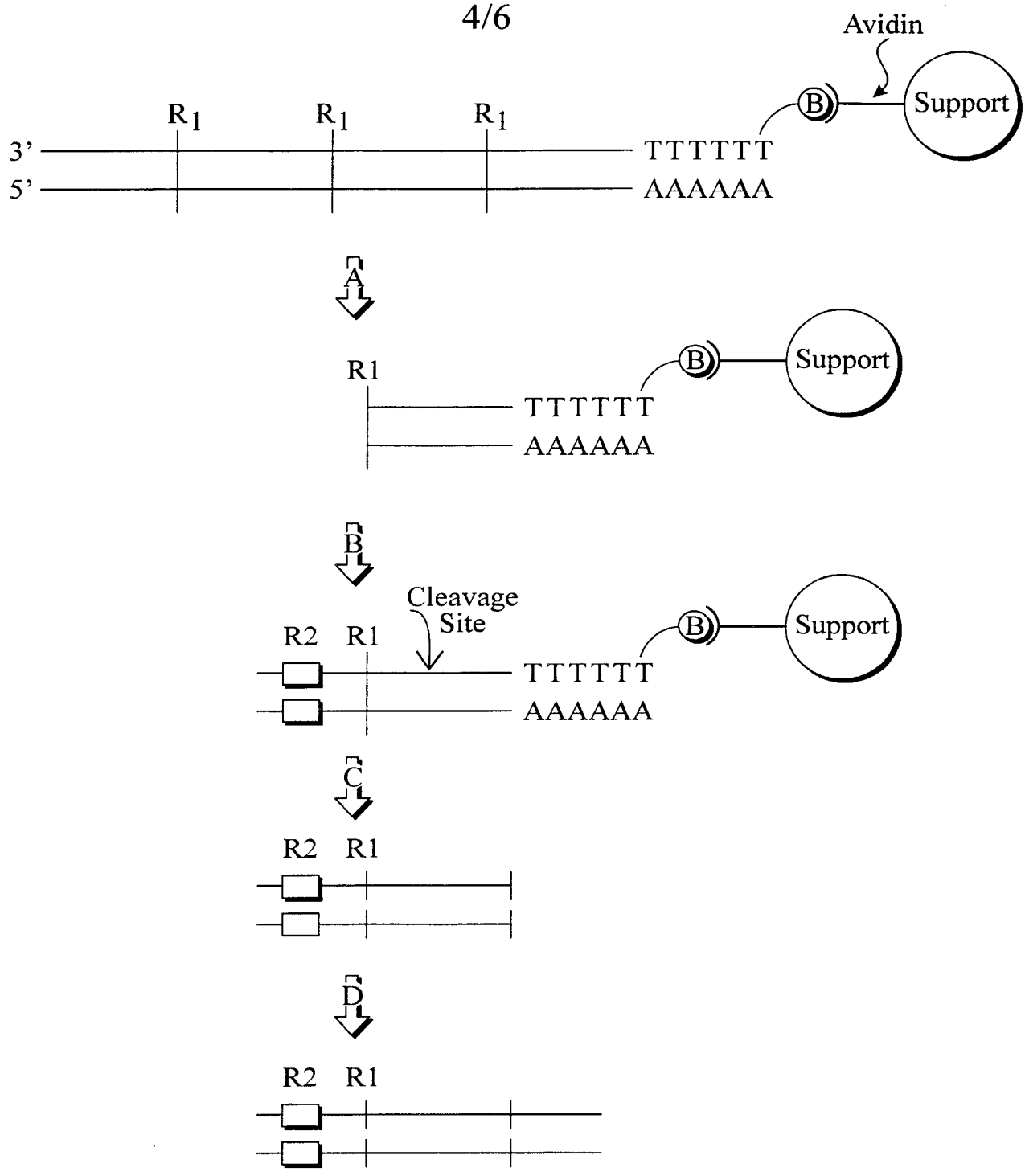


Fig. 4

IDT Exhibit 1020.8700

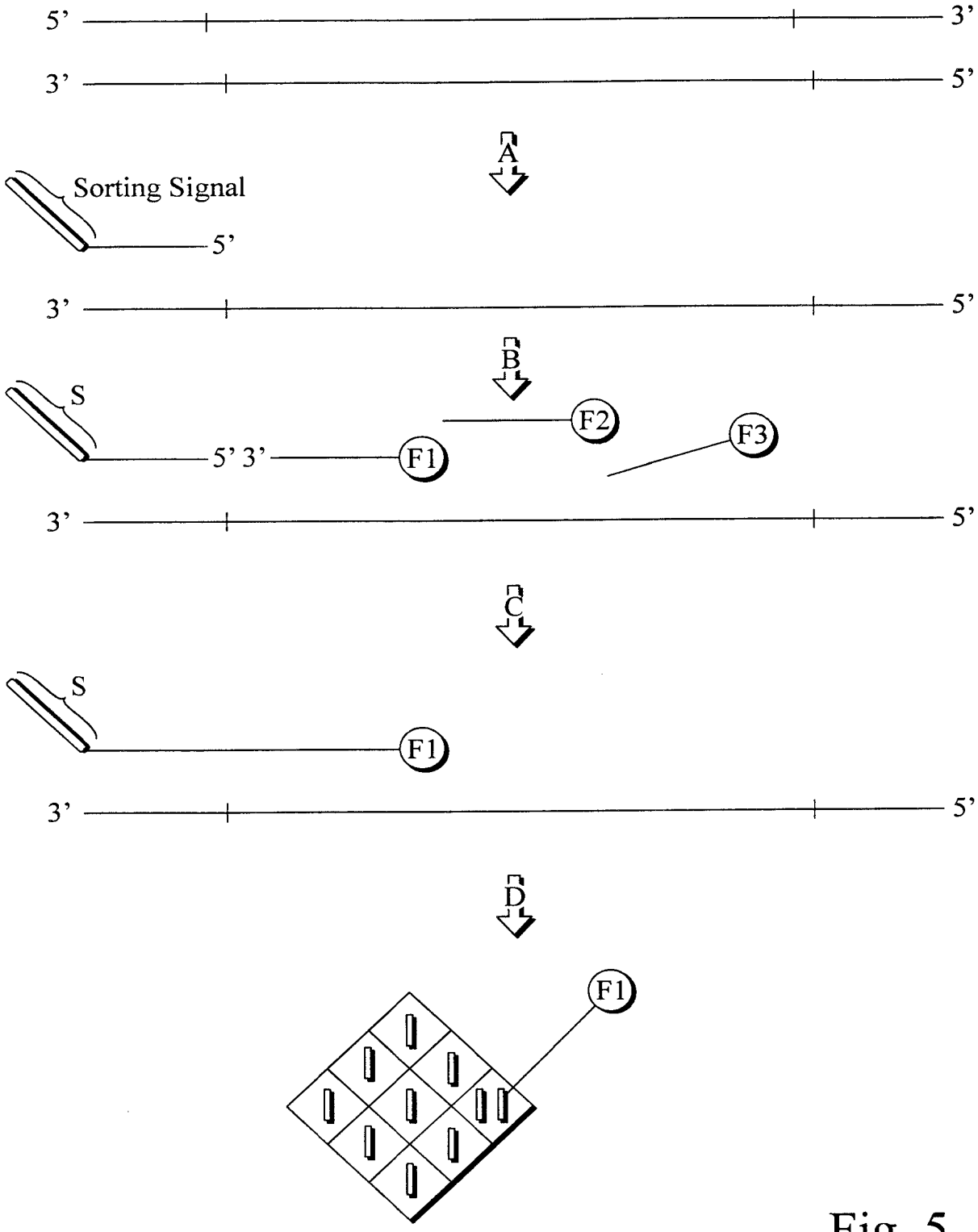


Fig. 5

IDT Exhibit 1020.8701

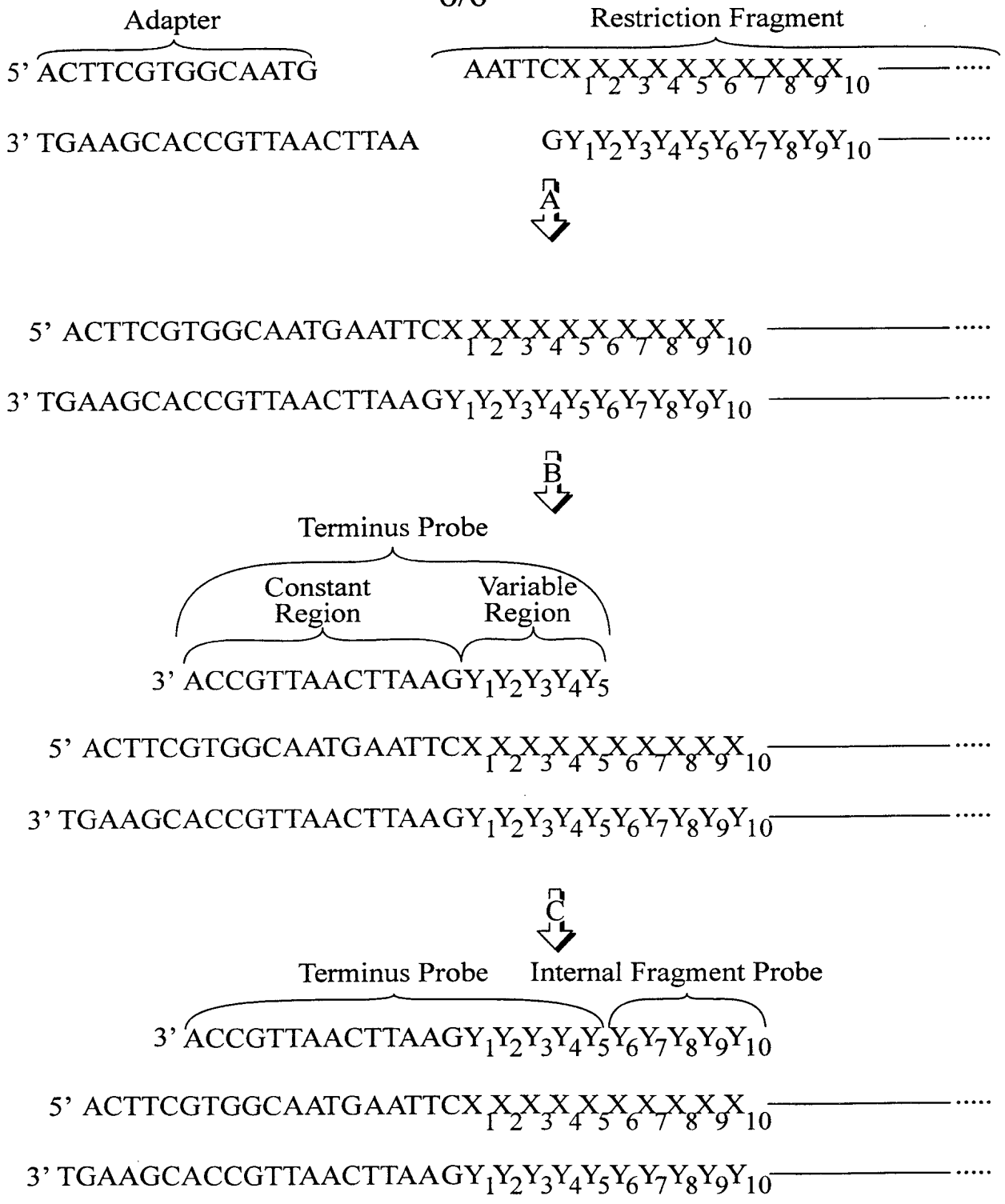


Fig. 6

IDT Exhibit 1020.8702

INTERNATIONAL SEARCH REPORT

ernational Application No
PCT/US 99/18422

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12Q

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 97 27317 A (CHEE MARK ;LAI CHAOQIANG (US); LEE DANNY (US); AFFYMETRIX INC (US)) 31 July 1997 (1997-07-31) See "VIII. Ligation enhanced signal detection", page 70-page 83 the whole document	1-35
Y	US 5 710 000 A (GINGERAS THOMAS R ET AL) 20 January 1998 (1998-01-20) the whole document	1-35
Y	WO 96 41011 A (SPECTRAGEN INC) 19 December 1996 (1996-12-19) the whole document	1-35
Y	EP 0 735 144 A (JAPAN RES DEV CORP) 2 October 1996 (1996-10-02) the whole document	1-35

-/--

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

[X] Patent family members are listed in annex.

° Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
17 January 2000	24/01/2000

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

1

Form PCT/ISA/210 (second sheet) (July 1992)

IDT Exhibit 1020.8703

INTERNATIONAL SEARCH REPORT

ernational Application No
PCT/US 99/18422

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 97 31256 A (BLOK HERMAN ;BARANY GEORGE (US); KEMPE MARIA (US); ZIRVI MONIB (US) 28 August 1997 (1997-08-28) the whole document	1-35
A	US 5 242 794 A (HUNKAPILLER MICHAEL W ET AL) 7 September 1993 (1993-09-07) the whole document	
A	WO 95 04160 A (ISIS INNOVATION ;SOUTHERN EDWIN (GB); CUMMINS WILLIAM JONATHAN (GB) 9 February 1995 (1995-02-09) the whole document	
A	UNRAU P ET AL: "NON-CLONING AMPLIFICATION OF SPECIFIC DNA FRAGMENTS FROM WHOLE GENOMIC DNA DIGEST USING DNA INDEXERS" GENE,NL,ELSEVIER BIOMEDICAL PRESS. AMSTERDAM, vol. 145, 1994, page 163-169 XP002056703 ISSN: 0378-1119 the whole document	
A	WO 94 01582 A (MEDICAL RES COUNCIL ;SIBSON DAVID ROSS (GB)) 20 January 1994 (1994-01-20) the whole document	

1

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

IDT Exhibit 1020.8704

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 99/18422

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9727317 A	31-07-1997	AU 2253397 A	20-08-1997
		EP 0880598 A	02-12-1998
US 5710000 A	20-01-1998	NONE	
WO 9641011 A	19-12-1996	US 5635400 A	03-06-1997
		AU 6102096 A	30-12-1996
		CA 2222581 A	19-12-1996
		CN 1193357 A	16-09-1998
		CZ 9703926 A	17-06-1998
		EP 0832287 A	01-04-1998
		HU 9900910 A	28-07-1999
		JP 11507528 T	06-07-1999
		NO 975744 A	05-02-1998
		PL 324000 A	27-04-1998
		AU 712929 B	18-11-1999
		AU 4277896 A	06-05-1996
		AU 7717596 A	30-04-1997
		CZ 9700866 A	17-09-1997
		EP 0793718 A	10-09-1997
		EP 0931165 A	28-07-1999
		FI 971473 A	04-06-1997
		JP 10507357 T	21-07-1998
		NO 971644 A	02-06-1997
		WO 9713877 A	17-04-1997
EP 0735144 A	02-10-1996	JP 2763277 B	11-06-1998
		JP 9028399 A	04-02-1997
		JP 2763278 B	11-06-1998
		JP 8322598 A	10-12-1996
		AU 692685 B	11-06-1998
		AU 5031196 A	10-10-1996
		US 5707807 A	13-01-1998
WO 9731256 A	28-08-1997	AU 2799797 A	10-09-1997
		CA 2244891 A	28-08-1997
		EP 0920440 A	09-06-1999
US 5242794 A	07-09-1993	US 4883750 A	28-11-1989
		US 5521065 A	28-05-1996
		US 5962223 A	05-10-1999
		DE 3586090 A	25-06-1992
		EP 0185494 A	25-06-1986
		JP 1919077 C	07-04-1995
		JP 6044880 B	15-06-1994
		JP 61191964 A	26-08-1986
WO 9504160 A	09-02-1995	AT 159767 T	15-11-1997
		AU 695349 B	13-08-1998
		AU 7269194 A	28-02-1995
		CA 2168010 A	09-02-1995
		CN 1131440 A	18-09-1996
		DE 69406544 D	04-12-1997
		DE 69406544 T	26-02-1998
		DK 711362 T	22-12-1997
		EP 0711362 A	15-05-1996
		EP 0778280 A	11-06-1997
		ES 2108479 T	16-12-1997

IDT Exhibit 1020.8705

INTERNATIONAL SEARCH REPORT

Information on patent family members

ernational Application No
PCT/US 99/18422

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9504160 A		FI 960403 A	29-01-1996
		HU 73802 A	30-09-1996
		JP 9501830 T	25-02-1997
		NO 960370 A	28-03-1996
		US 5770367 A	23-06-1998
WO 9401582 A	20-01-1994	AT 159986 T	15-11-1997
		AU 686563 B	12-02-1998
		AU 4575893 A	31-01-1994
		CA 2139944 A	20-01-1994
		DE 69315074 D	11-12-1997
		DE 69315074 T	05-03-1998
		EP 0650528 A	03-05-1995
		JP 7508883 T	05-10-1995
		US 5728524 A	17-03-1998

IDT Exhibit 1020.8706

Espacenet

Bibliographic data: WO0200938 (A2) — 2002-01-03

METHODS AND COMPOSITIONS FOR TRANSCRIPTION-BASED NUCLEIC ACID AMPLIFICATION

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

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

Classification: - international: ***C12M1/00; C12N15/09; C12Q1/68; G01N33/53; G01N37/00;*** (IPC1-7): C12Q1/68
- cooperative: **C12Q1/6865**; C12Q1/6876 more

Application number: WO2001US20660 20010626 Global Dossier

Priority number (s): US20000213908P 20000626 ; US20010277748P 20010321

Also published as: WO0200938 (A3) WO0200938 (A9) AT455186 (T) AU7159501 (A) CA2412721 (A1) more

Abstract of WO0200938 (A2)

Methods for isothermal exponential amplification of a target polynucleotide are disclosed. The methods employ two transcription modules, the first module providing linear amplification resulting in RNA transcripts, and a second module providing for further (generally cyclical) amplification resulting in more RNA transcripts. In one aspect, the amplification for the first module is composite primer based. In a second aspect, the amplification of the first module is based on target switching to generate a primer extension product comprising a promoter sequence. In all aspects, the RNA transcripts of the first transcription module are subjected to further amplification by creating an intermediate product comprising a double stranded promoter region from which transcription can occur. The invention further provides compositions and kits for practicing said methods, as well as methods which use the amplification results.

IDT Exhibit 1020.8707

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 January 2002 (03.01.2002)

PCT

(10) International Publication Number
WO 02/00938 A2

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/US01/20660

(22) International Filing Date: 26 June 2001 (26.06.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/213,908 26 June 2000 (26.06.2000) US
60/277,748 21 March 2001 (21.03.2001) US

(71) Applicant *(for all designated States except US)*: **NUGEN TGECHNOLOGIES, INC.** [US/US]; 11 Bennett Road, Redwood City, CA 94062 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **KURN, Nurith** [IL/US]; 2876 Ramona Street, Palo Alto, CA 94306 (US).

(74) Agents: **POLIZZI, Catherine, M.** et al.; Morrison & Foerster LLP, 755 Page Mill Road, Palo Alto, CA 94304-1018 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, 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 02/00938 A2

(54) Title: METHODS AND COMPOSITIONS FOR TRANSCRIPTION-BASED NUCLEIC ACID AMPLIFICATION

(57) Abstract: Methods for isothermal exponential amplification of a target polynucleotide are disclosed. The methods employ two transcription modules, the first module providing linear amplification resulting in RNA transcripts, and a second module providing for further (generally cyclical) amplification resulting in more RNA transcripts. In one aspect, the amplification for the first module is composite primer based. In a second aspect, the amplification of the first module is based on target switching to generate a primer extension product comprising a promoter sequence. In all aspects, the RNA transcripts of the first transcription module are subjected to further amplification by creating an intermediate product comprising a double stranded promoter region from which transcription can occur. The invention further provides compositions and kits for practicing said methods, as well as methods which use the amplification results.

IDT Exhibit 1020.8708

METHODS AND COMPOSITIONS FOR TRANSCRIPTION-BASED NUCLEIC ACID AMPLIFICATION

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the priority benefit of the provisional patent applications U.S. Serial No. 60/213,908, filed June 26, 2000, and U.S. Serial No. 60/277,748, filed March 21, 2001, both of which are incorporated herein by reference in their entirety.

TECHNICAL FIELD

This invention relates to the filed of polynucleotide amplification. More particularly, the invention provides methods, compositions and kits for amplifying (i.e., making multiple copies) target polynucleotide sequences which involve transcription, and employ either an RNA/DNA composite primer and/or a target switch polynucleotide.

BACKGROUND ART

The development of methods for nucleic acid amplification and detection of amplification products have advanced the detection, identification, quantification and sequence analyses of nucleic acids in recent years. Use of these methods hascontributed to rapid advances in the areas of genomics, cell biology, molecular medicine, genetics and the like.

Nucleic acid analysis is widely used for detection and identification of pathogens, detection of gene alterations leading to defined phenotypes, diagnosis of genetic diseases or susceptibility to such disease, assessment of gene expression in development, in disease and in response to defined stimuli, as well as in various genome projects. Other applications of nucleic acid amplification methods include the detection of rare cells, detection of pathogens and the detection of altered gene expression in malignancy, and the like. Nucleic acid amplification is potentially useful for both qualitative analysis such as the detection of the presence of defined nucleic acid sequences, and quantification of defined gene sequences. The latter is useful for assessing and determining the amount of pathogenic sequences in a sampleas well as for the determination of gene multiplication or deletion, as often found in cell transformation from normal to malignant type.

Although detection of the presence of a defined nucleic acid sequence, and its sequence analysis, can be carried out by direct probe hybridization to target nucleic acid

sequences, the method generally lacks sensitivity when amounts of the target nucleic acid sequence present in the test sample are low. One solution to this obstacle was the development of methods for generation of multiple copies of the defined nucleic acid sequence to make them more accessible to further analysis. Methods for generating multiple copies of a specific nucleic acid sequence in a sample are generally defined as target amplification methods. Other methods for increasing the sensitivity of detection of hybridization analysis are based on the generation of multiple products from the hybridized probe(s), for example, cleaving the hybridized probe to form multiple products or ligating adjacent probes to form a unique hybridization-dependent product. Similarly, increased sensitivity of hybridization reaction is achieved by methods for amplifying signals generated by the hybridization event, such as methods based on hybridization of branched DNA probes.

Various target nucleic acid amplification methods have been described in recent years. Target nucleic acid amplification is carried out through multiple cycles of incubations at various temperatures (thermal cycling) or alternatively, carried out by an isothermal process. The discovery of thermostable nucleic acid modifying enzymes has also contributed to rapid advances in nucleic acid amplification technology. Thermostable nucleic acid modifying enzymes, such as DNA and RNA polymerases, ligases, nucleases and the like, are used both in methods dependent on thermal cycling and isothermal amplification methods. For example, a method for "homogeneous isothermal amplification and detection of nucleic acids using a template switch oligonucleotide" is described in WO/070095A2 (Liu, et al.)

The most commonly used target amplification method is the polymerase chain reaction (PCR) which is based on multiple cycles of denaturation, hybridization of two oligonucleotide primers, one to each strand of a double stranded target , and primer extension by a nucleotide polymerase to produce multiple double stranded copies of the target sequence. Many variations of PCR have been described, and the method is being used for amplification of DNA or RNA nucleic acid sequences, sequence determination, mutation analysis and others. (*see* PCR Protocols: A Guide to Methods and Applications (Innis, M., Gelfand, D., Sninsky, J. and White, T., eds.) Academic Press (1990); Mullis et al., Methods in Enzymology, 155:335-350 (1987)). Thermocycling-based methods that employ a single primer are also described. Other methods that depend on thermal cycling include the ligase chain reaction (LCR) and the related repair chain reaction (RCR).

IDT Exhibit 1020.8710

Isothermal nucleic acid amplification methods based on strand displacement, are described. See, for e.g., Fraiser et al. in U.S. Patent No. 5,648,211; Cleuziat et al. in U.S. Patent No. 5,824,517; and Walker et al. *Proc. Natl. Acad. Sci. U.S.A.* 89:392-396 (1992). Other isothermal target amplification methods are the transcription-based amplification methods, in which an RNA polymerase promoter sequence is incorporated into primer extension products at an early stage of the amplification (WO/01050), and target sequence, or target complementary sequence, is further amplified by transcription and digestion of the RNA strand in a DNA/RNA hybrid intermediate product. See, for example, U.S. Patent Nos. 5,169,766 and 4,786,600. Target nucleic acid amplification may be carried out through multiple cycles of incubations at various temperatures, i.e. thermal cycling, or at one temperature (an isothermal process). These methods include transcription-mediated amplification (TMA), self-sustained sequence replication (3SR), nucleic acid sequence-based amplification (NASBA), and variations thereof. See, for example, Guatelli et al. *Proc. Natl. Acad. Sci. U.S.A.* 87:1874-1878 (1990); U.S. Patent Nos. 5,766,849 (TMA); and 5,654,142 (NASBA). Other amplifications methods use template switching oligonucleotides (TSOs) and blocking oligonucleotides. For example, a template switch amplification method in which chimeric DNA primers are utilized is disclosed in U.S. Patent 5,679,512 and by Patel et al. (*Proc. Natl. Acad. Sci. U.S.A.* 93:2969-2974 (1996)), and a method that uses blocking oligonucleotides is disclosed by Laney et al. in U.S. Patent No. 5,679,512.

Since isothermal target amplification methods do not require a thermocycler, they are easier to adapt to common instrumentation platforms. However, previously known isothermal target amplification methods have severe drawbacks. Amplification according to the strand displacement amplification (SDA) process requires the presence of sites for defined restriction enzymes, thus limiting its applicability. Existing transcription-based amplification methods, such as the NASBA and TMA, on the other hand, are limited by the need for incorporation of the RNA polymerase promoter sequence into the amplification product by means of a primer, a process prone to causing non-specific amplification. Moreover, the mechanism of amplification of a DNA target by these transcription-based amplification methods is not well established.

The completion of sequencing of a number of genomes, and the sequencing of the human genome in particular, has tremendous implications for advances in molecular and cell biology in general and molecular medicine in particular. Development of methods for

IDT Exhibit 1020.8711

isothermal gene sequencing will greatly enhance the application of this information in various testing facilities, as isothermal sequencing markedly simplifies the process of genetic identification compared to current methods that require thermocycling.

The methods of the present invention provide for isothermal, high efficiency nucleic acid sequence amplification and methods that use these amplification methods and products, such as in sequence determination.

All references cited herein, including patent applications and publications, are incorporated by reference in their entirety.

DISCLOSURE OF THE INVENTION

The invention provides methods, compositions and kits for amplifying polynucleotide sequences. The methods generally comprise use of a primer (which in some aspects of the invention is an RNA/DNA composite primer), optionally a termination sequence, and a propromoter oligonucleotide sequence.

Accordingly, in one aspect the invention provides methods for generating multiple copies of a nucleic acid sequence of interest, comprising the steps of: (a) hybridizing a single stranded target polynucleotide comprising the nucleic acid sequence of interest with a first primer, wherein said first primer is a composite primer comprising an RNA portion and a 3' DNA portion; (b) optionally hybridizing a polynucleotide comprising a termination polynucleotide sequence to a region of the target polynucleotide 5' with respect to hybridization of the first primer to the target polynucleotide; (c) extending the first primer with a DNA-dependent DNA polymerase to generate a complex comprising a first primer extension product and target polynucleotide; (d) cleaving the RNA portion from the composite primer in the complex of first primer extension product and target polynucleotide with an enzyme that cleaves RNA from an RNA/DNA hybrid such that another composite primer (which is generally and preferably the same as the first primer) can hybridize to the target polynucleotide and repeat primer extension by strand displacement to produce displaced primer extension product; (e) hybridizing a propromoter polynucleotide comprising a propromoter and a region which hybridizes to the displaced primer extension product under conditions which allow transcription to occur by RNA polymerase (these conditions generally but not necessarily include extension of the 3' end of the primer extension product to generate a double stranded promoter region), such that

IDT Exhibit 1020.8712

an RNA transcript comprising sequences complementary to the displaced primer extension products is produced; (f) hybridizing a second primer to the RNA transcript of step (e); (g) extending the second primer with RNA-dependent DNA polymerase to generate a complex comprising a second primer extension product and the RNA transcript; (h) cleaving RNA in the complex of step (g) with an enzyme that cleaves RNA in an RNA/DNA hybrid; (i) hybridizing the single stranded second primer extension product with a propromoter polynucleotide, wherein the propromoter polynucleotide comprises a propromoter and a region which hybridizes to the single stranded second primer extension product under conditions which allow transcription to occur by RNA polymerase (these conditions generally but not necessarily include extension of the 3' end of the primer extension product to generate a double stranded promoter region), such that an RNA transcript comprising the sequence of interest is produced; whereby multiple copies of the nucleic acid sequence of interest are produced.

In another aspect the invention provides methods for generating multiple copies of a nucleic acid sequence of interest, comprising (a) combining: the complex comprising a first primer extension product and target polynucleotide as described above; a composite primer (which is generally and preferably the same as the first primer) that is hybridizable to the target polynucleotide, wherein the composite primer is a primer that comprises an RNA portion and a 3' DNA portion; an enzyme that cleaves RNA from an RNA/DNA hybrid; a propromoter polynucleotide comprising a propromoter and a region which hybridizes to displaced composite primer extension product; an RNA polymerase; a second primer that is hybridizable to a sense RNA transcript comprising the sequence of interest; an RNA-dependent DNA polymerase; and a propromoter polynucleotide comprising a propromoter and a region which hybridizes to a second primer extension product; and (b) incubating the mixture of step (a) under conditions that permit primer hybridization and extension, RNA cleavage, displacement of the first primer extension product from the complex comprising a first primer extension product and target polynucleotide when its RNA is cleaved and a composite primer binds to the target polynucleotide in the complex, hybridization of a propromoter polynucleotide to a first primer extension product to form a complex comprising a first primer extension product and a propromoter polynucleotide, hybridization of a propromoter polynucleotide to a second primer extension product to form a complex comprising a second primer extension product and a propromoter

IDT Exhibit 1020.8713

polynucleotide, and RNA transcription, whereby multiple copies of the nucleic acid sequence of interest are generated.

In another aspect the invention provides methods for generating multiple copies of a nucleic acid sequence of interest, comprising: (a) combining: the displaced primer extension product generated as described above; a propromoter polynucleotide comprising a propromoter and a region which hybridizes to displaced first primer extension product; an RNA polymerase; a second primer that is hybridizable to a sense RNA transcript comprising the sequence of interest; an RNA-dependent DNA polymerase; an enzyme that cleaves RNA from an RNA/DNA hybrid; and a propromoter polynucleotide comprising a propromoter and a region which hybridizes to a second primer extension product; and (b) incubating the mixture of step (a) under conditions that permit primer hybridization and extension, RNA cleavage, hybridization of a propromoter polynucleotide to a first primer extension product to form a complex comprising a first primer extension product and a propromoter polynucleotide, hybridization of a propromoter polynucleotide to a second primer extension product to form a complex comprising a second primer extension product and a propromoter polynucleotide, and RNA transcription, whereby multiple copies of the nucleic acid sequence of interest are generated.

In another aspect the invention provides methods for generating multiple copies of a nucleic acid sequence of interest, comprising: (a) combining: an RNA transcript comprising sequences complementary to the displaced primer extension products as described above; a second primer that is hybridizable to a sense RNA transcript comprising the sequence of interest; an RNA-dependent DNA polymerase; an enzyme that cleaves RNA from an RNA/DNA hybrid; a propromoter polynucleotide comprising a propromoter and a region which hybridizes to a second primer extension product; and an RNA polymerase; and (b) incubating the mixture of step (a) under conditions that permit primer hybridization and extension, RNA cleavage, hybridization of a propromoter polynucleotide to a primer extension product to form a complex comprising a primer extension product and a propromoter polynucleotide, and RNA transcription, whereby multiple copies of the nucleic acid sequence of interest are generated.

In yet another aspect the invention provides methods for generating multiple copies of a nucleic acid sequence of interest, comprising: (a) combining: a first primer, wherein the first primer is a composite primer that is hybridizable to a target polynucleotide, and wherein the composite primer comprises an RNA portion and a 3' DNA portion; optionally

IDT Exhibit 1020.8714

a polynucleotide comprising a termination polynucleotide sequence that is hybridizable to a region of the target polynucleotide which is 5' with respect to hybridization of the composite primer to the target polynucleotide; a DNA-dependent DNA polymerase; an enzyme that cleaves RNA from an RNA/DNA hybrid; a propromoter polynucleotide comprising a propromoter and a region which hybridizes to a first primer extension product; an RNA polymerase; a second primer that is hybridizable to a sense RNA transcript comprising the sequence of interest; an RNA-dependent DNA polymerase; and a propromoter polynucleotide comprising a propromoter and a region which hybridizes to a second primer extension product; and (b) incubating the mixture of step (a) under conditions that permit primer hybridization and extension, RNA cleavage, displacement of a first primer extension product from a complex comprising a first primer extension product and target polynucleotide when its RNA is cleaved and a composite primer binds to the target polynucleotide in the complex, hybridization of a propromoter polynucleotide to a primer extension product to form a complex comprising a primer extension product and a propromoter polynucleotide, and RNA transcription, whereby multiple copies of the nucleic acid sequence of interest are generated.

In another aspect, the invention provides methods for generating multiple copies of a nucleic acid sequence of interest, comprising: (a) hybridizing a second primer to an RNA transcript, said RNA transcript produced by a process comprising (i) hybridizing a single stranded target polynucleotide comprising the nucleic acid sequence of interest with a first primer, wherein said first primer is a composite primer comprising an RNA portion and a 3' DNA portion; (ii) optionally hybridizing a polynucleotide comprising a termination polynucleotide sequence to a region of the target polynucleotide 5' with respect to hybridization of the first primer to the target polynucleotide; (iii) extending the first primer with a DNA-dependent DNA polymerase to generate a complex comprising a first primer extension product and target polynucleotide; (iv) cleaving the RNA portion from the composite primer in the complex of first primer extension product and target polynucleotide with an enzyme that cleaves RNA from an RNA/DNA hybrid such that another composite primer (which is generally and preferably the same as the first primer) can hybridize to the target polynucleotide and repeat primer extension by strand displacement to produce displaced primer extension product; (v) hybridizing a propromoter polynucleotide comprising a propromoter and a region which hybridizes to the displaced primer extension product under conditions which allow transcription to occur by RNA

IDT Exhibit 1020.8715

polymerase (these conditions generally but not necessarily include extension of the 3' end of the primer extension product to generate a double stranded promoter region), such that an RNA transcript comprising sequences complementary to the displaced primer extension product is produced; and (b) extending the second primer with RNA-dependent DNA polymerase to generate a complex comprising a second primer extension product and the RNA transcript; (c) cleaving RNA in the complex of step (b) with an enzyme that cleaves RNA in an RNA/DNA hybrid; (d) hybridizing the single stranded second primer extension product with a propromoter polynucleotide, wherein the propromoter polynucleotide comprises a propromoter and a region which hybridizes to the single stranded second primer extension product under conditions which allow transcription to occur by RNA polymerase (these conditions generally but not necessarily include extension of the 3' of the primer extension product to generate a double stranded promoter region), such that an RNA transcript comprising the sequence of interest is produced; whereby multiple copies of the nucleic acid sequence of interest are produced.

Various embodiments of the composite primer used in the composite primer-based methods of the invention are described herein. For example, in some embodiments, the RNA portion of the composite 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. For the composite primer-based methods described herein, one or more composite primers can be used. In preferred embodiments of the composite primer-based methods, the first and second primers are different (for example, the first primer is a composite primer and the second primer is not a composite primer). In preferred embodiments wherein the first and second primers are different, the primers are hybridizable to similar, preferably identical, sequences. In still other embodiments of the composite primer-based methods, the first and second primers are hybridizable to different sequences.

Various exemplary embodiments of polynucleotides comprising a termination sequence are also described herein. In some embodiments, the polynucleotide comprising a termination sequence is a template switch oligonucleotide (TSO), which may (but not necessarily) contain one or more modifications to enhance binding to template. Accordingly, in some embodiments, the TSO comprises a modification in the region which hybridizes to the template, wherein, under a given set of conditions, the TSO binds more tightly to the region as compared to a TSO without the modification. Examples of suitable modifications are provided herein. In some embodiments, the polynucleotide comprising a

IDT Exhibit 1020.8716

termination sequence is a blocking sequence, which, like the TSO, may contain one or more modifications to enhance binding to template. Accordingly, in some embodiments, the blocker sequence comprises a modification in the region which hybridizes to the template, wherein, under a given set of conditions, the blocker binds more tightly to the region as compared to a blocker without the modification. Examples of suitable modifications are provided herein.

The enzymes which may be used in the methods and compositions are described herein. For example, the enzyme that cleaves RNA may be an RNaseH.

In some aspects, a TSO provides propromoter function and also comprises a region (which may or may not be adjacent to the promoter) which hybridizes to the displaced primer extension product. In other embodiments, the polynucleotide comprising the propromoter comprises a region at the 3' end which hybridizes to the displaced primer extension product, whereby DNA polymerase extension of displaced extension product produces a double stranded promoter from which transcription occurs. In some embodiments, the polynucleotide comprising the propromoter is a propromoter template oligonucleotide (PTO). In some embodiments, the polynucleotide comprising the propromoter is a polynucleotide comprising: (a) a termination sequence that does not effect template switch under conditions wherein the termination sequence is hybridizable to a target polynucleotide; (b) a propromoter sequence, wherein the propromoter sequence is not hybridizable to a target polynucleotide under conditions wherein the termination sequence is hybridizable to the target polynucleotide; and (c) a sequence which is hybridizable to a complementary sequence of said target polynucleotide.

In another aspect the invention provides methods for generating multiple copies of a nucleic acid sequence of interest, comprising: (a) hybridizing a single stranded target polynucleotide comprising the sequence of interest with a first primer; (b) hybridizing a propromoter template switch oligonucleotide (TSO) comprising a propromoter sequence and a region that is hybridizable to a region of the target polynucleotide which is 5' with respect to hybridization of the first primer to the target polynucleotide; (c) extending the first primer with DNA polymerase such that a first primer extension product comprising a sequence complementary to the propromoter sequence of the propromoter TSO is produced, whereby a complex of first primer extension product, target polynucleotide and propromoter TSO is generated, wherein said complex comprises a double stranded promoter region; (d) transcribing from the double stranded promoter region with a DNA-

IDT Exhibit 1020.8717

dependent RNA polymerase, to generate a sense RNA transcript; (e) hybridizing a second primer (which is generally and preferably the same as the first primer) to the sense RNA transcript of step (d); (f) extending the second primer with RNA-dependent DNA polymerase to generate a complex comprising a second primer extension product and an RNA transcript; (g) cleaving RNA in the complex of step (f) with an enzyme that cleaves RNA in an RNA/DNA hybrid; (h) hybridizing a single stranded second primer extension product with a propromoter polynucleotide, wherein the propromoter polynucleotide comprises a propromoter and a region which hybridizes to the single stranded second primer extension product under conditions which allow transcription to occur by RNA polymerase, such that sense RNA transcripts comprising the sequence of interest are produced; whereby multiple copies of the nucleic acid sequence of interest are produced.

In another aspect the invention provides methods for generating multiple copies of a nucleic acid sequence of interest, comprising: (a) combining: the complex described above of a first primer extension product, target polynucleotide and propromoter TSO, wherein said complex comprises a double stranded promoter region; an RNA polymerase; a second primer (which is generally and preferably the same as the first primer) that is hybridizable to a sense RNA transcript comprising the sequence of interest; an RNA-dependent DNA polymerase; an enzyme that cleaves RNA from an RNA/DNA hybrid; and a propromoter polynucleotide comprising a propromoter and a region which hybridizes to a second primer extension product; and (b) incubating the mixture of step (a) under conditions that permit primer hybridization and extension, RNA cleavage, hybridization of a propromoter polynucleotide to a primer extension product to form a complex comprising a primer extension product and a propromoter polynucleotide, and RNA transcription, whereby multiple copies of the nucleic acid sequence of interest are generated.

In another aspect the invention provides methods for generating multiple copies of a nucleic acid sequence of interest, comprising: (a) combining: a sense RNA transcript generated as described above; a second primer (which is generally and preferably the same as the first primer) that is hybridizable to the RNA transcript; an RNA-dependent DNA polymerase; an enzyme that cleaves RNA from an RNA/DNA hybrid; a propromoter polynucleotide comprising a propromoter and a region which hybridizes to a second primer extension product; and an RNA polymerase; and (b) incubating the mixture of step (a) under conditions that permit primer hybridization and extension, RNA cleavage, hybridization of a propromoter polynucleotide to a primer extension product to form a

IDT Exhibit 1020.8718

complex comprising a primer extension product and a propromoter polynucleotide, and RNA transcription, whereby multiple copies of the nucleic acid of interest are generated.

In another aspect the invention provides methods for generating multiple copies of a nucleic acid sequence of interest, comprising: (a) combining: a target polynucleotide; a first primer which is hybridizable to the target polynucleotide; a propromoter template switch oligonucleotide comprising a propromoter sequence and a region that is hybridizable to a region of the target polynucleotide which is 5' with respect to hybridization of the first primer to the target polynucleotide; optionally a DNA-dependent DNA polymerase; an RNA polymerase; a second primer (which is generally and preferably the same as the first primer) that is hybridizable to a sense RNA transcript comprising the sequence of interest; an RNA-dependent DNA polymerase; an enzyme that cleaves RNA from an RNA/DNA hybrid; and a propromoter polynucleotide comprising a propromoter and a region which hybridizes to a second primer extension product; and (b) incubating the mixture of step (a) under conditions that permit primer hybridization and extension, RNA cleavage, hybridization of a propromoter polynucleotide to a primer extension product to form a complex comprising a primer extension product and a propromoter polynucleotide, and RNA transcription, whereby multiple copies of the nucleic acid sequence of interest are generated.

In another aspect, the invention provides methods for generating multiple copies of a nucleic acid sequence of interest, comprising: (a) hybridizing a second primer to an RNA transcript, said RNA transcript produced by a process comprising (i) hybridizing a single stranded target polynucleotide comprising the sequence of interest with a first primer (which is generally and preferably the same as the second primer); (ii) hybridizing a propromoter template switch oligonucleotide (TSO) comprising a propromoter sequence and a region that is hybridizable to a region of the target polynucleotide which is 5' with respect to hybridization of the first primer to the target polynucleotide; (iii) extending the first primer with DNA polymerase such that a first primer extension product comprising a sequence complementary to the propromoter sequence of the propromoter TSO is produced, whereby a complex of first primer extension product, target polynucleotide and propromoter TSO is generated, wherein said complex comprises a double stranded promoter region; (iv) transcribing from the double stranded promoter region with a DNA-dependent RNA polymerase to generate a sense RNA transcript; (b) extending the second primer with RNA-dependent DNA polymerase to generate a complex comprising a second

IDT Exhibit 1020.8719

primer extension product and an RNA transcript; (c) cleaving RNA in the complex of step (b) with an enzyme that cleaves RNA in an RNA/DNA hybrid; (d) hybridizing a single stranded second primer extension product with a propromoter polynucleotide, wherein the propromoter polynucleotide comprises a propromoter and a region which hybridizes to the single stranded second primer extension product under conditions which allow transcription to occur by RNA polymerase, such that sense RNA transcripts comprising the sequence of interest are produced; whereby multiple copies of the nucleic acid sequence of interest are produced.

In preferred embodiments of the propromoter TSO-based methods, the first primer and the second primer are the same. In other embodiments of the propromoter TSO-based methods, the first and second primers are different. In certain embodiments wherein the first and second primers are different, the primers are hybridizable to similar, preferably identical, sequences. In still other embodiments of the propromoter TSO-based methods, the first and second primers are hybridizable to different sequences.

The methods are applicable to amplifying any DNA or RNA target, including, for example, genomic DNA or RNA and cDNA. One or more steps may be combined and/or performed sequentially (often in any order, as long as the requisite product(s) are formed).

The invention also provides methods which employ (usually, analyze) the products of the amplification methods of the invention, such as sequencing and detection of sequence alteration(s). A method of sequencing according to the invention comprises: (a) amplifying a target polynucleotide by the methods of the invention in the presence of a mixture of rNTPs and rNTP analogs such that transcription is terminated upon incorporation of an rNTP analog; and (b) analyzing the amplification products to determine sequence.

Another method of sequencing according to the invention comprises: (a) amplifying a target polynucleotide by the methods of the invention wherein RNA transcripts generated from a first primer extension product are amplified in the presence of a mixture of rNTPs and rNTP analogs such that transcription is terminated upon incorporation of an rNTP analog; and (b) analyzing the amplification products to determine sequence. Another method of sequence analysis according to the invention comprises: (a) amplifying a target polynucleotide by the methods of the invention; and (b) analyzing the amplification products for single stranded conformation, wherein a difference in conformation as

IDT Exhibit 1020.8720

compared to a reference single stranded polynucleotide indicates a mutation in the target polynucleotide sequence.

The invention provides methods of characterizing a sequence of interest in a target polynucleotide comprising (i) amplifying a target polynucleotide sequence containing the sequence of interest by the methods of the invention, wherein the sequence of the RNA portion of the composite primer is known, and (ii) comparing the amplification products if any from step (i) with the amount of amplification products from a reference template wherein (1) production of detectably fewer amplification products from the template as compared to the amount of amplification products from the reference template which comprises a region complementary to the RNA portion of the composite primer indicates that the target polynucleotide does not comprise a sequence complementary to the RNA portion of the composite primer and is a sequence variant with respect to the sequence complementary to the RNA portion of the composite primer; or (2) production of detectably more amplification products from the template as compared to the amount of amplification products from the reference template which does not comprise a region which is complementary to the RNA portion of the composite primer indicates that the target polynucleotide comprises a sequence complementary to the RNA portion of the composite primer and is not a sequence variant with respect to the sequence complementary to the RNA portion of the composite primer.

In the methods of the invention the sequence of the RNA portion of the composite primer may comprise a sequence complementary to the sequence of interest which may be a wild type sequence, a mutant sequence, or an allelic variant sequence.

The methods of the invention are useful for detection and quantification of a target nucleic acid sequence.

For example, in one aspect, the invention provides methods of sequencing a sequence of interest comprising sequencing a nucleic acid amplification product, wherein said nucleic acid amplification product is generated by an amplification method described herein. Sequencing of the amplification products reveals the sequence of the sequence of interest.

In yet another example, the invention provides methods of detecting presence of a nucleic acid sequence of interest in a sample, said method comprising detecting presence of the sequence of interest in a nucleic acid amplification product generated by an

IDT Exhibit 1020.8721

amplification method described herein. Detection of the sequence of interest in the amplification products is indicative of presence of the sequence of interest in a sample.

In various embodiments of methods of the invention, a sequence of interest comprises a mutation. In some embodiments wherein a sequence of interest comprises a mutation, the mutation is a single nucleotide polymorphism. In some embodiments, detecting of a sequence of interest comprises hybridizing an amplification product comprising the sequence of interest with a nucleic acid probe that is hybridizable to said nucleic acid sequence of interest. In other embodiments, detecting of a sequence of interest comprises hybridizing an amplification product comprising the complement of the sequence of interest with a nucleic acid probe that is hybridizable to said complement. In some embodiments, a nucleic acid probe comprises DNA. In other embodiments, a nucleic acid probe comprises RNA. In some embodiments, a probe is provided as a microarray. In some embodiments, a microarray comprises a probe immobilized on a substrate fabricated from a material selected from the group consisting of paper, glass, plastic, polypropylene, nylon, polyacrylamide, nitrocellulose, silicon, and optical fiber. In some embodiments, a sequence of interest or its complement is detected by conducting limited primer extension, wherein the limited primer extension comprises extending a primer hybridized to an amplification product such that a primer extension product is generated that has a characteristic that indicates presence or absence of the sequence of interest.

Thus, the methods of the invention are useful for detection and quantification of a target nucleic acid sequence. The single-stranded nucleic acid amplification products of the invention can also be used for the detection and quantification of a target nucleic acid sequence by hybridization to probes, such as immobilized probes. The invention also provides methods of producing a microarray, comprising (a) amplifying a target polynucleotide by the methods of the invention; and (b) immobilizing the amplification products on a substrate to fabricate a microarray comprising the amplification products.

One aspect of the invention provides methods of determining gene expression profile in a sample comprising: (a) amplifying at least two sequences of interest in the sample by the methods of the invention; and (b) determining amount of amplification products of each sequence of interest, wherein each said amount is indicative of amount of each sequence of interest in the sample, whereby the gene expression profile in the sample is determined. In one embodiment each target polynucleotide is a cDNA. In another embodiment each target polynucleotide is RNA.

IDT Exhibit 1020.8722

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. One aspect of the invention provides a system for amplifying a sequence of interest, comprising: (a) a first primer which is a composite primer; (b) a second primer; (c) a DNA-dependent DNA polymerase; (d) an RNA-dependent DNA polymerase; (e) a propromoter polynucleotide; (f) an RNA polymerase; and (g) an enzyme that cleaves RNA from an RNA/DNA hybrid; and optionally, (h) a polynucleotide comprising a termination polynucleotide sequence. Another aspect of the invention provides a system for amplifying a sequence of interest, comprising: (a) a propromoter TSO; (b) a first primer; (c) optionally a DNA-dependent DNA polymerase; (d) an RNA-dependent DNA polymerase; (e) an enzyme that cleaves RNA from an RNA/DNA hybrid; (f) an RNA polymerase; and optionally (g) a second primer.

In another aspect, the invention provides reaction mixtures (or compositions comprising reaction mixtures) which contain various combinations of components described herein. 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. For example, the invention provides kits that comprise a composite primer (for example, comprising a 3' DNA portion and an RNA portion, which may be 5' and may further be adjacent to the 3' DNA portion) that is hybridizable to a target polynucleotide, a propromoter polynucleotide, and instructions for amplifying the target polynucleotide according to any composite primer-based method described herein. The composite primer and propromoter polynucleotide in the kits can be any described herein. The kits can contain further components, such as any of (a) a polynucleotide comprising a termination polynucleotide sequence; (b) any of the enzymes described herein, such as an enzyme which cleaves RNA from an RNA/DNA hybrid (for example, RNaseH); (c) a propromoter TSO; and (d) a second primer. In another example, the invention provides kits comprising a propromoter TSO and a first primer (which may or may not be a composite primer), wherein both are hybridizable to a target polynucleotide, and instructions for amplifying the target polynucleotide according to any propromoter TSO-based method described herein. The kits can contain further components, such as any of (a) any of the enzymes described herein, such as an enzyme which cleaves RNA from an RNA/DNA hybrid (for

IDT Exhibit 1020.8723

example, RNaseH); (b) a polynucleotide comprising a propromoter and a region which hybridizes to a second primer extension product; and (c) optionally a second primer.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGURES 1A-F is a diagrammatic representation of the steps of a composite primer isothermal amplification process using a blocker sequence component as a termination polynucleotide.

FIGURES 2A-E is a diagrammatic representation of the steps of a composite primer isothermal amplification process using a template switch oligonucleotide sequence as a termination polynucleotide.

FIGURES 3A-B is a diagrammatic representation of the steps of a single primer isothermal amplification of a DNA target using template switching.

FIGURES 4A-B is a diagrammatic representation of the steps of a single primer isothermal amplification of an RNA target using template switching.

MODES FOR CARRYING OUT THE INVENTION

The invention provides methods, compositions and kits for amplifying nucleic acid sequences. Amplification of target nucleic acid sequence can be advantageous in many instances, for example in cases where the target polynucleotide is present in very small quantities, or where only a few molecules of a target polynucleotide is available, or for the purpose of nucleotide sequence determination of a defined target nucleic acid.

In one aspect, the invention provides methods generally using an RNA/DNA composite primer and two sets or rounds of transcription steps (each set or round referred to herein as a "transcription module") based on hybridization of a propromoter polynucleotide to intermediate polynucleotides. In another aspect, the invention provides methods comprising using, in a first transcription module, a propromoter template switch oligonucleotide to generate an intermediate polynucleotide complex capable of being transcribed to generate sense RNA transcripts and, in a second transcription module, further amplifying the RNA transcripts of the first transcription module by transcription.

In one aspect, the amplification methods generally work as follows: a composite RNA/DNA primer and a first propromoter polynucleotide form the basis for transcription

IDT Exhibit 1020.8724

of a target polynucleotide in the first transcription module. In some embodiments, a termination sequence provides an endpoint for the replication by either diverting or blocking further replication along the target strand. Generally, the termination sequence comprises a sequence that is sufficiently complementary to hybridize to the target strand. As described below, in some embodiments a polynucleotide comprising a termination sequence is a template switch oligonucleotide (TSO) which comprises sequences that are not sufficiently complementary to hybridize to the target strand in addition to the sequences which are sufficiently complementary to hybridize. A DNA polymerase uses the primer to initiate copying of the target polynucleotide to generate a primer extension product. An enzyme (such as RNaseH) which cleaves RNA from an RNA/DNA hybrid cleaves (removes) RNA sequences from the hybrid of template strand and primer extension product, leaving a sequence on the template strand available for binding by another composite primer. A second strand is then produced by DNA polymerase, which displaces the previously replicated strand, resulting in a displaced extension product. A polynucleotide comprising a propromoter and a region which hybridizes to the displaced primer extension product (which can be, for example, a template switch oligonucleotide or propromoter template oligonucleotide) is hybridized to the displaced primer extension product.

If the primer extension product comprises a propromoter sequence, hybridization of a propromoter polynucleotide (such as a template switch oligonucleotide (TSO)) to this sequence generates a double stranded promoter region. Where the displaced primer extension product does not comprise a suitable propromoter sequence, a propromoter sequence is incorporated using a propromoter polynucleotide, such as a propromoter template oligonucleotide (PTO). A propromoter polynucleotide binds to the displaced primer extension product. The double stranded promoter that is created from the displaced primer extension product is used by a DNA-dependent RNA polymerase to produce sense RNA products (i.e., RNA which comprises the target sequence, as opposed to its complement). A second primer is then hybridized to the sense RNA transcript and extended by an RNA-dependent DNA polymerase to generate a complex of a second primer extension product and the sense RNA transcript. A ribonuclease that cleaves RNA in an RNA/DNA hybrid (such as RNaseH) cleaves the RNA in the complex. A propromoter polynucleotide is then hybridized to a single stranded second primer extension product such that a double stranded promoter region is created (which, as described herein, can arise

IDT Exhibit 1020.8725

directly via hybridization or indirectly via extension of the primer extension product along the propromoter polynucleotide). Transcription from the double stranded promoter region produces sense RNA transcripts comprising the sequence of interest.

Accordingly, the invention provides methods of producing at least one copy of a nucleic acid sequence of interest (generally, methods of amplifying a target polynucleotide sequence) comprising combining and reacting the following: (a) a single-stranded target polynucleotide comprising a target sequence; (b) a first primer, which is a composite primer comprising an RNA portion and a 3' DNA portion; (c) a DNA polymerase; (d) deoxyribonucleoside triphosphates or suitable analogs; (e) an enzyme, such as RNaseH, which cleaves RNA from an RNA/DNA duplex; (f) optionally a polynucleotide comprising a termination sequence, such as any of those described herein, which comprises a portion (or region) which hybridizes to the template polynucleotide; (g) a polynucleotide comprising a propromoter sequence (which can be in any of a number of forms, as described herein) and a region which hybridizes to the displaced primer extension product; (h) ribonucleoside triphosphates or suitable analogs; (i) a second primer (which may or may not be the same as the first primer; and (j) RNA polymerase, under conditions such that transcription of the displaced strand can occur. The combination is subjected to suitable conditions such that (a) the composite primer (and, optionally, a polynucleotide comprising a termination sequence) hybridizes to the template; (b) primer extension occurs from the composite primer, to form a duplex; (c) RNaseH cleaves RNA of the composite primer from the RNA/DNA duplex; (d) another composite primer (which is generally and preferably the same as the first primer) hybridizes to the template, and another round of primer extension (mediated by DNA polymerase) occurs, displacing the strand already copied from the template; (e) transcription occurs to generate RNA transcripts complementary to a composite primer extension product; (f) a second primer hybridizes to the RNA transcript of step (e); (g) primer extension occurs from the second primer, to form a duplex; (h) a ribonuclease such as RNaseH cleaves RNA from the duplex of step (g); (i) a propromoter polynucleotide hybridizes to the single stranded second primer extension product that is generated by step (h) such that a double stranded promoter region is generated (either directly or indirectly as described herein); and (j) transcription occurs to generate sense RNA transcripts comprising the sequence of interest. Optional repetition of the second set of transcription steps using the sense transcripts of step (j) as the starting

IDT Exhibit 1020.8726

substrate (to which the second primer would hybridize) results in exponential amplification of the sequence of interest.

It is understood that, with respect to the combinations described herein, it is not necessary to react all components in the same vessel and/or simultaneously. Accordingly, the invention provides subcombinations, as long as the subcombinations expressly or implicitly reflect the amplification manipulations/methods described herein.

In another aspect, as a general summary, the amplification methods work as follows: template switching during extension of a first primer due to a propromoter TSO that is hybridized 5' with respect to the hybridization site of the first primer on a target polynucleotide results in generation of a polynucleotide complex that comprises a double stranded promoter region. The propromoter TSO comprises a sequence hybridizable (i.e., that hybridizes) to the target polynucleotide, and a sequence that is not hybridizable (i.e., does not hybridize under conditions that permit hybridization of the hybridizable sequence) to the target polynucleotide, wherein the sequence that is not hybridizable comprises a propromoter sequence. In a first transcription module, extension of the first primer along the target polynucleotide sequence is switched to the non-hybridizable sequence of the propromoter TSO at the point where the TSO is hybridized to the target polynucleotide. In some embodiments wherein the target polynucleotide is RNA, the RNA is cleaved following first primer extension. In other embodiments wherein the target polynucleotide is RNA, the RNA is not cleaved following first primer extension. The complex of the primer extension product and the non-hybridizable sequence of the TSO comprises a double stranded promoter region. Transcription occurs from the double stranded promoter region to generate sense RNA transcripts. In a second transcription module, these sense RNA transcripts are subjected to further amplification. In the second module, a second primer (which is generally and preferably the same as the first primer) is hybridized to a sense RNA transcript and extended by an RNA-dependent DNA polymerase to generate a complex of a second primer extension product and the sense RNA transcript. A ribonuclease that cleaves RNA in an RNA/DNA hybrid (such as RNaseH) cleaves RNA in the complex. A propromoter polynucleotide is then hybridized to a single stranded second primer extension product such that a double stranded promoter region is created (either directly or indirectly, as described herein). Transcription from the double stranded promoter region produces sense RNA transcripts comprising the sequence of interest.

IDT Exhibit 1020.8727

Accordingly, the invention provides methods of producing at least one copy of a nucleic acid sequence of interest (generally, methods of amplifying target polynucleotide sequence) comprising combining and reacting the following: (a) a single stranded target polynucleotide; (b) a first primer that is hybridizable to the target polynucleotide at a location 3' of the sequence of interest; (c) a propromoter TSO; (d) a DNA polymerase; (e) deoxyribonucleoside triphosphates or suitable analogs; (f) an RNA polymerase; and (g) ribonucleoside triphosphates or suitable analogs; (h) a second primer (which is generally and preferably the same as the first primer) that is hybridizable to an RNA transcript produced in the first transcription step; (i) an enzyme, such as RNaseH, which cleaves RNA from an RNA/DNA duplex; (j) a propromoter polynucleotide (which can be in any of a number of forms, as described herein) which hybridizes to a second primer extension product. The combination is subjected to suitable conditions such that: (a) the first primer and propromoter TSO hybridize to the target polynucleotide; (b) primer extension occurs from the first primer along the target polynucleotide, and the non-hybridizable sequence of the propromoter TSO following template switching, to form a complex comprising a double stranded promoter region; (c) transcription occurs from the double stranded promoter region of step (b); and (d) a second primer (which is generally and preferably the same as the first primer) hybridizes to the RNA transcript of step (c); (e) primer extension occurs from the second primer, to form a duplex; (f) a ribonuclease such as RNaseH cleaves RNA from the duplex of step (e); (g) a propromoter polynucleotide hybridizes to a single stranded second primer extension product such that a double stranded promoter region is generated (either directly or indirectly, as described herein); and (h) transcription occurs to generate sense transcripts comprising the sequence of interest.

Any of the methods of the invention can be used to generate RNA products that are labeled by incorporating labeled nucleotides into appropriate steps of the methods. These labeled products are particularly suitable for quantification and/or identification of the nucleic acids by methods known in the art, which include the use of arrays such as cDNA microarrays and oligonucleotide arrays.

In some embodiments, the invention provides methods of sequencing a nucleic acid sequence of interest. For sequencing methods based on methods described herein, the appropriate rNTPs, or analogs thereof, which may be labeled or unlabelled, are used. Accordingly, the invention provides methods of sequencing a target polynucleotide comprising a sequence of interest based on the methods described above, wherein rNTPs

IDT Exhibit 1020.8728

and rNTP analogs, which may be labeled or unlabelled primer elongation terminators, are used and the amplification products are analyzed for sequence information, as described below. Sequencing by transcription can be used for both DNA and RNA target sequences using rNTPs and rNTP analogs.

The invention provides methods of detecting nucleic acid sequence mutations in target sequence(s). In one embodiment of the composite primer-based methods, the presence or absence of a mutation in a target polynucleotide is detected based on the ability to amplify the target polynucleotide using a composite primer whose RNA portion either contains or lacks the mutated sequence using the methods of the invention. The amplified products of the methods of the invention can also be characterized for sequences of interest. For example, the amplification products can be sequenced to determine a sequence of interest. They can also be used for detection of mutations, such as by hybridization with specific probes. The amplified products can also be used to detect and/or identify sequence alterations or single nucleotide polymorphisms in the target nucleic acid sequence by determining single strand conformation polymorphisms. Thus, the invention provides methods to characterize (for example, detect and/or quantify) a sequence of interest by generating 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. Generally, but not necessarily, these amplified products are labeled.

In yet another embodiment, the invention provides methods for generating microarrays of nucleic acids (RNA) using products of the amplification methods of the invention.

Details regarding the various components of the methods of the present invention are provided below.

General Techniques

The practice of the present invention 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", 2d. edition (Sambrook et al., (1989)); "Oligonucleotide Synthesis" (M.J. Gait, ed., 1984); "Animal Cell Culture" (R.I. Freshney, ed., 1987); "Methods in

IDT Exhibit 1020.8729

Enzymology" (Academic Press, Inc.); "Current Protocols in Molecular Biology" (F.M. Ausubel et al., eds., (1987)), and periodic updates); "PCR: The Polymerase Chain Reaction", (K. Mullis et al., eds., (1994)).

Primers, oligonucleotides and polynucleotides employed in the present invention can be generated using standard techniques known in the art. The primers of the present invention can be prepared by methods known in the art. Preferably, the primers are synthesized by chemical methods. For example, Caruthers et al. (Methods In Enzymology, vol. 154, p. 287 (1987)) describe using standard phosphoramidite solid-phase chemistry to join nucleotides by phosphodiester linkages.

Definitions

A "target polynucleotide," as used herein, is a polynucleotide known or suspected of comprising a nucleic acid sequence of interest, for which amplification is desired. The nucleic acid sequence of interest may be known or not known, in terms of its actual sequence. Generally, a "template," as used herein, is a polynucleotide that contains a nucleic acid sequence of interest. In some instances, the terms "target sequence," "template DNA," "template polynucleotide," "target nucleic acid," and variations thereof, are used interchangeably.

"Amplification," 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), and/or sequence errors that occur during amplification.

A "sense" polynucleotide (for example, a sense RNA transcript), as used herein, refers to a polynucleotide that comprises a sequence of interest from a target nucleic acid sequence as opposed to its complement.

"Polynucleotide," or "nucleic acid," as used interchangeably herein, refer to polymers of nucleotides of any length, and include DNA and RNA. The nucleotides can be deoxyribonucleotides, ribonucleotides, modified nucleotides or bases, and/or their analogs, or any substrate that can be incorporated into a polymer by DNA or RNA polymerase. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and

IDT Exhibit 1020.8730

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, □-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.

A "first transcription module," as used herein, refers to a process comprising steps of the methods up to and including the first transcription step. A "second transcription

IDT Exhibit 1020.8731

module," as used herein, refers to a process comprising steps of the methods starting from priming of the RNA transcript generated in the first transcription module to creating a DNA product up to and including the second transcription step.

A "labeled rNTP," as used herein, refers, respectively, to an rNTP, or analogs thereof, that is directly or indirectly attached with a label. For example, a "labeled" 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" rNTP may also be indirectly labeled by its attachment to, for example, a moiety to which a label is/can be attached. An rNTP may also comprise a moiety (for example, an amine group) to which a label may be attached following incorporation of the 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 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.

The "type" of rNTP, as used herein, refers to the particular base of a nucleotide, namely adenine, cytosine, uridine, or guanine.

"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 present invention include the composite primer, TSO, PTO and blocker sequence. The terms "oligonucleotide" and "polynucleotide" are not mutually exclusive. The description above for polynucleotides is equally and fully applicable to oligonucleotides.

A "primer" is generally 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) that is hybridized to a sequence in the template itself (for example, as a hairpin loop), and that is capable of promoting nucleotide polymerization.

IDT Exhibit 1020.8732

A "random primer," as used herein, is a primer that comprises a sequence that is designed not based on a particular or specific sequence in a sample, but rather is based on a statistical expectation 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 existing, or 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 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 by a polymerase which does not require prior knowledge of the exact sequence of the target.

A "termination polynucleotide sequence" or "termination sequence," as used interchangeably herein, is a polynucleotide sequence which effects cessation of DNA replication by DNA polymerase with respect to the template comprising the target sequence. A termination sequence comprises a portion (or region) that generally hybridizes to the template at a location 5' to the termination point (site). The hybridizable portion may or may not encompass the entire termination sequence. Examples of suitable termination polynucleotide sequences (such as blocker sequences and TSOs) are provided herein. Other examples of suitable termination polynucleotide sequences include a combination polynucleotide comprising a termination sequence (which generally and preferably does not effect template switch under conditions in which the termination sequence is hybridizable to a target polynucleotide) and a propromoter sequence, as described herein.

"Blocker sequence," or "blocking sequence" as used interchangeably herein, is an example of a termination sequence, and refers to an oligonucleotide that binds, generally with high affinity, to the template nucleic acid at a location 5' to the termination site and effects cessation of DNA replication by DNA polymerase with respect to the template comprising the target sequence. Its 3' end may or may not be blocked for extension by DNA polymerase.

"Termination site," or "termination point," as used interchangeably herein, refers to the site, point or region of the template that is last replicated by the DNA polymerase

IDT Exhibit 1020.8733

before termination of polymerization (generally, primer extension) or template switch. For example, with respect to a TSO, it is the position or region in the target sequence that is complementary to the 3' end of the primer extension product prior to switching template from the template polynucleotide to the unhybridized portion of the TSO.

"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. A propromoter sequence may be provided by any of the propromoter polynucleotides described herein, for example TSO, PTO, or a combination polynucleotide comprising a termination sequence (which generally and preferably does not effect template switch under conditions in which the termination sequence is hybridizable to a target polynucleotide) and a propromoter sequence, as described herein.

"Template switch oligonucleotide," "target switch oligonucleotide" or "TSO," as used herein, refers to an oligonucleotide that comprises a portion (or region) that is hybridizable to a template at a location 5' to the termination site of primer extension and that is capable of effecting a template switch in the process of primer extension by a DNA polymerase, generally due to a sequence that is not hybridized to the template. TSOs are generally known in the art. "Template switch" refers to a change in template nucleic acid, generally from the target nucleic acid to the unhybridized portion of a TSO, during the course of a single round of primer extension.

A "propromoter TSO," as used herein, is a TSO that further comprises a propromoter sequence that is not hybridizable to a target polynucleotide (under a given set of conditions, which are generally those in effect when the hybridizable portion or region of the TSO is hybridized to the target). The propromoter sequence is generally, but not necessarily, in the 5' region of the TSO.

"Propromoter template oligonucleotide (PTO)," as used herein, refers to an oligonucleotide that comprises a propromoter sequence and a portion, generally a 3' portion, that is hybridizable 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 1020.8734

A first sequence which "corresponds" to a second sequence, such as an RNA portion of a composite primer, means that the first sequence has significant sequence identity with respect to the second sequence. This term is generally used in the context of detecting mutations, or characterizing sequences of a target.

To "inhibit" is to decrease or reduce an activity, function, and/or amount as compared to a reference.

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).

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 contiguous nucleotides.

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. For an illustration of this example, see Figure 1A-C.

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).

"A", "an" and "the", and the like, unless otherwise indicated include plural forms.

"Comprising" means including.

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.

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

IDT Exhibit 1020.8735

sequence, due to mechanisms such as substitution, deletion or insertion. Single nucleotide polymorphism (SNP) is also a sequence mutation as used herein.

"Single stranded conformation polymorphism," and "SSCP," as used herein, generally refers to the specific conformation of a single stranded nucleic acid as is affected by its specific nucleic acid sequence. Alteration of the sequence of the single stranded polynucleotide, such as single nucleotide substitution, deletions or insertions, results in change, or polymorphism, of the conformation of the single stranded polynucleotide. The conformation of the polynucleotide is generally detectable, identifiable and/or distinguishable using methods known in the art, such as electrophoretic mobility as measured by gel electrophoresis, capillary electrophoresis, and/or susceptibility to endonuclease digestion.

"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 positions on a substrate. Arrays are formed on substrates fabricated with materials such as paper, glass, plastic (e.g., polypropylene, nylon), 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 noncovalently immobilized on the substrate by hybridization to

IDT Exhibit 1020.8736

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.

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.

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.

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 20, more preferably from about 3 to about 18, even more preferably from about 5 to about 15 nucleotides.

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 20, more preferably from about 3 to about 18, even more preferably from about 5 to about 15 nucleotides.

"Detection" includes any means of detecting, including direct and indirect detection. For example, "detectably fewer" products may be observed directly or indirectly, and the term indicates any reduction (including no products). Similarly, "detectably more" product means any increase, whether observed directly or indirectly.

Amplification Methods

The following are examples of amplification methods of the invention. It is understood that various other embodiments may be practiced, given the general description

IDT Exhibit 1020.8737

provided. For example, reference to using a composite primer means that any of the composite primers described herein may be used.

In one aspect, the methods of the invention utilize two linked transcription modules: a first transcription module that provides a composite primer-based linear nucleic acid amplification that is transcription-based (also referred to herein as "enhanced linear amplification"), which is in turn linked to a second transcription module that enables further isothermal amplification of the amplification products of the first module. The methods of the invention generally utilize two primers, both of the same sense with respect to the target nucleic acid strand to be amplified. All amplification reactions can be isothermal, and do not require thermocycling.

In another aspect, methods for generating multiple copies of sense RNA transcripts of a nucleic acid of interest are based on using a propromoter template switch oligonucleotide (TSO) to generate an intermediate polynucleotide product capable of being transcribed in a first set of transcription steps (a first transcription module) to generate RNA transcripts that are further amplified through a second set of transcription steps (a second transcription module).

The products of the amplification methods can be detected by either homogenous or heterogeneous detection, including identification by size and/or migration properties in gel electrophoresis, or by hybridization to sequence-specific probes. The detection of the amplification product is indicative of the presence of the target nucleic acid. Quantitative analysis is also feasible. 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. Further extensions of this amplification method to the analysis

The ability to modulate the total amount of the amplification product as determined by the amount of components such as TSO in the reaction mixture can facilitate more accurate detection and/or quantification of target polynucleotides. Excessive amplification produced by other known amplification methods, such as PCR, NASBA, TMA, etc. can pose problems for detection and/or quantification using methods employing two or more detection probes. The "hook effect" from binding excess analyte is well known in the art and results in non-linearity and signal drop-offs leading to false negatives or reduced signal strengths relative to expected levels at high analyte concentrations. The present invention has the advantage of limiting the maximum amount of product generated as appropriate for the chosen method for detection of the amplified product.

Amplification methods of the invention have the additional advantage of being resistant to non-specific incorporation of a promoter sequence into primer extension products which can lead to non-specific amplification. In the composite primer-based methods, the promoter sequence is introduced by a TSO and/or another propromoter polynucleotide as described herein. In the propromoter TSO-based methods, the promoter sequence is introduced by a propromoter TSO in the first transcription module, and another propromoter polynucleotide which, as described herein, can be a TSO, in the second transcription module. In the first transcription module of the propromoter TSO-based methods, the formation of a double stranded functional promoter is dependent on primer extension and template switch. The formation of a functional promoter in the second module requires hybridization of the second primer extension product to a propromoter polynucleotide such as a propromoter TSO. The step of propromoter polynucleotide hybridization in the methods of the invention increases the specificity of the process. A non-specific extension product would not be likely to have a sequence (e.g., a 3' end sequence) which is complementary to a propromoter polynucleotide designed to be hybridizable to a particular sequence in a primer extension product.

The methods are applicable to amplification of either DNA or RNA target molecules. The amplification methods of the invention can be used for, for example, detection, quantification and sequencing of a target sequence.

IDT Exhibit 1020.8739

Composite primer-based amplification methods

In one aspect, methods for generating multiple copies of sense RNA transcripts of a nucleic acid sequence of interest using composite primers are provided. The composite primer-based amplification methods of the invention provide for isothermal exponential amplification of a nucleic acid sequence of interest. The methods utilize a composite primer and generally a second primer which is generally, but not necessarily, non-composite. In one embodiment, the methods also employ a termination sequence, such as a blocker sequence (as described in Method 1) or a TSO (as described in Method 2). Methods 1 and 2 are described below.

The termination sequence (either TSO or blocker sequence component, if used) is added for generating a product with a defined 3'-end. In some embodiments, natural sequence(s) within the template 5' of the primer binding site inhibits nucleic acid polymerization such that termination of primer extension is achieved. Such natural sequences are known in the art (for example, GC rich sequences) or can be empirically determined. Use of a termination sequence is particularly beneficial when amplifying genomic DNA so as to provide a defined end for primer extension. When this feature is not desired, the isothermal amplification according to the methods of the invention can be carried out without a termination sequence.

The isothermal amplification utilizes three enzymes, a DNA polymerase, a ribonuclease such as RNaseH, and an RNA polymerase. One embodiment of method 1, which utilizes a blocker sequence, is illustrated in Figure 1A-F. An embodiment of method 2, which utilizes a TSO, is illustrated in Figure 2A-E.

The composite primer-based methods of the invention are generally designed to amplify a single stranded DNA target. When the target nucleic acid to be amplified is double stranded DNA, the target is first denatured to produce a single stranded target. Target denaturation may be carried out using methods known in the art, such as heat denaturation or alkali treatment. When the target is single stranded RNA, such as mRNA or viral RNA, the target is first transcribed to produce a cDNA target by methods known in the art.

The target polynucleotide is combined with a composite primer, DNA polymerase, a ribonuclease such as RNaseH and optionally, a blocker sequence component or a TSO, as described above. In one embodiment, each amplification reaction includes composite

IDT Exhibit 1020.8740

primers of identical sequence. In another embodiment, each amplification reaction includes a mixture of composite primer variants, wherein the variants represent two or more homologous but non-identical sequences, and wherein all are capable of hybridizing to the same target nucleic acid sequence. The complementarity is preferably at least about 50%, more preferably at least about 75%, and most preferably at least about 90%. Advantages of this embodiment include the ability to introduce different sequences of interest into the primer extension products. In yet another embodiment, each amplification reaction includes a mixture of composite primers, wherein the primers represent two or more non-identical sequences that are of low or no homology, and wherein the primers preferentially hybridize to different target nucleic acid sequences or different sites along the same target nucleic acid strand. Advantages of this embodiment include multiplex detection and/or analysis of target nucleic acids through amplification of a plurality of target nucleic acid species in a single amplification reaction.

Methods that utilize a blocker sequence (or that do not utilize a termination polynucleotide) result in the generation of a displaced primer extension product which when hybridized to a propromoter polynucleotide forms a complex, which is a substrate for the RNA polymerase. The methods that utilize a TSO result in the generation of a unique intermediate amplification product comprising target and template switch oligonucleotide (TSO)-related portions. The complex formed by the hybridization of a TSO to a displaced primer extension product is a substrate for transcription by the RNA polymerase, which generates an RNA product of the same sense as the initial target sequence.

Two exemplary embodiments of the composite primer-based methods are described below:

Method 1: Blocker sequence-based polynucleotide amplification

In some embodiments, blocker sequence-based amplification methods are provided. An illustrative embodiment is shown in Figures 1A-F.

Figures 1A-F illustrate an embodiment of the method that utilizes a composite primer, as described herein; a blocker sequence component which is either an oligonucleotide or an oligonucleotide analog which, as described herein, is further able to hybridize to a sequence on the same target nucleic acid strand as the single primer; and a third oligonucleotide, a promoter template (PTO) which, as described herein, comprises a 3'-portion which is able to hybridize (and is preferably complementary) to the 3'-end of the

IDT Exhibit 1020.8741

displaced extension product and a 5'-portion which includes a sequence of a DNA-dependent RNA polymerase promoter. As in the TSO described herein, the sequence immediately adjacent to the promoter sequence can be designed to provide for preferably optimal transcriptional activity by the RNA polymerase used in the amplification according to the method of the invention. The blocker sequence component is generally designed to hybridize to the target sequence at a site which is upstream, i.e., located towards the 5' end of the target relative to the hybridization site of the single primer. Stated alternatively, and as described herein, the blocker sequence hybridizes to a segment of the target nucleic acid sequence which is 5' to the target sequence complementary to the 3' end of the primer extension product. The blocker sequence binds with sufficiently high affinity so as to terminate primer extension at the site where the blocker hybridizes to the target. This feature provides a strong stop for primer extension by the polymerase and defines the 3'-end of the primer extension product.

Suitable reaction medium and conditions are as described herein. In one embodiment, transcription is performed at a different temperature, generally lower, than that of the preceding steps. In another embodiment, all steps of the method are performed isothermally.

The single stranded nucleic acid target is combined with the composite primer, the blocker component, the propromoter template (PTO), DNA polymerase, ribonuclease such as RNaseH, a DNA dependent RNA polymerase, and nucleotides, such as NTPs (e.g., dNTPs and rNTPs). The composite primer and the blocker sequence component hybridize to the same target strand to form a tri molecular complex I (Figure 1A). The primer is extended along the target up to the site of hybridization of the blocker sequence, to form complex II (Figure 1A).

A ribonuclease, such as RNaseH, cleaves the RNA portion, generally the 5' RNA portion, of the single composite primer of complex II to form complex III (Figure 1B). As described herein, the ribonuclease is specific for cleaving the RNA strand of an RNA/DNA hybrid, and does not digest single stranded RNA. Thus, the ribonuclease does not degrade the free composite primer. Free primer hybridizes to the primer complementary site of the target nucleic acid in complex III (Figure 1B). This hybridization results in formation of complex IV (Figure 1B), in which only the 5' RNA portion of the primer is hybridized to the target strand. Displacement of the 5' most portion of the primer extension product by the 3' DNA portion of the partially hybridized primer will result in formation of complex V

IDT Exhibit 1020.8742

(Figure 1C), a substrate for a DNA polymerase. Extension of the primer along the target strand results in displacement of the first primer extension product (VI) from the complex. Cycles of the process of RNaseH cleavage, hybridization of a free composite primer and generation of strand displacement product proceeds to form multiple single stranded DNA products (VII). The displaced extension products are fully complementary to the target sequence and does not comprise a 3' end portion which is not complementary to the target.

The promoter template oligonucleotide (PTO) binds to the displaced extension product to form complex VIII (Figure 1C), by hybridization of the 3' end portion (A) of the propromoter template to the 3' end of the displaced primer extension product. As described herein, the 3' end of the PTO may or may not be blocked. When the 3' end of the propromoter template is not blocked, the template will be extended along the displaced primer extension product. The 3' end of the displaced extension product will be extended by the nucleotide (DNA) polymerase along the B portion of the hybridized propromoter template to form complex IX (see Figure 1D), which comprises at one end a double stranded promoter sequence for the DNA-dependent RNA polymerase. Complex IX is depicted in Figure 1D as the product of hybridization of a promoter template (PTO) in which the 3' end is blocked for extension by the polymerase. Alternatively, when the 3' end of the promoter template is not blocked, extension of the 3' end along the displaced primer extension product results in formation of a fully double stranded complex. The DNA-dependent RNA polymerase will transcribe both forms of the extended displaced primer extension product of complex IX. The choice of RNA polymerase takes into account its capability to transcribe either the partial duplex or the fully double stranded duplex forms of the complex. Multiple copies of a sense single stranded RNA (X) are produced by this transcription step. (Figure 1D).

The production of preferably at least about 1, more preferably at least about 50, even more preferably at least about 75, still more preferably at least about 100, and most preferably at least about 1000, RNA transcript products from each primer extension product is expected.

As illustrated in Figure 1E, the second set of steps (second transcription module) utilize a second oligonucleotide primer, 4, which is the same sense as the composite primer relative to the target nucleic acid sequence, the promoter template oligonucleotide as used in the first module, an RNA-dependent DNA polymerase or reverse transcriptase (RT), a DNA-dependent DNA polymerase, RNaseH and a DNA-dependent RNA polymerase. The

IDT Exhibit 1020.8743

DNA polymerase, reverse transcriptase, and the RNaseH activities may be contained in the same enzyme, as is known in the art.

The two modules may be combined or carried out sequentially. For the purpose of simplicity, the processes are described herein as sequential steps. An aliquot of the reaction mix of the first set of steps is combined with a mixture containing the above mentioned components and the amplification reaction carried out isothermally.

The second primer, 4, hybridizes to a sequence at the 3' end of the RNA product of the first module. The second primer is thus generally designed to be complementary to a sequence which is upstream of the sequence which is complementary to the composite primer used in the first set of steps, e.g. a nested primer. The second primer is generally, although not necessarily, a DNA oligonucleotide (which is generally, but not necessarily, non-composite). The hybridized primer is extended by RT to form a RNA/DNA double stranded complex XI (Figure 1E). The RNA strand of complex XI is cleaved by RNaseH to produce a single stranded DNA product XII (Figure 1E), which is anti-sense relative to the target nucleic acid sequence (i.e., is complementary), similar to the primer extension product of the first set of steps.

The 3'-end of primer extension product XII is generally complementary to the 3'-end of the promoter template oligonucleotide (as is the primer extension product of the first set of steps). Hybridization of product XII to the promoter template oligonucleotide results in formation of complex XIII (Figure 1E). The 3'-end of the primer extension product is extended along the promoter template oligonucleotide by RT to form the partial duplex XIV (Figure 1F) which has a double stranded promoter sequence at one end. Partial duplex XIV is a substrate for transcription by a DNA-dependent RNA polymerase, to form sense RNA products X (Figure 1F).

Hybridization of primer 4 to product X will reinitiate the process described for the second transcription module, thus creating a cyclical amplification process, or exponential amplification. In the second set of steps, the extent of amplification of the initial sequence of interest will be limited by the concentration of the promoter template oligonucleotide and the second primer.

As evident from the above description, methods of the invention are suitable for highly efficient amplification of target nucleic acid as well as for applications of the amplification methods for purposes such as isothermal sequencing described herein.

IDT Exhibit 1020.8744

Method 2: TSO-based enhanced linear nucleic acid amplification

In one embodiment, the TSO-based linear amplification method of the present invention is linked to transcription from the primer extension products to provide enhanced nucleic acid amplification. A schematic description of one embodiment of this novel amplification method, Method 2, is shown in Figures 2A-2E.

The TSO-based nucleic acid amplification method of the invention employs a composite primer, as described herein. A second oligonucleotide used in the amplification method of the invention is a template switch oligonucleotide (TSO), as described above. The nucleic acid target to be amplified can be DNA or RNA. Amplification of an RNA target will require an initial cDNA synthesis step according to methods known in the art. The TSO-based enhanced linear amplification method of the present invention produces multiple copies of an RNA product homologous (i.e., sense) to the target DNA sequence.

The single stranded target nucleic acid is combined with the composite primer, a TSO oligonucleotide, DNA polymerase, ribonuclease such as RNaseH, a DNA dependent RNA polymerase, and nucleotides, such as deoxyribonucleoside triphosphates (dNTPs) and ribonucleoside triphosphates (rNTPs) in a reaction medium suitable for nucleic acid hybridization and amplification, as is known in the art. In one embodiment, transcription is performed at a different temperature, generally lower, than that of the preceding steps. In another embodiment, all the steps of the methods are performed isothermally.

In one embodiment, the TSO functions as a termination sequence and provides a propromoter sequence. In another embodiment, the TSO does not comprise a propromoter sequence. In this embodiment, a propromoter sequence is provided separately by another oligonucleotide, such as a PTO, that comprises a propromoter sequence and is hybridizable to the 3' portion of the primer extension product such that transcription of the primer extension product can occur.

The single composite primer and the TSO then hybridize to the same strand of the nucleic acid to be amplified. The two oligonucleotides may be added to the sample suspected of containing the nucleic acid target prior to the target nucleic acid denaturation step. Hybridization of the two oligonucleotides to the target strand results in the formation of the tri molecular complex XV (Figure 2A).

A DNA polymerase carries out primer extension. The primer is extended along the target nucleic acid strand of complex XV (Figure 2A) and terminates at the site of TSO

IDT Exhibit 1020.8745

hybridization. Template switching from the target strand to the 5' unhybridized portion of the TSO, and further primer extension along the TSO template results in the formation of the tri molecular complex XVI which comprises a target nucleic acid, the TSO and the first primer extension product. The first primer extension product is a unique DNA comprising both a target dependent portion (i.e., sequence complementary to the target nucleic acid) and a TSO dependent portion (i.e., sequence complementary to the unhybridized portion of the TSO) .

Complex XVI (Figure 2A) is a substrate for both an RNA polymerase and a ribonuclease such as RNaseH. The DNA dependent RNA polymerase binds to the functional double stranded promoter of complex XVI and transcribes the first primer extension product to produce a sense RNA product X (Figure 2A). A ribonuclease, such as RNaseH, which is specific for degradation of the RNA strand of an RNA/DNA heteroduplex, degrades the 5' portion of the primer extension product in complex XVI to form the tri molecular complex XVII (Figure 2B).

Free composite primer hybridizes to the primer complementary site of the target nucleic acid in complex XVII (Figure 2B). This hybridization results in formation of complex XVIII (Figure 2B) in which only the RNA portion, generally the 5' RNA portion, of the primer is hybridized to the target strand. Displacement of the 5' most portion of the primer extension product by the 3' DNA portion (XVIII) of the partially hybridized primer will result in formation of complex XIX (Figure 2B), which is a substrate for a DNA polymerase. Extension of the primer along the target strand (XIX; Figure 2B) results in displacement of the first primer extension product XX from the complex. Repeated primer extensions and strand displacements result in generation of multiple copies of polynucleotides (XX; Figure 2C) that are substantially complementary to the target nucleic acid.

The primer extension product generated as described above is used as a template for transcription in the embodiment wherein TSO that comprises a propromoter sequence is provided. The displaced primer extension product (XX; Figure 2C) hybridizes to free TSO oligonucleotide to form the partial duplex XXI (Figure 2C). Complex (duplex) XXI comprises a double stranded portion at one end and two non-complementary single strands respectively derived from the primer extension product and the TSO. The double stranded portion of this partial duplex contains a fully functional double stranded promoter for the DNA dependent RNA polymerase. The latter recognizes the promoter of the partial duplex

IDT Exhibit 1020.8746

XXI and transcribes the primer extension product to form multiple copies of a sense RNA product X (Figure 2C).

As described in method 1 and shown in Figures 2D-E, the isothermal amplification described above is linked to further amplification by the combination of the reaction mixture with a second primer, 4, RT, RNaseH, a DNA dependent RNA polymerase and the required nucleotide triphosphates. The second primer 4 hybridizes at the 5' portion of the sense RNA product, as in method 1, and is extended by RT to form an RNA/DNA hybrid duplex XXII (Figure 2D). RNaseH cleaves the RNA strand of XXII to generate a single stranded anti sense DNA primer extension product XXIII (Figure 2D). The 3'-end of the primer extension product is complementary to a sequence on the TSO. Hybridization of the primer extension product and free TSO results in formation of complex XXIV (Figure 2D). DNA polymerase extends the 3' end of the primer extension product along the 5'-end of the TSO to form partial duplex XXV (Figure 2E). The last is a substrate for DNA dependent RNA polymerase insofar as it comprises a double stranded promoter sequence at one end. Transcription of the primer extension product results in the generation of multiple copies of sense single stranded RNA product X (Figure 2E).

Hybridization of the second primer 4 to the RNA product and extension by RT reinitiates the process as described above, thus enabling a cyclical amplification of the target nucleic acid sequence, as described for method 1.

Propromoter TSO-based amplification methods

Propromoter TSO-based methods of the invention are based on the formation of a unique primer extension product using a template switch mechanism. The nucleic acid amplification methods of the invention encompass both primer extension by nucleotide polymerase and primer-independent transcription. The methods employ a single primer and a propromoter TSO capable of effecting template switching during primer extension. Both the primer and propromoter TSO are complementary to the same target nucleic acid strand. The propromoter TSO comprises a 3' portion which is complementary to the target polynucleotide; and a 5' portion which is not hybridizable to the target (under a given set of conditions), wherein the non-hybridizable 5' portion comprises a sequence (generally, a 5' most sequence) that is a single strand of a promoter for a DNA dependent RNA polymerase used in the amplification steps according to the methods of the invention. In some embodiments the sequence of the propromoter TSO which is immediately adjacent to the

IDT Exhibit 1020.8747

propromoter sequence (for example, 3' with respect to the propromoter sequence), is designed to maximize transcription by the RNA polymerase. Such sequences are known in the art and are described herein. TSO sequences capable of effecting template switch can be designed according to criteria known in the art (see Patel et. al., *Proc. Natl. Acad. Sci. USA* (1996), 93:2969-2974); U.S. Pat. Nos. 5,679,512; 5,683,879; & 6,030,774.

The propromoter TSO-based methods of amplification generally utilize several enzymes: a DNA-dependent DNA polymerase, a DNA-dependent RNA polymerase, an enzyme that cleaves RNA in an RNA/DNA hybrid (such as RNaseH), and an RNA-dependent DNA polymerase (such as reverse transcriptase). The enzymatic activities may be provided in individual enzymes or in various combinations (i.e., in one or more enzymes with multiple activities). The enzymes used for the amplification according to the methods of the invention are commercially available, as known to people skilled in the art.

The primer and the propromoter TSO are generally provided at high concentrations similar to those commonly used in PCR methods. The level of amplification of the target polynucleotide is generally limited by the concentration of these oligonucleotides. Whereas the primer is consumed by incorporation into the extension products, the propromoter TSO is not incorporated into the extension products. As illustrated in Figures 3A-B & 5A-B, the propromoter TSO participates in the initial template switch for the formation of a unique complex that is a substrate for DNA-dependent RNA polymerase (a transcription process). In certain embodiments, the propromoter TSO also participates, as illustrated in Figure 3B, in binding to a single stranded DNA product (reverse transcribed from an RNA transcript from the first transcription module) for the formation of a second substrate for transcription. It is understood that any polynucleotide that comprises a propromoter sequence and a region capable of hybridizing to said single stranded DNA product can be used. For example, a PTO may be used.

Unlike other transcription-based isothermal amplification schemes, such as NASBA, TMA and the like, the propromoter TSO-based methods of the invention do not require the formation of a fully double stranded DNA product, which is an essential intermediate in these previously described transcription-based nucleic acid amplification methods. Since a double stranded intermediate DNA product is not required, forward and reverse primer pairs (for primer extensions along a target polynucleotide strand and its complement) are not used. This reduces the complexity of the primer extension reactions by eliminating the formation of primer-dimers where two or more primers hybridize and

IDT Exhibit 1020.8748

are extended to form double stranded substrates for transcription comprising double stranded promoters in a target-independent manner. The ability to use a single primer as described in the present invention reduces the complexity of the process. The process is further simplified by the ability to use a promoter template that does not have to be extended, for example, by using the same propromoter TSO in both the first primer extension step and in hybridization to a second primer extension product.

An illustrative embodiment of propromoter TSO-based methods of the invention

The amplification of target nucleic acid is generally based on a single stranded target, and so is typically preceded by rendering a target single stranded (if it is not already single stranded). Therefore, if the target nucleic acid is a double stranded DNA, the first step of the amplification is denaturation to yield a single stranded target. When the target nucleic acid is RNA, amplification can proceed directly from the single stranded target RNA, as shown below. Alternatively, the amplification may be preceded by synthesis of cDNA from the RNA target by methods known in the art.

As illustrated in Figure 3A-B, in one embodiment of the methods of the invention, a TSO and a single primer hybridize to a target nucleic acid strand to form the tri molecular complex I. The target strand of the complex may be DNA, as illustrated in Figure 3A-B, or RNA, as illustrated in Figure 4A-B.

A DNA polymerase carries out primer extension along the target strand of complex I. When the target is DNA, the enzyme is generally a DNA-dependent DNA polymerase, and when the target is RNA, the enzyme is generally an RNA-dependent DNA polymerase such as reverse transcriptase.

As illustrated in Figures 3A-B & 4A-B, primer extension is switched from the target to the 5' non-hybridized portion of the propromoter TSO at the point where the TSO is bound to the target thus effectively terminating replication of the target-specific sequence. Extension along the 5' non-hybridized portion of the TSO results in the formation of a unique extension product which is composed of a target-related portion and a TSO-related portion. The unique tri-molecular complex II (Figures 3A & 4A) of the target, TSO and the first primer extension product, comprises a double stranded promoter region, thus making it a substrate for a DNA-dependent RNA polymerase.

IDT Exhibit 1020.8749

As illustrated in Figure 4A-B, when the target is RNA, the RNA strand of complex II can be (but is not necessarily) degraded by RNaseH to form complex IIa, comprising the propromoter TSO and the first primer extension product. This complex further comprises a double stranded promoter and is a substrate for an RNA polymerase.

A DNA-dependent RNA polymerase binds to the double stranded promoter of complexes II (Figure 3) and IIa (Figure 4), and transcription of the unique primer extension product is initiated. This process results in the generation of multiple copies of a single stranded RNA product which are of the same sense as the target polynucleotide.

A second primer hybridizes to the sense RNA product to form complex III (Figures 3A & 4B). The second primer is generally and preferably the same as the first primer. In embodiments wherein the first and second primers are different, the second primer is generally of the same polarity as the first primer. Primer extension along the RNA product is carried out by an RNA-dependent DNA polymerase such as reverse transcriptase, yielding a DNA/RNA heteroduplex IV (Figures 3A & 4B).

RNaseH digestion of the RNA strand of complex IV results in the formation of a single stranded DNA product of primer extension by the RNA-dependent DNA polymerase (the second primer extension product). A free propromoter polynucleotide (shown in Figure 3B as a propromoter TSO) hybridizes to the 3' portion of the single stranded DNA product to form complex V (Figure 3B). If the propromoter polynucleotide is a TSO used in the first step, a sequence on the second primer extension product (DNA) which is complementary to a sequence on the TSO would have been incorporated into the second primer extension product by the template switch process.

The binding of a propromoter polynucleotide (illustrated in Figure 3B as a TSO) to the DNA product results in a 5' overhang of the TSO, so the 3' end of the single stranded DNA product of complex V is extended along the propromoter polynucleotide (TSO) to replicate the promoter sequence. This results in the formation of complex VI (Figure 3B) which contains a functional double stranded promoter sequence. Complex VI is a substrate for a DNA-dependent RNA polymerase. The polymerase binds to the double stranded promoter and transcribes the second primer extension product to produce multiple copies of the sense RNA product. The RNA products are substrates for further primer hybridization and extension as described above. Thus, cyclical amplification proceeds to form multiple copies of the sense RNA products.

IDT Exhibit 1020.8750

Unless terminated deliberately, amplification can be expected to proceed until all free TSO is hybridized to the second primer extension product, or the primer is fully consumed.

As illustrated in Figures 3A-B, once formed following the transcription of the first primer extension products, the RNA product of amplification of either DNA or RNA target is the same, and the amplification schemes from that step on are the same.

Each transcription step (i.e., transcription of the first or second primer extension product) is expected to result in the production of preferably at least about 1, more preferably at least about 100, more preferably at least about 500, and most preferably at least about 1000, copies of a single stranded RNA product from a starting template (in the first transcription module, the template would be the target polynucleotide; in the second transcription module, the template would be an RNA transcript generated in the first transcription module).

The amplification products of the amplification method of the invention are single stranded and are thus readily detectable by any of a variety of methods known in the art. Both homogeneous and heterogeneous methods for detection of single stranded nucleic acids are known and would be suitable for the detection and quantification of the reaction products of the amplification method of the invention.

Components and reaction conditions used in the methods of the invention

Template nucleic acid

The nucleic acid 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. Obtaining and purifying nucleic acids use standard techniques in the art. In some embodiments, amplification of an RNA target will require initial cDNA synthesis, as known in the art. Amplification of a DNA-RNA hybrid would require denaturation of the hybrid to obtain a ssDNA or ssRNA, or denaturation followed by reverse transcription to obtain a cDNA. The target nucleic acid

IDT Exhibit 1020.8751

can be only a minor fraction of a complex mixture such as a biological sample and can be obtained from various biological material by procedures well known in the art.

The initial step of the amplification of a target polynucleotide is rendering the polynucleotide single stranded. If the target polynucleotide is double stranded, the initial step is target denaturation. The denaturation step may be thermal denaturation or any other method known in the art, such as alkali treatment.

Composite primer

In one aspect, methods of the invention employ a composite primer that is composed of RNA and DNA portions. The composite design of the 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 primer, as described below.

Composite primers for use in the methods and compositions of the present invention comprise at least one RNA portion that is capable of (a) binding (hybridizing) to a sequence on the target nucleic acid (template) independent of hybridization of the DNA portion(s) to a sequence on the target nucleic acid; and (b) being cleaved with a ribonuclease when hybridized to the target DNA. The composite primers bind to the target nucleic acid to form a partial heteroduplex in which only the RNA portion of the primer is cleaved upon contact with a ribonuclease such as RNaseH, while the target strand remains intact, thus enabling annealing of another composite primer.

The composite primers also comprise a 3' DNA portion that is capable of hybridization to a sequence on the target nucleic acid (template) such that its hybridization to the target sequence (template) is favored over that of the nucleic acid strand that is displaced from the target nucleic acid 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 target nucleic acid is favored over the hybridization of the homologous sequence in the 5' end of the displaced strand to the target nucleic acid.

IDT Exhibit 1020.8752

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 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).

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, 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 1, 3, 4, 5 nucleotides, with an upper limit of about any of 10, 15, 20, 25, 3, 50 nucleotides.

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, 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 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 nucleotides, with an upper limit of about any of 15, 17, 18, 20, 50 nucleotides.

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

IDT Exhibit 1020.8753

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.

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, 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 certain 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 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, 50 nucleotides.

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,

IDT Exhibit 1020.8754

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.

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, 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, 7, 10 nucleotides, with an upper limit of about any of 10, 12, 15, 18, 20, 22 nucleotides.

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, 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, 7, 10 nucleotides, with an upper limit of about any of 10, 12, 15, 18, 20, 22 nucleotides.

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.

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, 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, 7, 10 nucleotides, with an upper limit of about any of 10, 12, 15, 18, 20, 22 nucleotides.

IDT Exhibit 1020.8755

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 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.

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.

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. 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.

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

IDT Exhibit 1020.8756

primer includes the 5' most nucleotide of the primer. In other embodiments, the 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. It is understood that the length can be greater or less, as appropriate under the reaction conditions of the methods of this invention.

To achieve hybridization to a target nucleic acid (which, as is well known and understood in the art, depends on other factors such as, for example, ionic strength and temperature), the portion of the primer that is hybridizable to the target polynucleotide 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 target polynucleotide.

As described herein, one or more composite primers may be used in an amplification reaction.

Depending on the method of amplification, the first primer may or may not be a composite primer. The composite primer-based methods generally require the use of a first primer that is a composite primer. The propromoter TSO-based methods may or may not utilize a first primer that is a composite primer. In some embodiments, the first primer of propromoter TSO-based methods consists of DNA.

Second primer

The second primer in the methods of the invention 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 sense RNA transcript generated in a first transcription module at a site on the transcript such that the second primer extension product would include the sequence of interest. In some embodiments, the hybridizable sequence of the second primer is designed based on a known sequence of the desired binding site on a sense RNA transcript. In other

IDT Exhibit 1020.8757

embodiments, the hybridizable sequence is based on random sequences. In some embodiments, the second primer is a random primer.

In preferred embodiments of propromoter TSO-based methods, the first and second primers are the same. In other embodiments of propromoter TSO-based methods, the first and second primers are different. In yet other embodiments of propromoter TSO-based methods, the first and second primers hybridize to different complementary sequences.

To achieve hybridization to a sense RNA transcript (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 sense RNA transcript 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 sense RNA transcript.

In one embodiment, the second primer comprises DNA. In another embodiment, the second primer consists of DNA. In another embodiment, the second primer comprises RNA. In still another embodiment, the second primer consists of RNA. In yet another embodiment, the second primer comprises DNA and RNA.

In some embodiments, the second primer is provided by self priming (for example, by a hairpin loop) at the 3' end of the sense RNA transcript. In these embodiments, a sequence at the 3' end of the sense RNA transcript hybridizes to another sequence in the transcript itself. Self priming of polynucleotides has been described in, for example, U.S. Pat. No. 6,132,997. In these embodiments, said sequence at the 3' of the sense RNA transcript is generally cleaved following its hybridization to the transcript and/or its extension along the transcript.

As described herein, one or more second primers may be used in an amplification reaction.

Template switch oligonucleotide (TSO)

One aspect of the invention employs a template switch oligonucleotide (TSO) to effect template switching during primer extension along a target polynucleotide. In some embodiments of composite primer-based methods of the invention, the TSO functions as a termination sequence. In certain embodiments, the TSO functions as a propromoter polynucleotide.

IDT Exhibit 1020.8758

A TSO comprises a 3' portion that can hybridize to the target and a 5' portion which is designed for strand switch during polymerization. Design of a TSO that would effect strand switch is known in the art, such as was previously described by Patel et al., (Proc. Nat'l Acad. Sci. USA , 93:2969-2974 (1996)).

The 3' portion hybridizes to the template at a location 5' to the position or region in the target polynucleotide that is complementary to the 3' end of the primer extension product prior to switching template from the target polynucleotide to the unhybridized portion of the TSO ("termination site").

In one embodiment, strand switch is promoted by the presence of mutually complementary short sequences in the TSO segments immediately 5' and 3' to the junction between the hybridized and non-hybridized portions of the TSO. Without intending to be bound by theory, one explanation is that in the event that the primer extension product is extended into the portion of the target nucleic acid that is hybridized to the TSO (through displacement of the hybridized portion of the TSO), the 3' end of the primer extension product would comprise a short sequence that can bind to its complementary short sequence in the segment of the TSO immediately adjacent to the junction between the hybridized and non-hybridized portions of the TSO. This increases the efficiency of template switching by increasing the probability that the primer extension product would switch to the TSO tail portion as a template. The length of the short complementary sequences is preferably from about 3 to about 20 nucleotides, more preferably from about 5 to about 15 nucleotides, and most preferably from about 7 to about 10 nucleotides. In some embodiments, length is at least about any of 1, 3, 5, 7, 10 nucleotides, with an upper limit of about any of 10, 15, 20, 25 nucleotides. It is understood that the length can be greater or less, as appropriate under the reaction conditions of the methods of this invention.

In some embodiments, the 5' portion of the TSO (referred to herein as a "propromoter TSO") comprises a propromoter sequence that is designed for formation of a double stranded promoter of an RNA polymerase. This embodiment of the TSO can function both as a termination sequence and to provide a promoter template. In this embodiment, the propromoter sequence of the TSO serves as a template for incorporation of a propromoter sequence (generally complementary to the propromoter sequence of the template TSO) into the primer extension product. Subsequent hybridization of a TSO comprising a propromoter sequence that is hybridizable to the propromoter sequence of the primer extension product results in formation of a double stranded promoter capable of

IDT Exhibit 1020.8759

effecting transcription by a suitable RNA polymerase. Promoter sequences that allow transcription of a template DNA are known in the art, as are methods of obtaining and/or making them. 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, are 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.

In one embodiment, the promoter sequence is adjacent to a sequence that is designed to provide for enhanced, or more optimal, transcription by the RNA polymerase used. In some embodiments, 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 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.

In a preferred embodiment, a segment of the 3' portion of the TSO (including the entire 3' portion that hybridizes to target) that hybridizes to the template DNA is attached to the template DNA such that displacement of the TSO by the polymerase that effects primer extension is substantially, or at least sufficiently, inhibited. Suitable methods for achieving such attachment includes techniques known in the art, such as using a cytosine analog that contains a G-clamp heterocyclic modification (described in Flanagan et al., Proc. Natl. Acad. Sci. USA 1999, 96(7):3513-8); and locked nucleic acids (described, e.g., in Kumar et al., Bioorg. Med. Chem. Lett. 1998, 8(16):2219-22; and Wahlestedt et al., Proc. Natl. Acad. Sci. USA 2000, 97(10):5633-8). Other suitable methods include using, where appropriate, sequences with a high GC content and/or cross-linking. Any of these methods for obtaining enhanced attachment may be used alone or in combination. Displacement of the TSO is substantially or sufficiently inhibited if the polymerase switches template from the target nucleic acid strand to the unhybridized portion of the TSO in at least about 25%, preferably at least about 50%, more preferably at least about 75%, and most preferably at least about 90%, of the events of primer extension. Substantially or sufficiently inhibited TSO displacement can also be empirically indicated if the amplification methods lead to a satisfactory result in terms of amount of the desired

IDT Exhibit 1020.8760

product. Generally, under a given set of conditions, the "modified" TSO binds more tightly to template as compared to a TSO not so modified.

The length of the TSO portion that hybridizes to the target nucleic acid strand is preferably from about 15 to 50 nucleotides, more preferably from about 20 to 45 nucleotides, and most preferably from about 25 to 40 nucleotides. In other embodiments, the length is at least about any of the following: 10, 15, 20, 25, 30; and less than about any of the following: 35, 40, 45, 50, 55. It is understood that the length can be greater or less, as appropriate under the reaction conditions of the methods of this invention. The complementarity of the TSO portion that hybridizes to the target nucleic acid strand 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 target nucleic acid.

Polynucleotide comprising a termination polynucleotide sequence

In some embodiments of the methods of the present invention, a polynucleotide comprising a termination sequence is included, examples of which are provided below.

(i) TSO

As described above, a TSO can function as a polynucleotide comprising a termination sequence to affect primer extension along a target polynucleotide.

(ii) Blocker sequence

In some embodiments, the primer extension termination sequence is provided by a blocker sequence. The blocker sequence is a polynucleotide, usually a synthetic polynucleotide, that is single stranded and comprises a sequence that is hybridizable, preferably complementary, to a segment of target nucleic acid sequence 5' of the position in the target sequence that is complementary to the 3' end of the primer extension product ("termination site"). The blocker comprises nucleotides that bind to the target nucleic acid with an affinity, preferably a high affinity, such that the blocker sequence resists displacement by DNA polymerase in the course of primer extension, in preferably more than about 30%, more preferably more than about 50%, even more preferably more than about 75%, and most preferably more than about 90%, of primer extension events. The length and composition of the blocker polynucleotide should be such that excessive random non-specific hybridization is avoided under the conditions of the methods of the present

IDT Exhibit 1020.8761

invention. The length of the blocker polynucleotide is preferably from about 3 to about 30 nucleotides, more preferably from about 5 to about 25 nucleotides, even more preferably from about 8 to about 20 nucleotides, and most preferably from about 10 to about 15 nucleotides. In other embodiments, the blocker polynucleotide is at least about any of the following: 3, 5, 8, 10, 15; and less than about any of the following: 20, 25, 30, 35. It is understood that the length can be greater or less as appropriate under the reaction conditions of the methods of this invention. The complementarity of the blocker polynucleotide 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 target nucleic acid.

In one embodiment, the blocker sequence comprises a segment that is attached to the target DNA such that displacement of the blocker sequence by the polymerase that effects primer extension is substantially, or at least sufficiently, inhibited. Suitable means for achieving such attachment and determining substantial, or sufficient, inhibition of displacement are as described above for TSO used in the methods of the present invention.

In one embodiment, the blocker polynucleotide cannot function efficiently as a primer for nucleic acid extension (i.e., extension from the blocker sequence is reduced, or inhibited). Techniques for blocking the primer function of the blocker polynucleotide include any that prevent addition of nucleotides to the 3' end of the primer 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 blocker polynucleotide that is not capable of anchoring addition of nucleotides by a DNA polymerase.

Polynucleotide comprising a propromoter and a region which hybridizes to a primer extension product

Some embodiments 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. In other embodiments, the propromoter polynucleotide is provided as a TSO (as described herein). In yet other embodiments, a propromoter polynucleotide comprises a portion (generally a 3' portion) that comprises a termination sequence (which does not substantially effect template switch), and a portion (generally a 5' portion) that comprises a

IDT Exhibit 1020.8762

propromoter sequence, wherein the portion that comprises a propromoter sequence is generally not hybridizable to a target polynucleotide (under conditions wherein the portion that comprises a termination sequence is hybridizable to a target polynucleotide). In some embodiments wherein a propromoter polynucleotide that comprises a portion comprising a termination sequence (which does not substantially effect template switch) and a portion comprising a propromoter sequence is provided, the propromoter polynucleotide may function both to effect termination of primer extension and to provide a propromoter sequence in the same amplification reaction.

Propromoter template oligonucleotide

In some embodiments, the methods employ a promoter sequence for transcription which is provided by a propromoter template oligonucleotide (PTO).

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 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 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 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.

Promoter sequences that allow transcription of a template polynucleotide 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

IDT Exhibit 1020.8763

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.

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 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. 5766849 and 5654142, and can also be empirically determined.

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.

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.

IDT Exhibit 1020.8764

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.

Combination polynucleotide comprising a termination sequence and a propromoter sequence

In some embodiments of methods of the invention, a termination sequence and a propromtoer sequence are provided in a single combination polynucleotide. The combination polynucleotide comprises a portion (generally a 3' portion) that comprises a termination sequence that does not effect template switch under conditions wherein the termination sequence is hybridizable to a target polynucleotide, and a portion (generally a 5' portion) that comprises a propromoter sequence, wherein the portion that comprises a propromoter sequence is generally not hybridizable to the target polynucleotide (under conditions wherein the portion that comprises a termination sequence is hybridizable to the target polynucleotide). A termination sequence can be designed so as not to effect template switch using techniques known in the art, for example by ensuring that design characteristics that are known to promote template switch (such as described in Patel et al., Proc. Nat'l Acad. Sci. USA 1996, 93:2969-2974) are not present in the combination polynucleotide. The combination polynucleotide is hybridizable to the sequence of the template which is in the 5' direction with respect to the template sequence which is hybridizable to the primer. The polynucleotide further comprises a sequence which is hybridizable to a complementary sequence of the target polynucleotide. The sequence that is hybridizable to a complementary sequence of the target polynucleotide may be non-overlapping, overlapping or co-extensive with the termination sequence and/or propromoter sequence of the combination polynucleotide. Generally and preferably the sequence that is hybridizable to a complementary sequence of the target polynucleotide is hybridizable to a 3' portion of the complementary sequence of the target polynucleotide. Thus, in some

IDT Exhibit 1020.8765

embodiments of methods of the invention, a combination polynucleotide that comprises a portion comprising a termination sequence and a portion comprising a propromoter sequence functions both to effect termination of primer extension and to provide a propromoter sequence in the same amplification reaction.

5

DNA polymerase, ribonuclease and RNA polymerase

The amplification methods of the invention employ some or all of the following enzymes: an RNA-dependent DNA polymerase, a DNA-dependent DNA polymerase, a ribonuclease such as RNaseH, and a DNA-dependent RNA polymerase. One or more of
10 these activities may be found and used in a single enzyme. For example, RNaseH 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 RNaseH activity.

One aspect of the invention is the formation of a single stranded cDNA from a
15 primer-RNA complex. This process generally utilizes the enzymatic activities of an RNA-dependent DNA polymerase and a ribonuclease activity.

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 RNA polymerase is one that is capable
20 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. 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,
25 polymerase activity and ribonuclease activity) and can function in the formation of cDNA molecules. However, in some instances, it is preferable to employ a reverse transcriptase which lacks the RNaseH activity. Reverse transcriptase devoid of RNaseH activity are known in the art, including those comprising a mutation of the wild type reverse transcriptase where the mutation eliminates the RNaseH activity. In these cases, the
30 addition of an RNaseH from other sources, such as that isolated from *E. coli*, can be employed for the formation of cDNA.

IDT Exhibit 1020.8766

DNA-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 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. Amplification of a polynucleotide according to certain 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 polymerase preferably has little or no 5'->3' exonuclease activity so as to 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. 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.

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), and thermostable DNA polymerases from Thermoanaerobacter thermohydrosulfuricus.

The ribonuclease for use in the methods and compositions of the invention is capable of cleaving ribonucleotides in an RNA/DNA hybrid. Preferably, the ribonuclease cleaves ribonucleotides regardless of the identity and type of nucleotides adjacent to the

IDT Exhibit 1020.8767

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 (RNaseH).

The DNA-dependent RNA polymerase 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.

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

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. 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

IDT Exhibit 1020.8768

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 propromoter polynucleotide) 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, 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.

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. Nucleotides and/or analogs, such as ribonucleoside triphosphates, that can be employed for synthesis of the RNA transcripts in the methods of the invention 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.

IDT Exhibit 1020.8769

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 propromoter polynucleotide 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.

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. Thus, for example, it is possible to mix a sample containing the target with a primer and the propromoter TSO in the propromoter TSO-based methods, prior to formation of a single stranded species from a double stranded DNA target. Depending on the heat stability of the amplification enzymes used, the enzymes required for the amplification according to the present invention can be added to the sample either prior to the initial step of formation of single stranded target, or at a later stage. In some embodiments, it may be desirable to include a nucleic acid modification enzyme (as described herein) early in the amplification process, which would generally require the use of thermostable nucleic acid modification enzymes, which are known in the art. These enzymes include, for example, thermostable DNA ligases (such as Thermus thermophilus (Tth) ligase; Thermus sp. AK16D ligase; Aquifex aeolicus ligase); thermostable DNA polymerases (such as Taq, Tth or Pfu DNA polymerase); and thermostable reverse transcriptase (such as rTth RNA-dependent DNA polymerase).

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 first primer extension. Another timepoint is at the end of second primer extension. Methods for stopping the reactions are known in the art, including, for

IDT Exhibit 1020.8770

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. Alternatively, the reaction can be allowed to proceed (i.e., from start to finish) without interruption.

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.

Compositions and kits of the invention

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.

For example, the invention provides compositions comprising any two, preferably three, more preferably four, of the following: (a) a first primer which is a composite primer; (b) a second primer; (c) a DNA-dependent DNA polymerase; (d) an RNA-dependent DNA polymerase; (e) a propromoter polynucleotide; (f) an RNA polymerase; and (g) an enzyme that cleaves RNA from an RNA/DNA hybrid. In some embodiments, the composite primer, if included in the compositions, comprises an RNA portion and a 3' DNA portion (in some embodiments, the RNA portion is adjacent to the DNA portion). In other embodiments, the composite primer comprises 5'- and 3'-DNA portions with at least one intervening RNA portion. In other embodiments, the compositions further comprise a TSO (i.e., any of the TSO embodiments described herein, including TSOs containing one or

IDT Exhibit 1020.8771

more modifications which enhance binding to template). In some embodiments, the compositions further comprise a polynucleotide comprising a termination polynucleotide sequence.

In another embodiment, the invention provides compositions comprising any two, preferably three, more preferably four, of the following: (a) a propromoter TSO; (b) a first primer (which may consist of DNA); (c) a DNA-dependent DNA polymerase; (d) an RNA-dependent DNA polymerase; (e) an enzyme that cleaves RNA from an RNA/DNA hybrid; (f) optionally a second primer (which may consist of DNA); and (g) an RNA polymerase.

In some embodiments of the compositions described herein, the compositions further comprise single enzymes comprising multiple enzymatic activities. For example, the DNA-dependent DNA polymerase and the RNA-dependent DNA polymerase may be the same enzyme (such as reverse transcriptase); the DNA-dependent DNA polymerase and enzyme that cleaves RNA from an RNA/DNA hybrid may be the same enzyme; the RNA-dependent DNA polymerase and enzyme that cleaves RNA from an RNA/DNA hybrid may be the same enzyme; or all three activities of RNA-dependent DNA polymerase, DNA-dependent DNA polymerase and enzyme that cleaves RNA from an RNA/DNA hybrid may reside in a single enzyme.

Any of the above compositions may further comprise template (which comprises a target sequence) and/or any of the enzymes described herein (such as DNA polymerase, RNaseH, and/or RNA polymerase). The compositions are generally in aqueous form, preferably in a suitable buffer.

The invention also provides compositions comprising the amplification products described herein. Accordingly, the invention provides a population of sense RNA molecules which are copies 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).

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: amplifying a polynucleotide (RNA and/or DNA)

IDT Exhibit 1020.8772

sequence; sequencing of a nucleic acid sequence of interest; and detection of sequence mutation based on amplifying a nucleic acid sequence.

The kits of the invention comprise one or more containers comprising any combination of the components described herein, and the following are examples of such kits. For example, the invention provides kits comprising any two, preferably three, more preferably four, of the following: (a) a first primer which is a composite primer; (b) a second primer; (c) a DNA-dependent DNA polymerase; (d) an RNA-dependent DNA polymerase; (e) a propromoter polynucleotide; (f) an RNA polymerase; and (g) an enzyme that cleaves RNA from an RNA/DNA hybrid. In some embodiments, the composite primer, if included in the kits, comprises an RNA portion and a 3' DNA portion (in some embodiments, the RNA portion is adjacent to the DNA portion). In other embodiments, the composite primer comprises 5'- and 3'-DNA portions with at least one intervening RNA portion. In other embodiments, the kits further comprise a TSO (i.e., any of the TSO embodiments described herein, including TSOs containing one or more modifications which enhance binding to template). In some embodiments, the kits further comprise a polynucleotide comprising a termination polynucleotide sequence. In some embodiments, the kits further comprise instructions for using any of the components provided in the kits to amplify a target polynucleotide according to composite primer-based methods described herein.

In another embodiment, the invention provides kits comprising any two, preferably three, more preferably four, of the following: (a) a propromoter TSO; (b) a first primer (which may consist of DNA); (c) a DNA-dependent DNA polymerase; (d) an RNA-dependent DNA polymerase; (e) an enzyme that cleaves RNA from an RNA/DNA hybrid; (f) optionally a second primer (which may consist of DNA); and (g) an RNA polymerase. In some embodiments, the kits further comprise instructions for using any of the components provided in the kits to amplify a target polynucleotide according to any of the propromoter TSO-based methods described herein.

Kits may also optionally include deoxynucleoside triphosphates and/or ribonucleoside triphosphates. Kits may also include one or more suitable buffers (as described herein). Kits useful for nucleic acid sequencing may optionally include labeled or unlabelled nucleotide analogs that upon incorporation into a primer extension product effect termination of nucleotide polymerization. One or more reagents in the kit can be provided as a dry powder, usually lyophilized, including excipients, which on dissolution

IDT Exhibit 1020.8773

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 kits of the invention may optionally include a set of instructions, generally written instructions 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. Electronic storage media (*e.g.*, magnetic diskette or optical disk) containing instructions are acceptable modes of providing the instructions. Instructions may also be provided by storing said instructions in a central server, which is accessible via wide area networks (WAN) such as the internet.

The instructions included with the kit generally include information as to reagents (whether or not included 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, the invention provides kits that comprise a first composite primer that comprises a sequence which is hybridizable to a target polynucleotide and instructions for using the primer to amplify a target polynucleotide according to composite primer-based methods described herein. In another example, kits can further comprise a second primer, and optionally instructions for using the primer to amplify a target polynucleotide according to composite primer-based methods described herein. In other examples, the kits can contain further components, such as any of (a) a propromoter polynucleotide (such as a PTO); and (b) any of the enzymes described herein, such as an enzyme which cleaves RNA from an RNA/DNA hybrid (for example, RNaseH), DNA polymerase (RNA-dependent or DNA-dependent) and RNA polymerase, and instructions for using the components to amplify a target polynucleotide according to composite primer-based methods described herein.

The component(s) of the kit may be provide 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, the RNA polymerase (if included) is preferably provided separately from the components used in the steps prior to the transcription steps.

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

IDT Exhibit 1020.8774

to occur to practice the methods disclosed herein and/or to further optimize the sensitivity of any assay. With respect to the compositions and kits of the invention, any component may comprise any of the embodiments of the components as described herein.

The invention also provides systems for effecting the methods described herein. These systems comprise various combinations of the components discussed above. A system may generally include 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 invention also provides reaction mixtures (or compositions comprising reaction mixtures) which contain various combinations of components described herein. An example of a reaction mixture is (a) a complex of a first primer extension product and a target polynucleotide; (b) a polynucleotide comprising a propromoter sequence (for example, a PTO or a TSO); (c) RNA polymerase; and (d) a second primer. Other reaction mixtures are described herein and are encompassed by the invention.

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 1-4.

Methods for using the amplification methods and compositions of the invention

The methods and compositions of the invention can be used for a variety of purposes. For purposes of illustration, methods of sequencing, genotyping (nucleic acid mutation), microarray preparation, and characterizing nucleic acid sequences using the amplified nucleic acid products generated by the methods of the invention, are described.

Sequencing of polynucleotides using the methods of the invention

The amplification methods of the invention are useful, for example, for sequencing of a sequence of interest. The sequencing process is carried out as described for the amplification methods described herein.

Sequencing using the methods of the invention can be based on effecting premature (deliberate) termination of RNA transcription. The inclusion of rNTP analogs, which may be labeled or unlabelled, that upon incorporation into an RNA transcript effects termination of rNTP polymerization in the reaction mixture, will result in production of truncated RNA

IDT Exhibit 1020.8775

products, which result from blocking of the RNA polymerase at sites of incorporation of the analogs.

In one embodiment, the methods of sequencing comprise (a) amplifying a target polynucleotide containing the sequence of interest by the methods described herein in the presence of a mixture of rNTPs and rNTP analogs such that transcription is terminated upon incorporation of an rNTP analog; and (b) analyzing the amplification products to determine sequence.

In another embodiment, the methods of sequencing comprise (a) amplifying a target polynucleotide containing the sequence of interest by the methods described herein, wherein RNA transcripts are generated from the second primer extension product in the presence of a mixture of rNTPs and rNTP analogs such that transcription is terminated upon incorporation of a rNTP analog; and (b) analyzing the amplification products to determine sequence.

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

Analysis of the reaction products for elucidation of sequence information can be carried out using any of various methods known in the art. Such methods include gel electrophoresis and detection of the labeled bands using appropriate scanner, sequencing gel electrophoresis and detection of the radiolabeled band directly by phosphorescence such as Molecular Dynamics reader, capillary electrophoresis adapted with a detector

IDT Exhibit 1020.8776

specific for the labels used in the reaction, and the like. The label can also be a ligand for a binding protein which is used for detection of the label in combination with an enzyme conjugated to the binding protein, such as biotin-labeled chain terminator and streptavidin conjugated to an enzyme. The label is detected by the enzymatic activity of the enzyme, which generates a detectable signal. As with other sequencing methods known in the art, the sequencing reactions for the various nucleotide types (A, C, G, T or U) are carried out either in a single reaction vessel, or in separate reaction vessels (each representing 1 of the various nucleotide types). The choice of method to be used is dependent on practical considerations readily apparent to one skilled in the art, such as the nucleotide tri phosphate analogs and/or label used. Thus, for example, when each of the analogs is differentially labeled, the sequencing reaction can be carried out in a single vessel. The considerations for choice of reagent and reaction conditions for optimal performance of sequencing analysis according to the methods of the invention are similar to those for other previously described sequencing methods. The reagent and reaction conditions should be as described above for the nucleic acid amplification methods of the invention.

Detection of mutation based on single stranded conformation polymorphism utilizing the amplification methods of the invention

The RNA amplification products generated according to the methods of the invention are also suitable for analysis for the detection of any alteration 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 RNA products of the amplification methods are suitable for single stranded conformation polymorphism (rSSCP) based mutation detection. The amplification methods of the invention can be directly linked to appropriate means for detecting single stranded conformation polymorphism, such as an electrophoretic separation method for the identification of specific mobility pattern of the single stranded RNA products for the elucidation of the presence of specific sequence feature(s), and/or the presence of any difference in a test nucleic acid as compared to a reference nucleic acid.

Methods based on gel electrophoresis or capillary electrophoresis can be used for the detection and analysis of the various single stranded conformational isomers. Alternatively, it is also likely that cleavage of the single stranded RNA product using nucleases which recognize sequence dependent secondary structures may be useful for the

IDT Exhibit 1020.8777

determination of sequence specific conformation polymorphism. Such secondary structure-specific nucleases are known in the art, such as the 5'-nucleases known as Cleavase™ enzymes (Third Wave). The electrophoretic methods are potentially more suitable for high throughput mutation, or genotyping, detection methods.

The determination of sequence specific electrophoretic pattern for a given nucleic acid sequence is useful for, for example, the detection of specific alleles of a test sequence. Furthermore, it is expected that an electrophoretic mobility pattern for the various alleles could be well differentiated, thus allowing the detection of two alleles in a nucleic acid sample from a single individual, as required for heterozygous genotype, or multiple alleles. Any alteration in the test nucleic acid sequence, such as base substitution, insertions or deletion, could be detected using this method. The method is expected to be useful for detection of specific single base polymorphism, SNP, and the discovery of new SNPs. Thus, the invention also provides methods for detecting a polynucleotide comprising a single nucleotide polymorphism, comprising: (a) amplifying a target polynucleotide using any of the methods described herein; and (b) analyzing the amplification products for single stranded conformation, wherein a difference in conformation as compared to a reference single stranded polynucleotide indicates a single nucleotide polymorphism in the target polynucleotide, whereby a polynucleotide comprising a single nucleotide polymorphism is detected.

Mutation detection utilizing the composite primer-based amplification methods of the invention

The unique properties of the composite primer for use in the isothermal amplification methods of the invention provide the basis for an isothermal method for the detection of defined mutations, or polymorphic sites (such as SNPs), in a target nucleic acid sequence. The method is useful for genotyping, detection of mutation leading to drug resistance and the like. These methods are applicable to characterizing sequences in a region in the template strand which generally hybridize to the RNA portion of the composite primer – hence reference to "defined" mutations, which are defined in terms of their location.

In one embodiment, the RNA portion(s) of the composite primer is designed to be complementary to the sequence of the test target nucleic acid in which the presence of a sequence alteration is suspected. Stated alternatively, the primer comprises an RNA

IDT Exhibit 1020.8778

portion(s) that comprises a sequence that is complementary to the reference sequence (for example, a wild type sequence) against which the sequence in the test target nucleic acid is to be compared. In some embodiments, the altered sequence (i.e., the sequence comprising a sequence alteration) and the reference sequence are alleles. The sequence alteration may be a single nucleotide substitution, a deletion or insertion.

In another embodiment, the RNA portion(s) of the composite primer is designed to be complementary to the altered sequence suspected to be present in the test target nucleic acid. Stated alternatively, the primer comprises an RNA portion(s) that comprises a sequence that is complementary to the test target nucleic, and thus is not complementary to the reference sequence (for example, a wild type sequence) against which the sequence in the test target nucleic acid is to be compared. In some embodiments, the altered sequence (i.e., the sequence comprising a sequence alteration) and the reference sequence are alleles.

The RNA portion, generally 5' RNA portion, of the composite primer comprises a sequence which is complementary to a known normal wild type sequence, or a known mutant or a polymorphic genotype. Generally, a suitable composite primer comprises an RNA portion that allows the primer to preferentially hybridize to a target nucleic acid if the target nucleic sequence comprises a sequence complementary to the RNA portion of the primer compared to if there is a mismatch (i.e., the primer has the mutated sequence and the target does not, or vice versa), wherein the target nucleic acid has a bound primer extension product and has had its 5'-RNA portion cleaved. The presence of sequence alteration does not generally prevent the initial step of the amplification. The composite primer hybridizes to the target sequence, to form a tri molecular complex, and is extended. A ribonuclease, such as RNaseH, then cleaves the RNA portion of the extended primer of the complex. While it is likely that the presence of a mismatched base pair will affect the pattern of cleavage of the RNA/DNA hybrid, the cleavage is nonetheless likely to take place. The next step of binding of a composite primer to the complex by hybridization of the 5' RNA portion will be inhibited, preferably prevented, by a mismatch. This effect is dependent on factors such as the size of the hybridizing oligonucleotide and the stringency of the reaction condition. These factors are considered in the design of the composite primer, according to techniques well known and routine in the art. It is also possible that the mismatch will inhibit cleavage of the RNA portion(s) of the composite primer, thus preventing the amplification of the target sequence. Another possibility is that the mismatch will result in lower efficiency of cleavage of the RNA portion of the primer thus

IDT Exhibit 1020.8779

resulting in lower efficiency of amplification or production of less amplification product. The inability of the composite primer to hybridize to the target at this step of the amplification prevents further steps of primer extension strand displacement and production of multiple copies of the amplification products. It is understood that the detection of mutation by the methods of the present invention can be based on absence or presence of an amplification product, or quantitative comparisons of amount of accumulated amplification products. For example, when the composite primer comprises the reference sequence (for example, wild type), the presence of a mutation in a target strand may lead to no detectable amplification products; alternatively, it may lead to detectable products, but less than those produced from a template strand without the mutation.

When the composite primer comprises an RNA portion, generally a 5' RNA portion, that is fully complementary to a mutant genotype, amplification of a sequence which is of the normal genotype will be prevented, while a mutant genotype target will be amplified. Thus, in this case the detection and/or quantitative determination of multiple copies of the amplification product will be indicative of the presence of a target sequence of the mutant genotype. For example, parallel reactions that include either the nucleic acid sample of interest or reference sample of target nucleic with a wild type sequence could be run. Accumulation of more amplification products in the former compared to the latter reaction would be indicative of the presence of a mutant genotype in the sample of interest. Alternatively, when the composite primer comprises a 5' RNA sequence that is fully complementary to a normal genotype sequence of the test target, amplification of a target sequence of the mutant genotype is prevented, and the detection and/or quantitative determination of amplification products is indicative of a normal genotype.

Any of the composite primer-based amplification methods of the present invention are suitable for detection of mutation as described above.

Method of preparing microarrays of nucleic acids

The single stranded nature of the products of the amplification methods of the invention are particularly suitable for preparing microarrays comprising the amplification products.

Amplification products can be attached to a solid or semi-solid support or surface, which may be made, e.g., from glass, plastic (e.g., polypropylene, nylon), polyacrylamide, nitrocellulose, or other materials.

IDT Exhibit 1020.8780

Several techniques are well-known in the art for attaching nucleic acids to a solid substrate such as a glass slide. One method is to incorporate modified bases or analogs that contain a moiety that is capable of attachment to a solid substrate, such as an amine group, a derivative of an amine group or another group with a positive charge, into the amplified nucleic acids. The amplified product is then contacted with a solid substrate, such as a glass slide, which is coated with an aldehyde or another reactive group which will form a covalent link with the reactive group that is on the amplified product and become covalently attached to the glass slide. Microarrays comprising the amplified products can be fabricated using a Biodot (BioDot, Inc. Irvine, CA) spotting apparatus and aldehyde-coated glass slides (CEL Associates, Houston, TX). Amplification products can be spotted onto the aldehyde-coated slides, and processed according to published procedures (Schena et al., *Proc. Natl. Acad. Sci. U.S.A.* (1996), 93: 10614-10619). Arrays can also be printed by robotics onto glass, nylon (Ramsay, G., *Nature Biotechnol.* (1998), 16:40-44), polypropylene (Matson, et al., *Anal Biochem.* (1995), 224(1):110-6), and silicone slides (Marshall, A. and Hodgson, J., *Nature Biotechnol.* (1998), 16:27-31). Other approaches to array assembly include fine micropipetting within electric fields (Marshall and Hodgson, *supra*), and spotting the polynucleotides directly onto positively coated plates. Methods such as those using amino propyl silicon surface chemistry are also known in the art, as disclosed at http://www.cmt.corning.com and http://cmgm.stanford.edu/pbrown/.

One method for making microarrays is by making high-density polynucleotide arrays. Techniques are known for rapid deposition of polynucleotides (Blanchard et al., *Biosensors & Bioelectronics*, 11:687-690). Other methods for making microarrays, e.g., by masking (Maskos and Southern, *Nuc. Acids. Res.* (1992),20:1679-1684), may also be used. In principle, and as noted above, any type of array, for example, dot blots on a nylon hybridization membrane, could be used. However, as will be recognized by those skilled in the art, very small arrays will frequently be preferred because hybridization volumes will be smaller.

The amplified polynucleotides may be spotted as a matrix on substrates comprising paper, glass, plastic, polypropylene, nylon, polyacrylamide, nitrocellulose, silicon, optical fiber or any other suitable solid or semi-solid (e.g., thin layer of polyacrylamide gel ((Khrapko, et al., *DNA Sequence* (1991), 1:375-388)) surface.

An array may be assembled as a two-dimensional matrix on a planar substrate or may have a three-dimensional configuration comprising pins, rods, fibers, tapes, threads,

IDT Exhibit 1020.8781

beads, particles, microtiter wells, capillaries, cylinders and any other arrangement suitable for hybridization and detection of target molecules. In one embodiment the substrate to which the amplification products are attached is magnetic beads or particles. In another embodiment, the solid substrate comprises an optical fiber. In yet another embodiment, the amplification products are dispersed in fluid phase within a capillary which, in turn, is immobilized with respect to a solid phase.

Characterization of nucleic acids

The amplification products obtained by the methods of the invention are particularly amenable to further characterization, in part because the products are single stranded. The amplified products can be analyzed using, for example, probe hybridization techniques known in the art, such as Northern blotting, and hybridizing to probe arrays. They can also be analyzed by electrophoresis-based methods, such as differential display and size characterization, which are known in the art.

Amplification products can be used for determining the sequence of a sequence of interest. For example, an amplification product comprising a sequence of interest generated by amplification methods of the invention can be subjected to sequencing by any suitable sequencing method. Suitable sequencing methods are known in the art, and include, for example, using nucleotide triphosphates that upon incorporation into a primer extension product terminates nucleotide polymerization.

Amplification products can also be used for detecting presence of and/or quantifying a nucleic acid sequence of interest in a sample. For example, presence of a nucleic acid sequence of interest in a sample can be detected by detecting the sequence of interest in amplification product resulting from amplifying polynucleotides in a sample suspected of comprising the sequence of interest. In some embodiments, a sequence of interest comprises a mutation, for example, a single nucleotide polymorphism, an insertion, a deletion or a substitution. A sequence of interest in an amplification product can be detected by any of a variety of methods known in the art, including, for example, hybridizing amplification product comprising (or suspected of comprising) the sequence of interest with a nucleic acid probe that is hybridizable to the sequence of interest. Suitable nucleic acid probes would be evident to one skilled in the art, and include, for example, probes that comprise DNA, RNA or DNA and RNA. These probes can be provided in any suitable form, including, for example, as microarrays, which may comprise the probe

IDT Exhibit 1020.8782

immobilized on a suitable substrate that can be fabricated from a material such as paper, glass, plastic, polypropylene, nylon, polyacrylamide, nitrocellulose, silicon and optical fiber. Detection of sequence of interest in an amplification product can also be achieved by methods such as limited primer extension, which are known in the art and described in, for example, U.S. Patent Nos. 5,888,819; 6,004,744; 5,882,867; 5,710,028; 6,027,889; 6,004,745; 5,763,178; 5,011,769; 5,185,243; 4,876,187; 5,882,867; WO US88/02746; WO 99/55912; WO92/15712; WO 00/09745; WO 97/32040; WO 00/56925, and in co-pending U.S. Application Ser. No. 60/255,638, filed 13 December, 2000.

In one embodiment, the amplification methods of the invention are utilized to generate multiple copies of single stranded RNA products that are labeled by the incorporation of labeled nucleotides during RNA polymerization. For example, amplification according to the methods of the invention can be carried out with suitable labeled rNTPs. These labeled rNTPs can be directly attached to a label, or can comprise a moiety which could be attached to a label. The label may be attached covalently or non-covalently to the amplification products. Suitable labels are known in the art, and include, for example, a ligand which is a member of a specific binding pair which can be detected/quantified using a detectable second member of the binding pair. Thus, amplification of a target polynucleotide according to the methods of the invention in the presence of, for example, Cye3-dATP or Cye5-dATP results in the incorporation of these nucleotides into the amplification products.

The labeled amplified products are particularly suitable for analysis (for example, detection and/or quantification) by contacting them with, for example, microarrays (of any suitable surface, which includes glass, chips, plastic), beads, or particles, that comprise suitable probes such as cDNA and/or oligonucleotide probes. Thus, the invention provides methods to characterize (for example, detect and/or quantify) a sequence of interest by generating labeled RNA products using amplification methods of the invention, and analyzing the labeled products. Analysis of labeled products can be performed by, for example, hybridization of the labeled amplification products to, for example, probes immobilized at, for example, specific locations on a solid or semi-solid substrate, probes immobilized on defined particles, or probes immobilized on blots (such as a membrane), for example arrays, which have been described above. Other methods of analyzing labeled products are known in the art, such as, for example, by contacting them with a solution comprising probes, followed by extraction of complexes comprising the labeled

IDT Exhibit 1020.8783

amplification products and probes from solution. The identity of the probes provides characterization of the sequence identity of the amplified products, and thus by extrapolation the identity of the target polynucleotide present in a sample. Hybridization of the labeled products is detectable, and the amount of specific labels that are detected is proportional to the amount of the labeled amplification products of a specific sequence of interest. This measurement is useful for, for example, measuring the relative amounts of the various polynucleotide species in a sample. The amount of labeled products (as indicated by, for example, detectable signal associated with the label) hybridized at defined locations on an array can be indicative of the detection and/or quantification of the corresponding target polynucleotide species in the sample.

Determination of gene expression profile

The amplification methods of the invention are particularly suitable for use in determining the levels of expression of multiple genes in a sample since the methods described herein are capable of amplifying multiple target polynucleotides (RNA as a direct target or cDNA prepared from an RNA sample), 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 sequences of interest present in a sample, as determined by quantifying amplification products of the sequences, can provide for determination of the gene expression profile of the sample source.

The following Examples are provided to illustrate, but not limit, the invention.

EXAMPLES

Example 1: The first transcription module of composite primer-based methods

The general methods that are described in this example are also utilized in other examples of the first transcription module of composite primer-based methods provided herein.

IDT Exhibit 1020.8784

Buffers

Buffers that were used throughout the composite primer-based examples are made with the following materials.

TE Buffer: 10 mM Tris-HCl, pH 8.0, 1 mM EDTA

TBE Buffer: 89 mM Tris base, 89 mM boric acid, 2 mM EDTA, pH 8.3

FX Buffer: 20 mM Tris-SO_4, pH 9.0, 20 mM $(NH_4)_2SO_4$, 0.1% NP-40

A first illustration of the first transcription module of amplification methods using a single composite primer, DNA polymerase and Rnase H

Sequence amplification was performed in 15 µl reactions containing 20 mM Tris-HCl, pH 8.5, 6.0 mM $MgCl_2$, 1.0 mM dATP, 1.0 mM dCTP, 1.0 mM dTTP, 0.8 mM dGTP, 0.2 mM dITP (dNTP's from Amersham), 0-6% DMSO, 0-8% glycerol, 0-100 ug/ml acetylated BSA (Ambion, Austin, TX), 0.6 Units/µl recombinant ribonuclease inhibitor (rRNasin, Promega, Madison, WI), 0.5-5 uM composite primer IA005, and 100-200 nM promoter-template oligonucleotide (PTO) IA015C. Composite primer IA005 is a 20-mer primer with the sequence of: ACGGAUGCGGUCUCCAGTGT (SEQ ID NO:1). Promoter-template oligonucleotide (PTO) IA015C is a 55-mer oligonucleotide with the sequence of:

ggAATTCTAATACgACTCACTATAgggAgAgCggTACgCTgATCAAAgATCCgTg-biotin (SEQ ID NO:12).

Reactions were assembled with all components except the two enzymes. After heating to 70°C or 99°C for 10 sec. in a programmable thermal cycler (GeneAmp 9600, Perkin Elmer), the reaction mixtures were cooled down to 55°C, 60°C or 65°C, as described in the individual examples. Upon attaining the lower temperature, 0.05-0.24 Unit of RNase H (diluted from the 5 U/µl stock solution using a diluent/storage solution: 10 mM Tris-HCl, pH 8.5, 30% glycerol; Hybridase thermostable RNase H, Epicentre Technologies, Madison, WI) and 1.0-5.0 Units Bca DNA polymerase (2 U/µl; Panvera, Madison, WI) were added. The reactions were incubated at 55°C-65°C for 30 minutes At the end of the incubation, reactions were cooled to 4°C until RNA transcription step was desired.

RNA transcription was performed at 37°C for 3 hours in 10 ul reactions containing 2.5 ul of the linear amplification reaction mixtures above, and 40 mM Tris-HCl, pH 8.5, 70 mM KCl, 5.0 mM DTT, 12 mM $MgCl_2$, 110 µg/ml BSA, 3 mM each rNTP (ATP, UTP,

IDT Exhibit 1020.8785

CTP, GTP, Amersham), 7.5% DMSO, 1 Unit/μl rRNasin (Promega, Madison, WI), and 20 Units T7 RNA polymerase (Ambion, Austin, TX).

DNA Templates

A sequence from the J-gene region of *E. coli* K12 was chosen as a DNA template for several of the following Examples. Three DNA templates were used for these experiments: a synthetic DNA target (IA013), a primarily single-stranded DNA (351 bases) template produced by PCR amplification, and genomic DNA from the K12 strain of *E. coli* (preparation described in Example 4). Synthetic DNA target IA013 comprises:

Spacer18-Spacer18CGGTACGCTGATCAAAGATCCGTGCAACAAATGTCA TGGTCATGGTCGTGTTGAGCGCAGCAAAACGCTGTCCGTTAAAATCCCGGCAG GGGTGGACACTGGAGACCGCATCCGT. Spacer18 refers to polyoxyethylene spacers. These were added to the oligo in order to retard its mobility with respect to the 100-bp ssDNA product. The sequence of the aforementioned primarily single-stranded DNA (351 bases) template produced by PCR amplification is:

CGGTACGCTGATCAAAGATCCGTGCAACAAATGTCATGGTCATGGTCGTGTT GAGCGCAGCAAAACGCTGTCCGTTAAAATCCCGGCAGGGGTGG**ACACT*GGAG ACCGCATCCGT***CTTGCGGGCGAAGGTGAAGCGGGCGAGCATGGCGCACCGGCA GGCGATCTGTACGTTCAGGTTCAGGTTAAACAGCACCCGATTTTCGAGCGTGAA GGCAACAACCTGTATTGCGAAGTCCCGATCAACTTCGCTATGGCGGCGCTGGG TGGCGAAATCGAAGTACCGACCCTTGATGGTCGCGTCAAACTGAAAGTGCCTG GCGAAACCC**AGACCGGTAAGCTATTCCGTATGCG** wherein the PCR primers are bolded and underlined and the composite primers are bolded, with RNA portion in italics.

Preparation of ssDNA target from PCR amplification product

Single-stranded DNA template for amplification was prepared by PCR amplification of a 351-bp segment of the *E. coli* J gene using the primers IA006 and IA004. Primer IA006 is a 23-mer with the sequence of: CGGTACGCTGATCAAAgATCCGT. Primer IA004 is a 26-mer with the sequence of: CGCATACGGAATAGCTTACCGGTCT.

PCR was performed in 100 μl reactions containing 20 mM Tris-SO_4, pH 9.0, 20 mM $(NH_4)_2SO_4$, 0.1% NP-40, 2.0 mM $MgCl_2$, 300 μM each dNTP (dATP, dTTP, dCTP, dGTP), 5 Units Taq DNA polymerase, 400 nM primer IA006, 400 nM 5'-phosphate-

IDT Exhibit 1020.8786

primer IA004, and 0.2 ul of a crude lysate of *E. coli* K12 strain. A modified "touchdown PCR" protocol was used with the following parameters: 95°C for 5 seconds, 68°C for 1 minute for 5 cycles; 94°C for 5 seconds, 60°C for 30 seconds, 72°C for 1 minute for 5 cycles; 94°C for 5 seconds, 55°C for 30 seconds, 72°C for 1 minute for 40 cycles; 72°C for 15 minutes and then held indefinitely at 4°C. Primer IA004 was synthesized with a 5'-phosphate to protect the sense-strand from digestion by lambda exonuclease (Strandase kit, Novagen, Madison, WI). The Strandase digestion was performed according to the manufacturer's recommendation. Briefly, PCR product prepared as described above, was precipitated from the reaction mixture by the addition of 1/10 volume 3M sodium acetate, pH 5.2 and 0.6 volumes isopropanol, cooling to -20°C for 1 hour, and centrifuged at maximum speed in a microcentrifuge for 30 minutes. The DNA pellet was washed once with 75% ethanol, then air-dried briefly before resuspension in 80 µl water. Concentration was estimated from O.D. at 260 nm, and 60 Units of lambda exonuclease (Strandase, Novagen) was added. Digestion was allowed to proceed at 37°C for 20 minutes, reactions were then heated to 75°C for 10 minutes, and cooled to 4°C. Incubations were performed in a programmable thermal cycler (GeneAmp 9600, Perkin Elmer). Remaining DNA was purified using QiaQuick Nucleotide Removal Columns (Qiagen, Valencia, CA) following the manufacturer's recommended procedure and using the buffers provided with the kit (Qiagen, Valencia CA). Briefly, 10 volumes of Buffer PN (Qiagen) were added to the sample. The entire volume was then applied to a Qiagen spin column and centrifuged (6000 rpm 1 minute in a microcentrifuge). The filtrate was discarded, and the column was washed twice with 500 ul of Buffer PE (Qiagen). The column was then dried thoroughly by centrifugation at maximum speed for 3 minutes. The DNA was eluted in 50 µl Buffer EB (10 mM Tris-HCl, pH 8.5)(Qiagen). The concentration was estimated to be about 2.5 X 10^{12} copies/ 5µl from OD at 260 nm. Gel analysis revealed that significant double stranded DNA (less than half the total) remained, but the error in concentration was less than 2-fold. The DNA was diluted to 10^{10} copies/ 5µl in TE Buffer. Serial dilutions were prepared from the 10^{10} copy stock solution as needed. Concentration based on O.D. measurement was confirmed by limiting dilution PCR analysis.

Gel Electrophoresis

IDT Exhibit 1020.8787

Amplification products were electrophoretically separated on Novex pre-cast 4-20% polyacrylamide gradient gels (Invitrogen, Carlsbad, CA; part no. EC62255) in 1X TBE Buffer (89 mM Tris base, 89 mM boric acid, 2 mM EDTA, pH 8.3) in a Novex electrophoresis apparatus (EI9001-XCell II Mini Cell, Novex). Reaction mixtures (5 µl) were mixed with 1 µl of 6X Gel Loading Solution (40% sucrose, 0.25% bromophenol blue, 0.25% xylene cyanole), and the entire sample was immediately loaded into each well. Gels were subjected to 250V for approximately 5 minutes, until all samples had entered the gel, and the voltage was lowered to 175V for 45 minutes. Gels were removed from between the plastic plates and stained in 0.5 ug/ml ethidium bromide in 1X TBE Buffer (89 mM Tris base, 89 mM boric acid, 2 mM EDTA, pH 8.3). A double stranded DNA molecular size marker (Hi-Lo DNA Marker, Bionexus, San Leandro, CA) was included in one lane of each gel run. This marker contains 16 fragments of the following sizes: 50, 100, 200, 300, 400, 500, 750, 1000, 1400, 1550, 2000, 3000, 4000, 6000, 8000, and 10000 bp. Typically, 50-2000 bp could be resolved on the gels used.

Hybridization

Oligonucleotide probes for hybridization examples(IA010 for ssDNA products; IA014 for ssRNA products) were 5'-end-labelled using T4 polynucleotide kinase (New England Biolabs, Beverly, MA) and γ-^{32}P-ATP (adenosine 5'-[γ-^{32}P] triphosphate, triethylammonium salt, Amersham, Piscataway, NJ; PB10218, >5000 Ci/mmol, 10 mCi/ml). Primer IA010 is a 21-mer with the sequence of: ATGTCATGGTCATGGTCGTGT. Primer IA014 is a 31-mer with the sequence of: CTCAACACGACCATGACCATGACATTTGTTG. Labelling reactions (50 µl total volume) contained 70 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 5 mM DTT, 1 µg oligo (147 pmol for primer IA010; 101 pmol for primer IA014), 250 µCi γ-^{32}P-ATP, and 30 Units T4 polynucleotide kinase. Incubation was at 37°C for 30 minutes, followed by removal of unincorporated nucleotide using QIAquick Nucleotide Removal Kit (Qiagen, Valencia, CA). The decay rate (cpm) was determined in a Packard Minaxi Tri-Carb 4000 Series liquid scintillation counter by Cherenkov counting of 1 µl of the labelled oligo.

Hybridization was performed in 30 µl reactions. Product DNA (or RNA) (10 µl) was added to 20 µl of probe mix. Reactions contained 100 mM NaCl and 10^6 cpm of probe (correcting for decay using a half-life of 14.3 days). After heating to 65°C 15 seconds,

IDT Exhibit 1020.8788

hybridization was allowed to proceed at 42°C for 30 minutes, followed by cooling to 4°C. These steps were performed in a programmable thermal cycler with a heated cover (GeneAmp 9600, Perkin Elmer). The entire volume of hybridization reaction was electrophoresed in 10% polyacrylamide gels in 1X TBE Buffer (89 mM Tris base, 89 mM boric acid, 2 mM EDTA, pH 8.3) at 150V for 3 hours. Gels were removed from the glass plates, wrapped in plastic wrap, and exposed to autoradiography film (BioMax MR, Kodak) at -20°C overnight (~16 hours) with two intensifying screens.

Example 2: A second illustration of the first transcription module of a composite primer-based amplification method

Amplification using a single composite primer, DNA polymerase, RNase H, and TSO or blocker was performed. Reaction mixtures containing all reaction components, as described above, as well as reaction mixtures without one of the key reagents such as composite primer, RNase H, or Bca DNA polymerase (Panvera, Madison, WI) were spiked with 10^{10} copies of synthetic ssDNA target (IA010 - sequence listed in Example 1). A negative control reaction containing all the reagents and no target ssDNA, was also included. The amplification of target DNA sequence was carried out as described above. Target denaturation was carried out at 70°C and the isothermal amplification was carried out at 65°C.

The amplification products were resolved by gel electrophoresis. No amplification products were detected in reaction mixtures without primer, RNase H or Bca DNA polymerase.

Probe IA010 hybridization and autoradiography to the ssDNA product of the linear amplification method, verified the identity of the amplification product. The linear amplification of the synthetic oligonucleotide target in this experiment was done using a non blocked promoter-template oligonucleotide (IA015b). Promoter-template oligonucleotide IA015b is a 55-mer with a sequence of: GGAATTCTAATACGACTCACTATAGGGAGAGCGGTACGCTGATCAAAGATCCG TG. The standard reaction components used for this amplification reaction are as given above. The initial denaturation step was performed at 70°C for 10 seconds. The reactions were cooled down to 65°C, and further incubated at this temperature for 30 minutes

IDT Exhibit 1020.8789

following the addition of Bca polymerase and RNase H. No hybridization was detected in the control reactions (no DNA, no primer, no RNase H, no Bca).

Example 3: Promoter template oligonucleotide-based first transcription module of a composite primer-based amplification method

The promoter-template oligonucleotide (PTO) contains two essential sequence motifs: a T7 promoter sequence (5'-TAATACGACTCACTATAGGGAgGAG)) and a sequence complementary to the ssDNA template. Four versions of a PTO were designed (IA012, IA012b, IA015, IA015b). IA012 PTO is a 67-mer and has a sequence of: GGAATTCTAATACGACTCACTATAGGGAGAGATCGAGTAGCTCCGGTACGCTG ATCAAAGATCCGTG. IA012 PTO contains two sequences in addition to the core T7 promoter: a 5'-extension (5'-GGAATTC) and a spacer (5'-ATCGAGTAGCTC) between the promoter and the target DNA-complementary sequence. IA015 is the shorter PTO (48-mer), lacking both the 5'-extension and the spacer. IA015 PTO has the sequence of: TAATACGACTCACTATAGGGAGAGCGGTACGCTGATCAAAGATCCGTG. IA012b PTO is a 60-mer which contains the spacer, but not the extension. IA012b PTO has the sequence: TAATACGACTCACTATAGGGAGAGATCGAGTAGCTCCGGTACGCTGATCAAAG ATCCGTG. IA015b contains the extension, but not the spacer. The sequence of IA015b is disclosed in Example 2. All primers other than the chimeric oligonucleotides IA005, IA019, and IA020 were synthesized by Keystone (Division of BioSource International, Camarillo, CA) and were PAGE purified.

The ability of IA012, IA012b, IA015, and IA015b to convert the ssDNA template into a substrate for T7 RNA polymerase was assessed by comparing the amount of RNA produced after transcription of overlap-extension products formed between a synthetic oligo product (IA009) and each of the PTO's. Synthetic oligo product IA009 is a 100-mer with the sequence of: AGTGTCCACCCCTGCCGGGATTTTAACGGACAGCG TTTTGCTGCGCTCAACACGACCATGACCATGACATTTGTTGCACGGATCTTTGA TCAGCGTACCG. Overlap-extension was performed in 15 ul reactions containing 20 mM Tris-HCl, pH 8.5, 6 mM $MgCl_2$, 1 mM each dNTP (dATP, dTTP, dCTP, dGTP), 100 nM IA009, 100 nM PTO, and 1 Unit Bca DNA polymerase. Reactions were constituted without Bca DNA polymerase, heated to 95°C then cooled over 10 minutes to 60°C. After

IDT Exhibit 1020.8790

addition of DNA polymerase, reactions were incubated at 60°C for 30 minutes. A portion (2.5 µl) of the reaction mixture was added to the standard RNA transcription reaction mixture and the transcription reactions were assessed by gel electrophoresis.

Significantly more RNA was produced by the transcription substrate produced with the shorter PTO's (IA015, IA015b) than by either of IA012 or IA012b. The PTO containing the 5'-extension but not the spacer (IA015b) produced demonstrably higher yields of RNA. In all cases, however, multiple products appeared in addition to the major RNA band. A fifth PTO was designed having the same sequence as IA015b, but with a 3'-blocking group (biotin) to eliminated the free 3'-OH, to demonstrate the improved performance of a 3'-blocked PTO. Blocking the free 3'-OH of the PTO eliminates its ability to initiate non-specific incorporation of a functional promoter sequence to the amplifiction products leading to non-specific generation of transcription products. The performance of 3-blocked and unblocked PTO in the enhanced isothermal linear amplification was assessed from amplification of a synthetic oligonucleotide target, using the standard conditions. The 3'-blocked PTO (IA015c) produced comparable yields of specific RNA as IA015b, but with significantly less background. Negative control reactions (no DNA template) and reactions containing 10^{10} copies of oligo target (IA013) were amplified by strand displacement for 30 minutes, 1 hour, or 2 hours at 55C, with either IA015b or IA015c included in the strand-displacement reaction. When the 3'-OH was not blocked, non-specific RNA was produced in all reactions and obscured identification of the specific RNA band. In contrast, the blocked PTO produced primarily a single RNA product.

Example 4: Amplification of a J gene sequence of *E. coli* genomic DNA by the first transcription module of a composite primer-based amplification method

DNA was isolated from 25 ml of *E. coli* K12 (ATCC 10798) grown overnight in Tryptone-NaCl medium. Genomic DNA was isolated by lysozyme digestion and solubilization in a chaotropic lysis solution (Bactozol Kit, Molecular Research Center, Cincinnati, OH) following the manufacturer's recommended procedure. Briefly, bacteria were collected by centrifugation at 6000 x g for 5 minutes. Cell pellets were resuspended in Bactozyme digestion buffer and incubated at 50°C for 30 minutes. The resulting lysate was clear at the end of the digestion, without any visible clumps of undigested cells. The

IDT Exhibit 1020.8791

lysate was mixed with 4 volumes of DNazol reagent (Molecular Research Center, Cincinnati, OH) and incubated for 15 minutes at room temperature. DNA was precipitated from the solution by addition of 0.6 volume ice-cold ethanol. After incubation for 5 minutes at room temperature, the precipitated DNA was collected by centrifugation for 5 minutes at maximum speed in a micro-centrifuge. The DNA pellet was washed with 75% ethanol, centrifuged again, and resuspended in 1.5 ml TE Buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) by heating at 50°-55°C for 30 minutes, with frequent agitation. The resulting solution was passed repeatedly through a 22 gauge syringe needle to shear the DNA and reduce the viscosity of the solution. DNA was precipitated again (EPI005-35) by the addition of 1/10 volume of 5M ammonium acetate and 3 volumes ice-cold ethanol. After incubation at -20°C for 1 hr, the DNA was collected by centrifugation at maximum speed in a micro-centrifuge. The pellet was washed with 75% ethanol, centrifuged again, and resuspended in 150 µl TE Buffer. Two dilutions in TE Buffer were prepared for O.D. measurement (Beckman DU640 spectrophotometer), from which DNA concentration was calculated by assuming 50 µg/ml dsDNA produces an O.D. at 260 nm of 1. DNA concentrations of the two dilutions were 24.2 ug/10 µl and 24.6 ug/10 µl. The average of these two measurements (24.4 ug/ 10 µl) corresponds to approximately 2.5 X 10^9 genome copies / 5 µl (5 fg of *E. coli* genomic DNA = 1 copy).

DNA amplification

DNA was serially diluted in TE Buffer to 10^9, 10^8, or 10^7 copies/ 5 ul, and denatured by heating to 95°C for 5 minutes followed by rapid cooling on ice. Single-stranded template DNA also was diluted to 10^9 copies/ 5ul. Reactions were assembled to contain no DNA, 10^7, 10^8, 10^9, or 2.5 X 10^9 copies of genomic DNA.

Amplification was performed in 15 µl reactions containing 20 mM Tris-HCl, pH 8.5, 6.0 mM $MgCl_2$, 1.0 mM dATP, 1.0 mM dCTP, 1.0 mM dTTP, 0.8 mM dGTP, 0.2 mM dITP (dNTP's from Amersham), 6% DMSO, 8% glycerol, 100 µg/ml acetylated BSA (Ambion, Austin, TX), 0.6 Units/µl recombinant ribonuclease inhibitor (rRNasin, Promega, Madison, WI), 5 uM composite primer IA005 (sequence disclosed in Example 1), 200 nM promoter-template oligonucleotide (PTO) IA015C (sequence disclosed in Example 1). Reactions were assembled with all components except the two enzymes. After heating to 99°C for 10 seconds in a programmable thermal cycler (GeneAmp 9600, Perkin Elmer),

IDT Exhibit 1020.8792

reactions were incubated at 60°C for 30 minutes Upon attaining 60°C, 0.6 µl of RNase H (0.05 Units diluted from the 5 U/µl stock solution in 10 mM Tris-HCl, pH 8.5, 30% glycerol) (Hybridase, Epicentre Technologies, Madison, WI) and 1.0 µl Bca DNA polymerase (2.0 Units, Panvera, Madison, WI) were added. At the end of the 60°C incubation, reactions were cooled to 4°C. A volume of 5.0 µl of strand-displacement product was added to each RNA transcription reaction (total volume 20 µl). RNA transcription was performed using the standard conditions and scaling up the reaction volume to 20 µl to provide sufficient material for direct gel analysis (5 µl) and probe hybridization (10 µl).

Unlike the amplification of defined single stranded synthetic target, the amplification of genomic DNA according to the composite primer-based method of the invention generally requires the formation of a defined stop for the formation of a ssDNA product with a defined 3'-end. The formation of a defined stop for primer extension can be achieved by a blocker, which hybridizes to a defined site on the target strand and can not be displaced by the polymerase. Alternatively, as in the present example, a GC rich sequence up stream of the primer site, provided a stop point for primer extension, thus leading to the formation of a ssDNA product with defined 3-end.

Amplification of a defined sequence of genomic DNA by the enhanced isothermal linear amplification method of the invention (first transcription module of a composite primer-based method) was achieved. The ssRNA product was found to hybridize to a specific oligonucleotide probe.

Example 5: Evaluation of the effect of composite primer design on the performance of the first transcription module of a composite primer-based amplification method

The performance of each of the three composite primers in the first transcription module of composite primer-based amplification methods of the invention was assessed. The isothermal linear amplification was performed in 15 µl reactions containing 20 mM Tris-HCl, pH 8.5, 6.0 mM $MgCl_2$, 1.0 mM dATP, 1.0 mM dCTP, 1.0 mM dTTP, 0.8 mM dGTP, 0.2 mM dITP (dNTP's from Amersham), 6% DMSO, 8% glycerol, 100 µg/ml acetylated BSA (Ambion, Austin, TX), 0.6 Units/µl recombinant ribonuclease inhibitor (rRNasin, Promega, Madison, WI), 5 µM composite primer, 200 nM promoter-template oligonucleotide (PTO) IA015C. The sequence of PTO IA015C is disclosed in Example 1.

IDT Exhibit 1020.8793

The sequence of composite primers IA005 (20-mer) is disclosed in Example 1. Other composite primer sequences with alphanumerical names are as follows:

IA019 (20-mer) ACGGAUGCGGUCUCCAGTGT

IA020 (21-mer) GACGGAUGCGGUCUCCAGTGT

Four other composite primer sequence were used that did not have alphanumerical names. Their sequences are, respectively:

(1) GCAAGACGGAUGCGGUCUCCAGTGT

(2) GACGATGCGUCTCCAGTGT

(3) GACGGATGCGGUCTCCAGUGT

(4) GACGGATGCGGUCTCCAGUGUCCA

These composite primers were synthesized by Dharmacon Research, Inc. (Boulder, CO). The RNA portion of the oligonucleotide was synthesized using a 5'-silyl protecting group in conjunction with an acid-labile 2'-orthoester protecting group (2'-bis(acetoxyethoxy)-methyl ether or "2'-ACE" (Scaringe, S.A., et al. *J. Am. Chem. Soc.* 120:11820-11821 (1998) and Scaringe, S.A. Advanced 5'-silyl-2'-orthoester approach to RNA oligonucleotide synthesis. Methods in Enzymology (in press)). Primers were PAGE purified.

Reactions were assembled with all components except the two enzymes. After heating to 70°C for 10 seconds in a programmable thermal cycler (GeneAmp 9600, Perkin Elmer), reactions were cooled to 55°C-65°C. Upon attaining the lower temperature, 0.05 Unit of RNase H (diluted from the 5 U/μl stock solution using a diluent/storage solution: 10 mM Tris-HCl, pH 8.5, 30% glycerol; Hybridase thermostable RNase H, Epicentre Technologies, Madison, WI) and 2.0 Units Bca DNA polymerase (2 U/μl; Panvera, Madison, WI) were added. The reactions were incubated at 55°C-65°C for 30 minutes. At the end of the incubation, reactions were cooled to 4°C until RNA transcription. RNA transcription was performed at 37°C for 3 hours in 10 μl reactions containing 2.5 μl of linear amplification reaction above, and 40 mM Tris-HCl, pH 8.5, 70 mM KCl, 5.0 mM DTT, 12 mM $MgCl_2$, 110 μg/ml BSA, 3 mM each rNTP (ATP, UTP, CTP, GTP, Amersham), 7.5% DMSO, 1 Unit/μl rRNasin (Promega, Madison, WI), and 20 Units T7 RNA polymerase (Novagen, Madison, WI).

The products of the enhanced linear amplification generated with each of the composite primers were resolved by gel electrophoresis. The composite primers were

IDT Exhibit 1020.8794

designed to hybridize at the same site on the target strand, and differed by the number of deoxynucleotides at the 3'-end. The highest yield of RNA product was produced with primer IA020, followed by IA005 and IA019 equally. The other four composite primers yielded less RNA products. The optimal temperature for the isothermal linear amplification step was different for the different primers, as expected.

Example 6: Genotyping using the first transcription module of a composite primer-based amplification method and genotype specific composite primer

Genomic DNA is isolated from test cells using methods described in previous examples or by other means known to the skilled artisan. Different organisms including, but not limited to, bacteria, viruses, fungi, yeast, plants, and animals are genotyped. Genotype specific primers are designed either to comprise a 3'-end nucleotide which hybridizes to one genotype of a specific nucleic acid sequence, or hybridize to the counterpart genotype. The sequence variation determining specific genotypes may be point mutation, single nucleotide polymorphism (SNP), insertions deletions and the like.

Amplification of the defined target nucleic acid sequence is carried out as described for amplification of genomic E. coli sequence, in the above example. Using the genotype specific primer and DNA polymerase which is devoid of proof reading activity, the generation of amplification product indicates the presence of target sequence of the defined genotype. Sequence variation that prevents hybridization of the 3'-end of the primer to the target nucleic acid sequence will prevent amplification. The amplification product is detected by any one of various methods for detection of single stranded nucleic acid sequence, which is known in the art. For example, the hybridization of specific labeled oligonucleotide probes to the amplification product is detected by gel electrophoresis.

In cases where the genotyping of diploid cells is required, such as the determination of homozygote or heterozygote genotype, it is feasible to carry out the amplification of the specific nucleic acid target sequence using specific primers which are designed for either the wild type and mutant genotype, or one genotype and the other genotype. The amplification reactions using the specific primers are carried out in separate reaction vessels. The detection of amplification product in only one reaction tube or less amplification product is indicative of a homozygote genotype, i.e. either wild type or

IDT Exhibit 1020.8795

mutant homozygote. The detection of amplification product in both amplification reactions indicates a heterozygote genotype.

Example7: Genotyping using a composite primer-based amplification method and genotype specific probe hybridization

Methods for sequencing by hybridization are described in previous examples. Determination of sequence identity by hybridization of specific probe is particularly advantageous using the isothermal method of the invention insofar as the amplification product generated by the method of the invention is single stranded and readily available to be used in hybridization of specific probes. Probes specific for a defined genotype are designed using methods known in the art. It is possible to determine hybridization criteria which will support selective probe hybridization to amplification products generated by amplification of one genotype and not the other. Sequence variation as small as a single nucleotide can prevent probe hybridization. The following factors are taken into consideration for hybridization criteria: probe length, temperature of the hybridization reaction, and buffer composition, in particular divalent ion concentration. The probes used for the ananlysis may be in solution, or may be attached to a solid surface. Further, the probes may be directly labeled or attached to a member of a specific binding pair and thus, able to specifically bind to another member of the specific binding pair which may be directly or indirectly labeled.

Genomic DNA is isolated from test samples by methods known in the art or as described in the above example. Test DNA is combined with the described amplification components, target-specific composite primer, and propromoter sequence (such as PTO). The combination is subjected to incubation conditions as described herein to generate single stranded RNA amplification product. Hybridization of the amplification product to genotype specific probes is carried out in solution or solid phase with attached genotype specific probes. Since the amplification products are single stranded, the products are ideally suited to be attached to a solid phase, such as glass slide, to generate an array of spatially resolved specific probes (i.e., gene chip). Alternatively, the solid phase comprises particles to which specific probes are attached. The detection of probe hybridization to the amplification products is carried out by various methods known in the art, for example, disclosed in Sambrook et al. *supra*. The specific probes is labeled, and the change in label spectral properties due to hybridization is detected and recorded by computer algorithms.

IDT Exhibit 1020.8796

Particle association due to hybridization of specific probes to amplification products is also used for the detection of probe hybridization. Labeled amplification products are generated and product hybridization to probes immobilized on solid surfaces is detected and recorded by computer algorithms. The generation of labeled amplification product is carried out by incorporation of labeled rNTPs during a transcription step by substituting of one of the four rNTPs by a rNTP analog, which is labeled. The label is a dye, or a small molecule such as biotin, which is then detected by binding to specific binding entity, such as labeled streptavidin. Methods for detecting probe hybridization on solid surfaces are known in the art.

Example 8: Genotyping by rSSCP (RNA Single Stranded Conformation Polymorphism) using a composite primer-based amplification method

Genotyping is carried out by amplification of the specific target nucleic acid sequence using a composite primer-based method described herein and by determination of the electrophoretic band pattern of the single stranded RNA product, which reflects the single stranded conformation. The use of SSCP for detection of sequence alteration is widely used. Genotype specific single stranded conformation is determined by subjecting samples to gel or capillary electrophoresis. The generation of single stranded product by amplification of target nucleic acid sequence according to the method of the invention, renders this method particularly suitable for genotype determination by combining the amplification method with rSSCP analysis.

Purified test genomic DNA is combined with components of the amplification method of the invention, as described above, and target specific composite primer and propromoter sequence, such as PTO. The combination is subjected to conditions for isothermal amplification of the target sequence. The reaction mixture containing the amplification product is subjected to either gel electrophoresis or capillary electrophoresis, using instrument and conditions known in the art. The electrophoretic band pattern of the amplification product is determined. The visualization of the oligonucleotide product is achieved by inclusion of a dye intercalator. The electrophoretic pattern of the amplification product is compared to that of amplification products generated by amplification of target nucleic acid sequence obtained from cells of known genotype. Any change in the electrophoretic mobility pattern is indicative of sequence variability. The combination of

IDT Exhibit 1020.8797

the amplification method of the invention and rSSCP provides a simple method for both the discovery of sequence polymorphism of defined target sequences, and detection of previously defined genotypes. The electrophoretic pattern of known nucleic acid sequences, or defined genotypes, can be predetermined, and the pattern generated by products of amplification of test DNA will be compared to known patterns for genotype determination.

Example 9: Amplification of *E. coli* J-gene target sequence using a composite primer-based method linking a first and second transcription module

This example describes the enhanced isothermal linear amplification of a defined sequence of the J-gene of *E. coli*, followed by isothermal exponential amplification of this sequence to yield multiple copies of ssRNA.

(i) In the first example, RNA products of enhanced linear isothermal amplification were generated using three different designs of the single composite primer. The performance of each of the three composite primers in the amplification method of the invention was assessed.

The isothermal linear amplification was performed in 15 µl reactions containing 20 mM Tris-HCl, pH 8.5, 6.0 mM $MgCl_2$, 1.0 mM dATP, 1.0 mM dCTP, 1.0 mM dTTP, 0.8 mM dGTP, 0.2 mM dITP (dNTP's from Amersham), 6% DMSO, 8% glycerol, 100 µg/ml acetylated BSA (Ambion, Austin, TX), 0.6 Units/µl recombinant ribonuclease inhibitor (rRNasin, Promega, Madison, WI), 5 µM composite primer (IA005 or IA019 or IA020), 200 nM promoter-template oligonucleotide (PTO) IA015C. All reactions had 10^9 copies of a 351-nucleotide single stranded DNA template. Reactions were assembled with all components except the two enzymes.

After heating at 70°C for 10sec. in a programmable thermal cycler (GeneAmp 9600, Perkin Elmer), reactions were cooled to 55°C-65°C. The composite primer IA005 was cooled to 60°C, the composite primers IA019 and IA020 were cooled to 55°C, 60°C and 65°C respectively in separate reactions. Upon attaining the lower temperature, 0.05 Unit of RNaseH (diluted from the 5 U/µl stock solution using a diluent/storage solution: 10 mM Tris-HCl, pH 8.5, 30% glycerol; Hybridase thermostable RNaseH, Epicentre Technologies, Madison, WI) and 2.0 Units Bca DNA polymerase (2 U/µl; Panvera, Madison, WI) were added. The reactions were incubated at 55°C-65°C for 30 min. At the end of the incubation, reactions were cooled to 4°C until RNA transcription which was performed at 37°C for 3 h

IDT Exhibit 1020.8798

in 10 µl reactions containing 2.5 µl of linear amplification reaction above, and 40 mM Tris-HCl, pH 8.5, 70 mM KCl, 5.0 mM DTT, 12 mM $MgCl_2$, 110 ugh/ml BSA, 3 mM each rNTP (ATP, UTP, CTP, GTP, Amersham), 7.5% DMSO, 1 Unit/µl rRNasin (Promega, Madison, WI), and 20 Units T7 RNA polymerase (Novagen, Madison, WI).

The products of the enhanced linear amplification generated with each of the composite primers were resolved by gel electrophoresis. The three composite primers were designed to hybridize at the same site on the target strand, and differed by the number of deoxynucleotides at the 3'-end. The highest yield of RNA product was produced with primer IA020. The optimal temperature for isothermal linear amplification step was different for the different primers, as expected. Further optimization of primer design would be expected to enhance the performance of amplification.

(ii) The product of the linear amplification produced with primer IA020, as described above was then subjected to isothermal exponential amplification.

The reaction mixture from the linear amplification described above with primer IA020 at 55°C, including the ssDNA product, was diluted 1:10, and 2.5 µl of this dilution was added to a 5 µl transcription mixture containing: 40 mM Tris-HCl, pH 8.5, 70 mM KCl, 5.0 mM DTT, 12 mM $MgCl_2$, 110 µg/ml BSA, 3 mM each rNTP (ATP, UTP, CTP, GTP, Amersham), 7.5% DMSO, 1 Unit/µl rRNasin (Promega, Madison, WI), and 20 units of T7 RNA polymerase (Novagen, Madison, WI), dNTPs (1.0 mM dATP, 1.0 mM dCTP, 1.0 mM dTTP, 0.8 mM dGTP, 0.2 mM dITP, Amersham), MMLV-reverse transcriptase (200U), RNaseH (0.05 U), PTO IA015c (200nM) and primer IA024 (1µM). The reactions were incubated at 37°C for 0, 5, 10, 20, 40, and 60 mins., respectively. Aliquots from these time points were subjected to gel electrophoresis (5-20% polyacrylamide gel, Novex), the gel was stained with SyberGreen™ and the product bands were visualized using an imaging system. Products were clearly detectable following a 20 minute reaction.

(iii) In an effort to determine conditions for optimizing reaction efficiency, the following conditions/factors were examined:

I. The RNaseH in the above protocol was a thermostable enzyme. It was postulated that replacement of this enzyme with an *E. coli* RNaseH, which is not thermostable, and may be more active at 37°C, would improve amplification conditions.

II. Lowering the reverse transcriptase concentration for improved specificity (reduced smearing).

IDT Exhibit 1020.8799

Reactions were set up as above with the following modifications: *E. coli* RNaseH (0.08 U) and MMLV-reverse transcriptase (25 U) were used in place of the RNaseH and reverse transcriptase in the above protocol. Two target sequences were tested using these conditions: the J gene target sequence as described above, and also a synthetic single stranded DNA target (IA009). The reactions were carried out for 60 min. at 37°C. The amplification efficiency appeared to improve under these conditions yielding more defined product bands and less overall streaking (which is indicative of high non-specific product generation).

Sequences:

The sequences used in this example are as follows:

composite RNA/DNA primers

IA005	r{ACggAUgCggUCUCC} AgTgT	(20-mer)
IA019	r{ACggAUgCggUCUCCAg} TgT	(20-mer)
IA020	r{gACggAUgCggUCU} CCAgTgT	(21-mer)

These primers were synthesized by Dharmacon Research, Inc. (Boulder, CO). The RNA portion of the oligonucleotide was synthesized using a 5'-silyl protecting group in conjunction with an acid-labile 2'-orthoester protecting group (2'-bis(acetoxyethoxy)-methyl ether or "2'-ACE" (Scaringe, SA, FE Wincott, and MH Caruthers. *J. Am. Chem. Soc.* 120:11820-11821 (1998); and Scaringe, SA. Advanced 5'-silyl-2'-orthoester approach to RNA oligonucleotide synthesis. *Methods in Enzymology* (Academic Press, in press)). Oligonucleotides were purified by polyacrylamide gel electrophoresis.

All oligonucleotides other than the composite oligonucleotides were synthesized by Keystone (Division of BioSource International, Camarillo, CA).

PTO

IA015c (55-mer):

ggAATTCTAATACgACTCACTATAgggAgAgCggTACgCTgATCAAAgATCCg Tg-biotin

J-gene target sequence (PCR product from J-gene)

CggTACgCTgATCAAAgATCCgTgCAACAAATgTCATggTCATggTCgTgTTg AgCgCAgCAAAACgCTgTCCgTTAAAATCCCggCAggggTgg**ACACT*ggAgACCgCATC***

IDT Exhibit 1020.8800

CgTCTTgCgggCgAAggTgAAgCgggCgAgCATggCgCACCggCAggCgATCTgTACgTTC
AggTTCAggTTAAACAgCACCCgATTTTCgAgCgTgAAggCAACAACCTgTATTgCgA
AgTCCCgATCAACTTCgCTATggCggCgCTgggTggCgAAATCgAAgTACCgACCCTTg
ATggTCgCgTCAAACTgAAAgTgCCTggCgAAACCC**AgACCggTAAgCTATTCCgTA**
TgCg

PCR primers are shown in **bold**/underlined. Chimeric primer is shown in **bold**, with RNA portion in ***italics***.

Synthetic oligo strand-displacement product (IA009)

IA009 (100-mer):

AgTgTCCACCCCTgCCgggATTTTAACggACAgCgTTTTgCTgCgCTCAACACg
ACCATgACCATgACATTTgTTgCACggATCTTTgATCAgCgTACCg

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

IDT Exhibit 1020.8801

CLAIMS

What is claimed is:

1. A method of generating multiple copies of a nucleic acid sequence of interest, said method comprising the steps of:

(a) hybridizing a single stranded target polynucleotide comprising the nucleic acid sequence of interest with a first primer, wherein said first primer is a composite primer comprising an RNA portion and a 3' DNA portion;

(b) optionally hybridizing a polynucleotide comprising a termination polynucleotide sequence to a region of the target polynucleotide 5' with respect to a site for hybridization of the first primer to the target polynucleotide;

(c) extending the first primer with a DNA-dependent DNA polymerase to generate a complex comprising a first primer extension product and target polynucleotide;

(d) cleaving the RNA portion from the composite primer in the complex of first primer extension product and target polynucleotide with an enzyme that cleaves RNA from an RNA/DNA hybrid such that another composite primer can hybridize to the target polynucleotide and repeating primer extension by strand displacement to produce a displaced primer extension product;

(e) hybridizing a propromoter polynucleotide comprising a propromoter and a region which hybridizes to the displaced primer extension product under conditions which allow transcription to occur by RNA polymerase, such that an RNA transcript comprising sequences complementary to the displaced primer extension products is produced;

(f) hybridizing a second primer to the RNA transcript of step (e);

(g) extending the second primer with RNA-dependent DNA polymerase to generate a complex comprising a second primer extension product and the RNA transcript;

(h) cleaving RNA in the complex of step (g) with an enzyme that cleaves RNA in an RNA/DNA hybrid;

(i) hybridizing the single stranded second primer extension product with a propromoter polynucleotide, wherein the propromoter polynucleotide comprises a propromoter and a region which hybridizes to the single stranded second primer extension product under conditions which allow transcription to occur by RNA polymerase, such that an RNA transcript comprising the sequence of interest is produced;

whereby multiple copies of the nucleic acid sequence of interest are produced.

IDT Exhibit 1020.8802

2. A method of generating multiple copies of a nucleic acid sequence of interest, said method comprising the steps of:

(a) combining:

the complex of step (c) of claim 1;

a composite primer that is hybridizable to the target polynucleotide, wherein the composite primer comprises an RNA portion and a 3' DNA portion;

an enzyme that cleaves RNA from an RNA/DNA hybrid;

a propromoter polynucleotide comprising a propromoter and a region which hybridizes to displaced composite primer extension product;

an RNA polymerase;

a second primer that is hybridizable to a sense RNA transcript comprising the sequence of interest;

an RNA-dependent DNA polymerase; and

a propromoter polynucleotide comprising a propromoter and a region which hybridizes to a second primer extension product; and

(b) incubating the mixture of step (a) under conditions that permit primer hybridization and extension, RNA cleavage, displacement of the first primer extension product from the complex of step (c) of claim 1 when its RNA is cleaved and a composite primer binds to the target polynucleotide in the complex, hybridization of a propromoter polynucleotide to a first primer extension product to form a complex comprising a first primer extension product and a propromoter polynucleotide, hybridization of a propromoter polynucleotide to a second primer extension product to form a complex comprising a second primer extension product and a propromoter polynucleotide, and RNA transcription, whereby multiple copies of the nucleic acid sequence of interest are generated.

3. A method of generating multiple copies of a nucleic acid sequence of interest, said method comprising the steps of:

(a) combining:

the displaced primer extension product of step (d) of claim 1;

a propromoter polynucleotide comprising a propromoter and a region which hybridizes to displaced first primer extension product;

an RNA polymerase;

a second primer that is hybridizable to a sense RNA transcript comprising the sequence of interest;

IDT Exhibit 1020.8803

an RNA-dependent DNA polymerase;

an enzyme that cleaves RNA from an RNA/DNA hybrid; and

a propromoter polynucleotide comprising a propromoter and a region which hybridizes to a second primer extension product; and

(b) incubating the mixture of step (a) under conditions that permit primer hybridization and extension, RNA cleavage, hybridization of a propromoter polynucleotide to a first primer extension product to form a complex comprising a first primer extension product and a propromoter polynucleotide, hybridization of a propromoter polynucleotide to a second primer extension product to form a complex comprising a second primer extension product and a propromoter polynucleotide, and RNA transcription, whereby multiple copies of the nucleic acid sequence of interest are generated.

4. A method of generating multiple copies of a nucleic acid sequence of interest, said method comprising the steps of:

(a) combining:

the RNA transcript of step (e) of claim 1;

a second primer that is hybridizable to a sense RNA transcript comprising the sequence of interest;

an RNA-dependent DNA polymerase;

an enzyme that cleaves RNA from an RNA/DNA hybrid;

a propromoter polynucleotide comprising a propromoter and a region which hybridizes to a second primer extension product; and

an RNA polymerase; and

(b) incubating the mixture of step (a) under conditions that permit primer hybridization and extension, RNA cleavage, hybridization of a propromoter polynucleotide to a primer extension product to form a complex comprising a primer extension product and a propromoter polynucleotide, and RNA transcription, whereby multiple copies of the nucleic acid sequence of interest are generated.

5. A method of generating multiple copies of a nucleic acid sequence of interest, said method comprising the steps of:

(a) combining:

a first primer, wherein the first primer is a composite primer that is hybridizable to a target polynucleotide, and wherein the composite primer comprises an RNA portion and a 3' DNA portion;

IDT Exhibit 1020.8804

optionally a polynucleotide comprising a termination polynucleotide sequence that is hybridizable to a region of the target polynucleotide which is 5' with respect to hybridization of the composite primer to the target polynucleotide;

a DNA-dependent DNA polymerase;

an enzyme that cleaves RNA from an RNA/DNA hybrid;

a propromoter polynucleotide comprising a propromoter and a region which hybridizes to a first primer extension product;

an RNA polymerase;

a second primer that is hybridizable to a sense RNA transcript comprising the sequence of interest;

an RNA-dependent DNA polymerase; and

a propromoter polynucleotide comprising a propromoter and a region which hybridizes to a second primer extension product; and

(b) incubating the mixture of step (a) under conditions that permit primer hybridization and extension, RNA cleavage, displacement of a first primer extension product from a complex comprising a first primer extension product and target polynucleotide when its RNA is cleaved and a composite primer binds to the target polynucleotide in the complex, hybridization of a propromoter polynucleotide to a primer extension product to form a complex comprising a primer extension product and a propromoter polynucleotide, and RNA transcription, whereby multiple copies of the nucleic acid sequence of interest are generated.

6. The method of claim 1, wherein the 5' RNA portion is adjacent to the 3' DNA portion.

7. The method of claim 1, wherein a plurality of composite primers are used.

8. The method of claim 1, wherein the polynucleotide comprising a termination sequence is a template switch oligonucleotide (TSO).

9. The method of claim 8, wherein the TSO comprises a modification in the region which hybridizes to the template, wherein, under a given set of conditions, the TSO binds more tightly to the region as compared to a TSO without the modification.

10. The method of claim 1, wherein the polynucleotide comprising a termination sequence is a blocking sequence.

11. The method of claim 10, wherein the blocking sequence comprises a modification in the region which hybridizes to the target polynucleotide, wherein, under a

IDT Exhibit 1020.8805

given set of conditions, the blocking sequence binds more tightly to the region as compared to a blocking sequence without the modification.

12. The method of claim 1, wherein the enzyme that cleaves RNA is RNaseH.

13. The method of claim 1, wherein the propromoter polynucleotide comprising a propromoter and a region which hybridizes to a displaced first primer extension product is a template switch oligonucleotide (TSO).

14. The method of claim 1, wherein the propromoter polynucleotide comprising a propromoter and a region which hybridizes to a single stranded second primer extension product is a TSO.

15. The method of claim 1, wherein the propromoter polynucleotide comprising a propromoter and a region which hybridizes to a displaced first primer extension product is a propromoter template oligonucleotide (PTO).

16. The method of claim 1, wherein the propromoter polynucleotide comprising a propromoter and a region which hybridizes to a single stranded second primer extension product is a PTO.

17. A method of generating multiple copies of a nucleic acid sequence of interest, said method comprising the steps of:

(a) hybridizing a single stranded target polynucleotide comprising the sequence of interest with a first primer;

(b) hybridizing a propromoter template switch oligonucleotide (TSO) comprising a propromoter sequence and a region that is hybridizable to a region of the target polynucleotide which is 5' with respect to hybridization of the first primer to the target polynucleotide;

(c) extending the first primer with DNA polymerase such that a first primer extension product comprising a sequence complementary to the propromoter sequence of the propromoter TSO is produced, whereby a complex of first primer extension product, target polynucleotide and propromoter TSO is generated, wherein said complex comprises a double stranded promoter region;

(d) transcribing from the double stranded promoter region with a DNA-dependent RNA polymerase to generate a sense RNA transcript;

(e) hybridizing a second primer to the sense RNA transcript of step (d);

IDT Exhibit 1020.8806

(f) extending the second primer with RNA-dependent DNA polymerase to generate a complex comprising a second primer extension product and an RNA transcript;

(g) cleaving RNA in the complex of step (f) with an enzyme that cleaves RNA in an RNA/DNA hybrid;

(h) hybridizing a single stranded second primer extension product with a propromoter polynucleotide, wherein the propromoter polynucleotide comprises a propromoter and a region which hybridizes to the single stranded second primer extension product under conditions which allow transcription to occur by RNA polymerase, such that sense RNA transcripts comprising the sequence of interest are produced;

whereby multiple copies of the nucleic acid sequence of interest are produced.

18. A method of generating multiple copies of a nucleic acid sequence of interest, said method comprising the steps of:

(a) combining:

the complex of step (c) of claim 17;

an RNA polymerase;

a second primer that is hybridizable to a sense RNA transcript comprising the sequence of interest;

an RNA-dependent DNA polymerase;

an enzyme that cleaves RNA from an RNA/DNA hybrid; and

a propromoter polynucleotide comprising a propromoter and a region which hybridizes to a second primer extension product; and

(b) incubating the mixture of step (a) under conditions that permit primer hybridization and extension, RNA cleavage, hybridization of a propromoter polynucleotide to a primer extension product to form a complex comprising a primer extension product and a propromoter polynucleotide, and RNA transcription, whereby multiple copies of the nucleic acid sequence of interest are generated.

19. A method of generating multiple copies of a nucleic acid sequence of interest, said method comprising the steps of:

(a) combining:

an RNA transcript of step (d) of claim 17;

a second primer that is hybridizable to the RNA transcript;

an RNA-dependent DNA polymerase;

an enzyme that cleaves RNA from an RNA/DNA hybrid;

IDT Exhibit 1020.8807

a propromoter polynucleotide comprising a propromoter and a region which hybridizes to a second primer extension product; and

an RNA polymerase; and

(b) incubating the mixture of step (a) under conditions that permit primer hybridization and extension, RNA cleavage, hybridization of a propromoter polynucleotide to a primer extension product to form a complex comprising a primer extension product and a propromoter polynucleotide, and RNA transcription, whereby multiple copies of the nucleic acid sequence of interest are generated.

20. A method of generating multiple copies of a nucleic acid sequence of interest, said method comprising the steps of:

(a) combining:

a target polynucleotide;

a first primer which is hybridizable to the target polynucleotide;

a propromoter template switch oligonucleotide comprising a propromoter sequence and a region that is hybridizable to a region of the target polynucleotide which is 5' with respect to hybridization of the first primer to the target polynucleotide;

optionally a DNA-dependent DNA polymerase;

an RNA polymerase;

a second primer that is hybridizable to a sense RNA transcript comprising the sequence of interest;

an RNA-dependent DNA polymerase;

an enzyme that cleaves RNA from an RNA/DNA hybrid; and

a propromoter polynucleotide comprising a propromoter and a region which hybridizes to a second primer extension product; and

(b) incubating the mixture of step (a) under conditions that permit primer hybridization and extension, RNA cleavage, hybridization of a propromoter polynucleotide to a primer extension product to form a complex comprising a primer extension product and a propromoter polynucleotide, and RNA transcription, whereby multiple copies of the nucleic acid sequence of interest are generated.

21. The method of claim 17, wherein the target polynucleotide is DNA.

22. The method of claim 17, wherein the target polynucleotide is RNA.

23. The method of claim 17, wherein the first and second primers are the same.

24. The method of claim 17, wherein the first and second primers are different.

IDT Exhibit 1020.8808

25. The method of claim 17, wherein the first and second primers hybridize to different complementary sequences.

26. The method of claim 17, wherein the RNA-dependent DNA polymerase is reverse transcriptase.

27. The method of claim 17, wherein the target polynucleotide is RNA, and RNA in the complex of step (c) is cleaved with an enzyme that cleaves RNA in an RNA/DNA hybrid.

28. The method of claim 17, wherein the propromoter polynucleotide comprising a propromoter and a region which hybridizes to a second primer extension product is a propromoter TSO.

29. The method of claim 17, wherein the propromoter polynucleotide comprising a propromoter and a region which hybridizes to a second primer extension product is a propromoter template oligonucleotide (PTO).

30. The method of claim 1 or 17, wherein at least one type of rNTP used is a labeled rNTP, whereby labeled RNA products are generated.

31. The method of claim 1, wherein the RNA-dependent DNA polymerase and DNA-dependent DNA polymerase are one enzyme.

32. The method of claim 17, wherein the DNA polymerase of step (c) is a DNA-dependent DNA polymerase, and wherein the DNA-dependent DNA polymerase and the RNA-dependent DNA polymerase are one enzyme.

33. The method of claim 1 or 17, wherein the RNA-dependent DNA polymerase and enzyme that cleaves RNA from an RNA/DNA hybrid are the same enzyme.

34. The method of claim 1, wherein the DNA-dependent DNA polymerase and enzyme that cleaves RNA from an RNA/DNA hybrid are the same enzyme.

35. The method of claim 17, wherein the DNA polymerase of step (c) is a DNA-dependent DNA polymerase, and wherein the DNA-dependent DNA polymerase and enzyme that cleaves RNA from an RNA/DNA hybrid are the same enzyme.

36. The method of claim 1, wherein the DNA-dependent DNA polymerase, the RNA-dependent DNA polymerase and the enzyme that cleaves RNA from an RNA/DNA hybrid are the same enzyme.

37. The method of claim 17, wherein the DNA polymerase of step (c) is a DNA-dependent DNA polymerase, and wherein the DNA-dependent DNA polymerase, the

IDT Exhibit 1020.8809

RNA-dependent DNA polymerase and enzyme that cleaves RNA from an RNA/DNA hybrid are the same enzyme.

38. A method of sequencing a target polynucleotide, said method comprising (a) amplifying a target polynucleotide by the method of claim 1 or 17 in the presence of a mixture of rNTPs and rNTP analogs such that transcription is terminated upon incorporation of an rNTP analog; and (b) analyzing the amplification products to determine sequence.

39. A method of sequencing a target polynucleotide, said method comprising (a) amplifying a target polynucleotide by the method of claim 1 or 17, wherein RNA transcripts generated from a first primer extension product are amplified in the presence of a mixture of rNTPs and rNTP analogs such that transcription is terminated upon incorporation of an rNTP analog; and (b) analyzing the amplification products to determine sequence.

40. A method of detecting a mutation in a target polynucleotide by single stranded conformation polymorphism, comprising (a) amplifying the target polynucleotide by the method of claim 1 or 17; and (b) analyzing the amplification products for single stranded conformation, wherein a difference in conformation as compared to a reference single stranded polynucleotide indicates a mutation in the target polynucleotide.

41. A method of characterizing a sequence of interest in a target polynucleotide, said method comprising

(i) amplifying a target polynucleotide sequence containing the sequence of interest by the method of claim 1, wherein the sequence of the RNA portion of the composite primer is known, and

(ii) comparing the amplification products if any from step (i) with the amount of amplification products from a reference template

wherein

(1) production of detectably fewer amplification products from the template as compared to the amount of amplification products from the reference template which comprises a region complementary to the RNA portion of the composite primer indicates that the target polynucleotide does not comprise a sequence complementary to the RNA portion of the composite primer and is a sequence variant with respect to the sequence complementary to the RNA portion of the composite primer; or

IDT Exhibit 1020.8810

(2) production of detectably more amplification products from the template as compared to the amount of amplification products from the reference template which does not comprise a region which is complementary to the RNA portion of the composite primer indicates that the target polynucleotide comprises a sequence complementary to the RNA portion of the composite primer and is not a sequence variant with respect to the sequence complementary to the RNA portion of the composite primer.

42. The method of claim 41, wherein the sequence of the RNA portion of the composite primer comprises a sequence complementary to a wild type sequence, and the sequence of interest is characterized in determining the presence or absence of the wild type sequence.

43. The method of claim 41, wherein the sequence of the RNA portion of the composite primer comprises a sequence complementary to a mutant sequence, and the sequence of interest is characterized in determining the presence or absence of the mutant sequence.

44. The method of claim 41, wherein the sequence of the RNA portion of the composite primer comprises a sequence complementary to an allelic sequence, and the sequence of interest is characterized in determining the presence or absence of the allelic sequence.

45. A method of producing a microarray, comprising (a) amplifying a target polynucleotide by the method of claim 1 or 17; and (b) immobilizing the amplification products on a substrate to fabricate a microarray comprising the amplification products.

46. A method of characterizing a sequence of interest, comprising (a) amplifying a target polynucleotide by the method of claim 30; and (b) analyzing the labeled RNA products.

47. The method of claim 46, wherein step (b) comprises contacting the labeled RNA products with at least one probe.

48. The method of claim 47, wherein the at least one probe is provided as a microarray.

49. The method of claim 48, wherein the microarray comprises at least one probe immobilized on a substrate fabricated from a material selected from the group consisting of paper, glass, plastic, polypropylene, nylon, polyacrylamide, nitrocellulose, silicon, and optical fiber.

IDT Exhibit 1020.8811

50. The method of claim 49, wherein the probe is immobilized on the substrate in a two-dimensional configuration or a three-dimensional configuration comprising pins, rods, fibers, tapes, threads, beads, particles, microtiter wells, capillaries, and cylinders.

51. The method of claim 46, wherein step (b) of analyzing the labeled RNA products comprises determining amount of said products, whereby the amount of the sequence of interest present in a sample is quantified.

52. A method of determining gene expression profile in a sample, said method comprising:

(a) amplifying at least two sequences of interest in the sample using the method of claim 1 or 17; and

(b) determining amount of amplification products of each sequence of interest, wherein each said amount is indicative of amount of each sequence of interest in the sample, whereby the gene expression profile in the sample is determined.

53. The method of claim 52, wherein each target polynucleotide is a cDNA.

54. A system for amplifying a sequence of interest, comprising: (a) a first primer which is a composite primer; (b) a second primer; (c) a DNA-dependent DNA polymerase; (d) an RNA-dependent DNA polymerase; (e) a propromoter polynucleotide; (f) an RNA polymerase; and (g) an enzyme that cleaves RNA from an RNA/DNA hybrid.

55. The system of claim 54, further comprising: (h) a polynucleotide comprising a termination polynucleotide sequence.

56. A system for amplifying a sequence of interest, comprising: (a) a propromoter TSO; (b) a first primer; (c) optionally a DNA-dependent DNA polymerase; (d) an RNA-dependent DNA polymerase; (e) an enzyme that cleaves RNA from an RNA/DNA hybrid; (f) optionally a second primer; (g) an RNA polymerase.

57. The system of claim 54 or 56, wherein the DNA-dependent DNA polymerase and the RNA-dependent DNA polymerase are the same enzyme.

58. The system of claim 54 or 56, wherein the DNA-dependent DNA polymerase and enzyme that cleaves RNA from an RNA/DNA hybrid are the same enzyme.

59. The system of claim 54 or 56, wherein the RNA-dependent DNA polymerase and enzyme that cleaves RNA from an RNA/DNA hybrid are the same enzyme.

60. The system of claim 54 or 56, wherein the RNA-dependent DNA polymerase, DNA-dependent DNA polymerase and enzyme that cleaves RNA from an RNA/DNA hybrid are the same.

IDT Exhibit 1020.8812

61. A kit comprising a composite primer that is hybridizable to a target polynucleotide, a propromoter polynucleotide, and instructions for amplifying the target polynucleotide according to the method of claim 1.

62. The kit of claim 61, further comprising a polynucleotide comprising a termination polynucleotide sequence.

63. The kit of claim 61, further comprising a second primer.

64. A kit comprising a propromoter TSO and a first primer, wherein both are hybridizable to a target polynucleotide, and instructions for amplifying the target polynucleotide according to the method of claim 17.

65. The kit of claim 64 further comprising a second primer.

66. A method of determining sequence of a sequence of interest comprising sequencing a nucleic acid amplification product, wherein said nucleic acid amplification product is generated by the method of claim 1 or 17.

67. A method of detecting presence of a nucleic acid sequence of interest in a sample, said method comprising detecting presence of the sequence of interest in a nucleic acid amplification product, wherein said nucleic acid amplification product is generated by the method of claim 1 or 17.

IDT Exhibit 1020.8813

Figure 1A

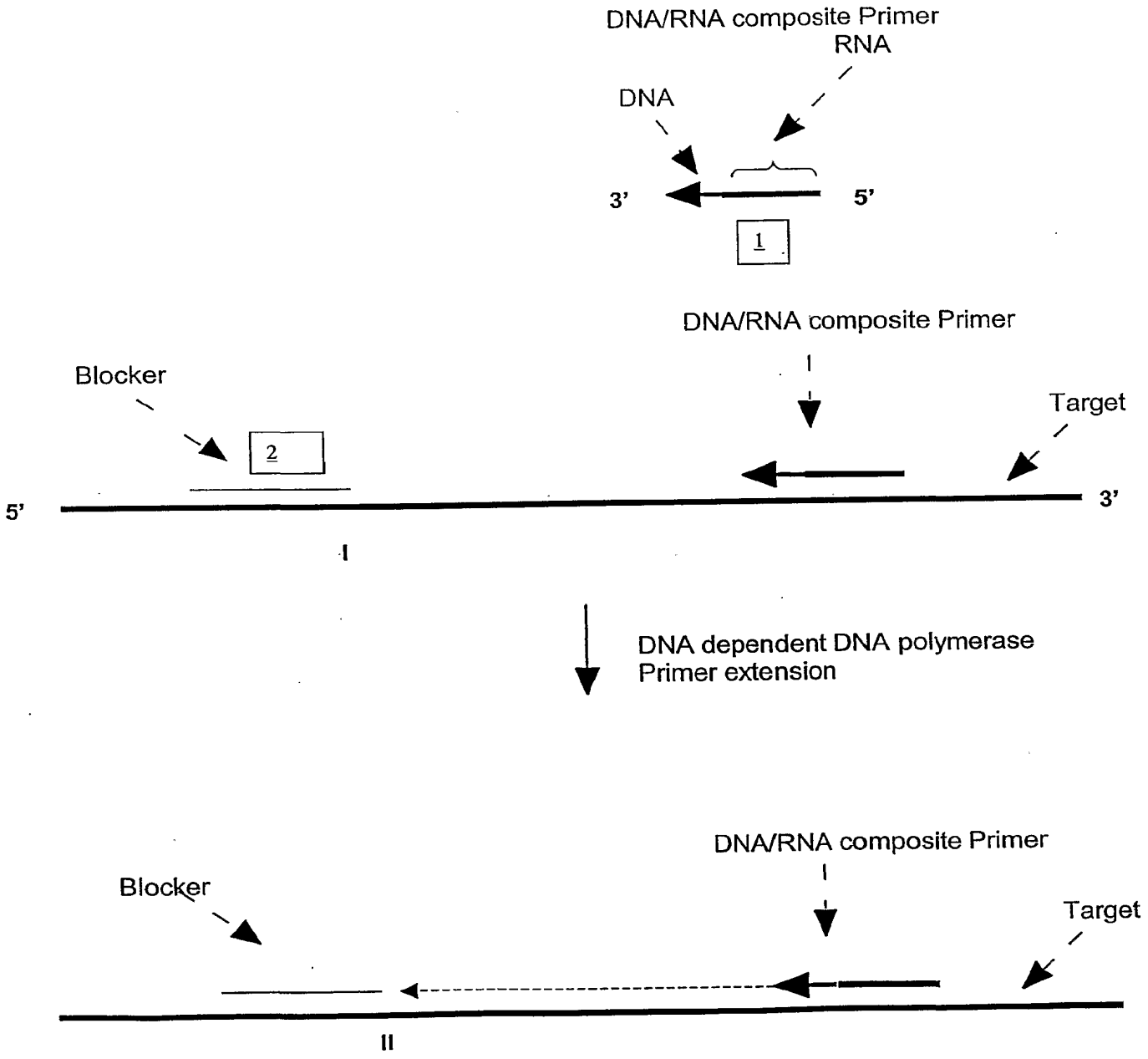

DNA/RNA composite Primer
RNA
DNA
3'
5'
1
DNA/RNA composite Primer
Blocker
Target
2
5'
3'
I
DNA dependent DNA polymerase
Primer extension
DNA/RNA composite Primer
Blocker
Target
II

Figure 1B

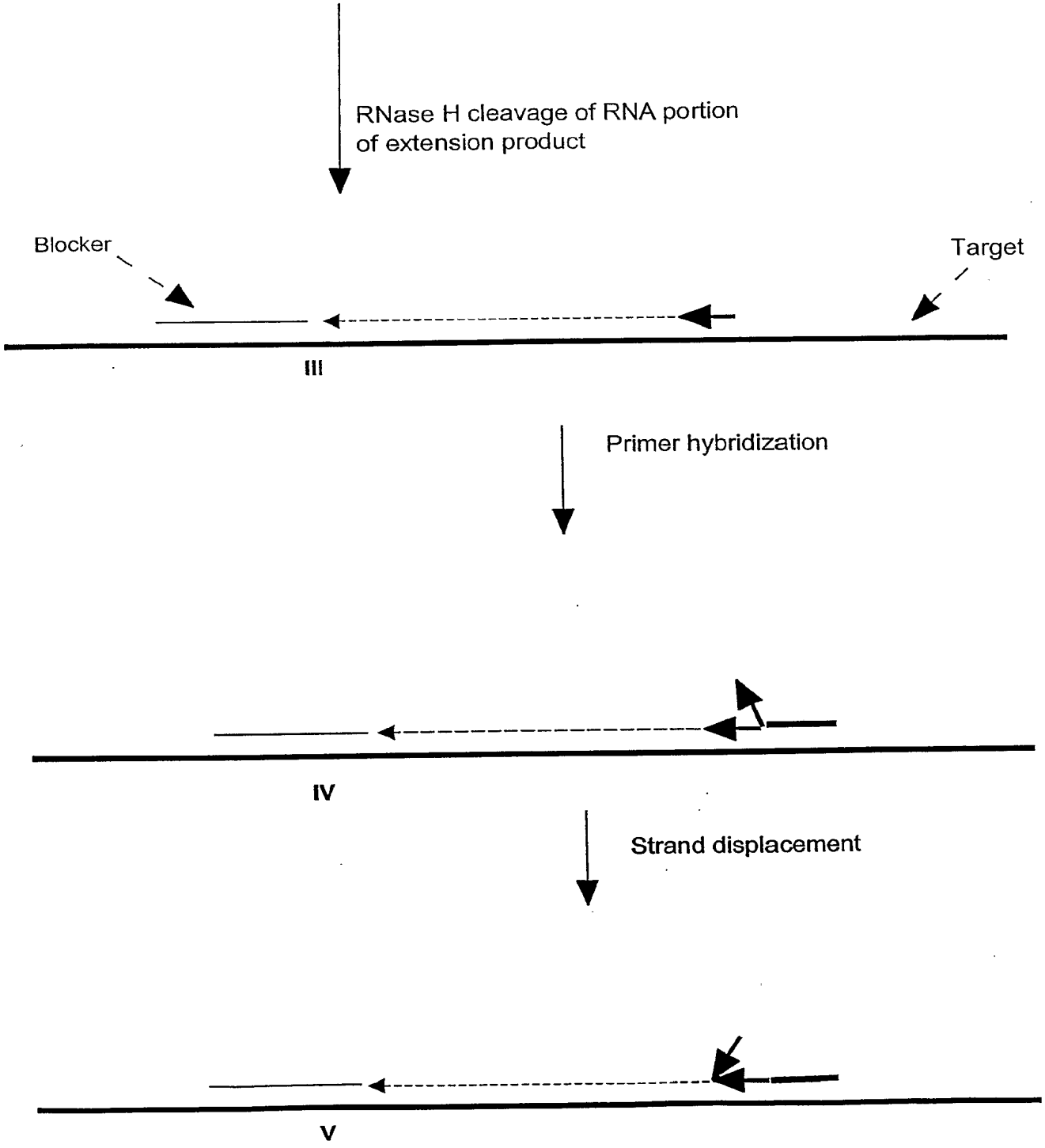

RNase H cleavage of RNA portion
of extension product
Blocker
Target
III
Primer hybridization
IV
Strand displacement
V

Figure 1C

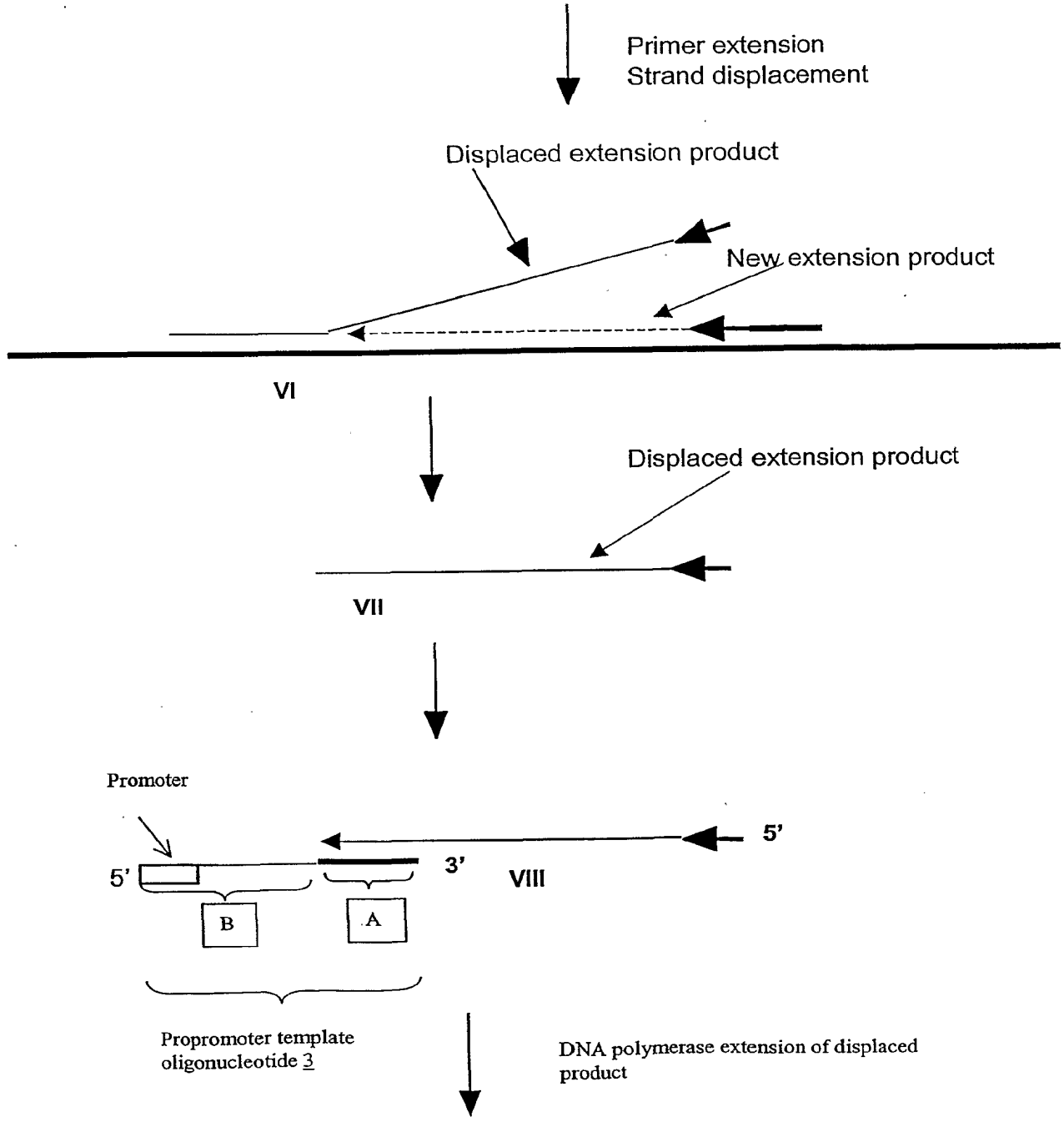

Primer extension
Strand displacement
Displaced extension product
New extension product
VI
Displaced extension product
VII
Promoter
5'
5'
3'
VIII
B
A
Propromoter template
oligonucleotide 3
DNA polymerase extension of displaced
product

Figure 1D

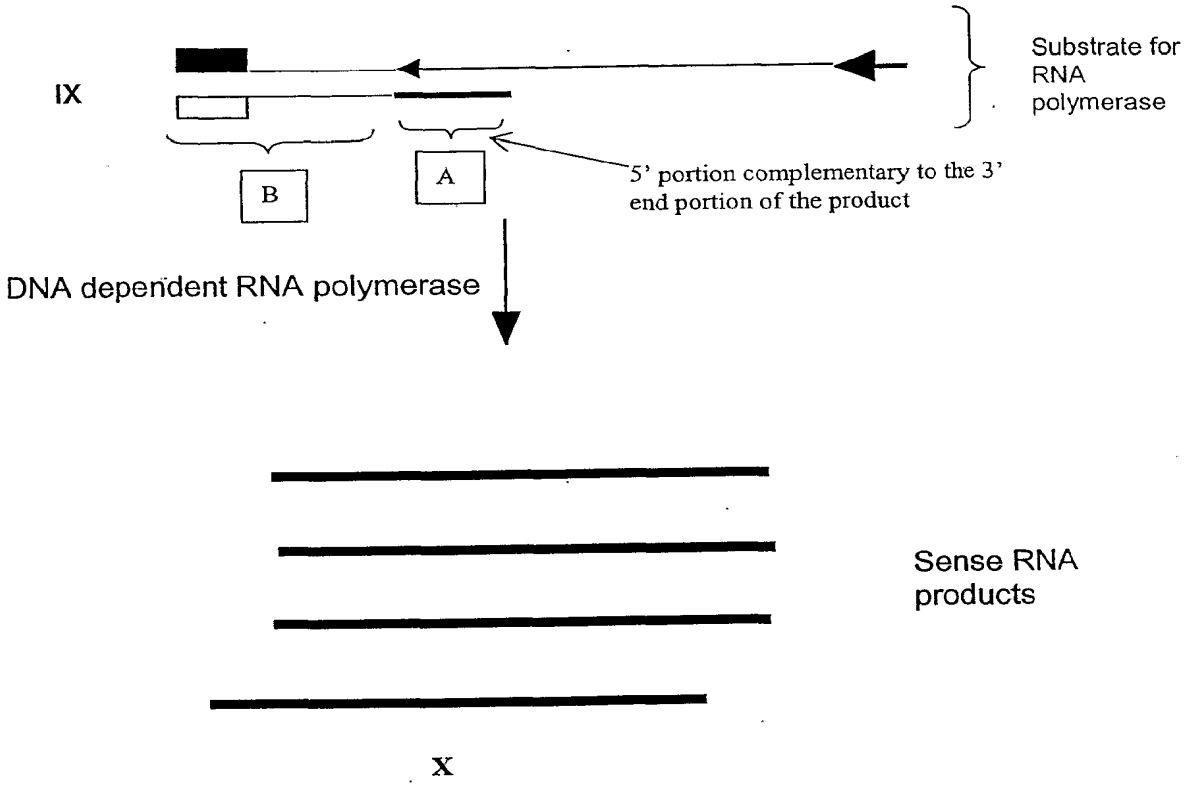

IX
Substrate for RNA polymerase
B
A
5' portion complementary to the 3' end portion of the product
DNA dependent RNA polymerase
Sense RNA products
X

Figure 1E

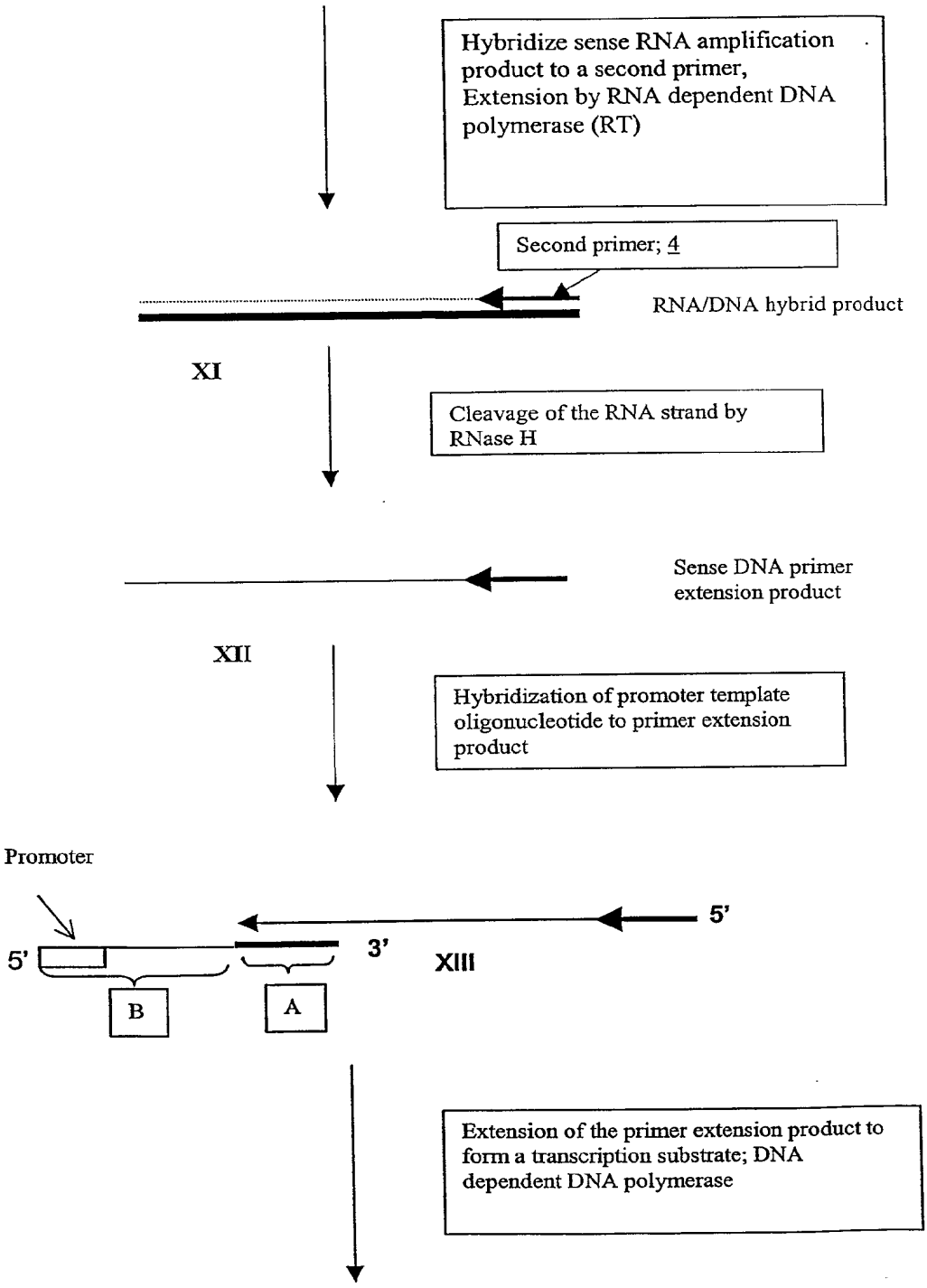

Hybridize sense RNA amplification product to a second primer, Extension by RNA dependent DNA polymerase (RT)
Second primer; 4
RNA/DNA hybrid product
XI
Cleavage of the RNA strand by RNase H
Sense DNA primer extension product
XII
Hybridization of promoter template oligonucleotide to primer extension product
Promoter
5'
3'
5'
B
A
XIII
Extension of the primer extension product to form a transcription substrate; DNA dependent DNA polymerase

Figure 1F

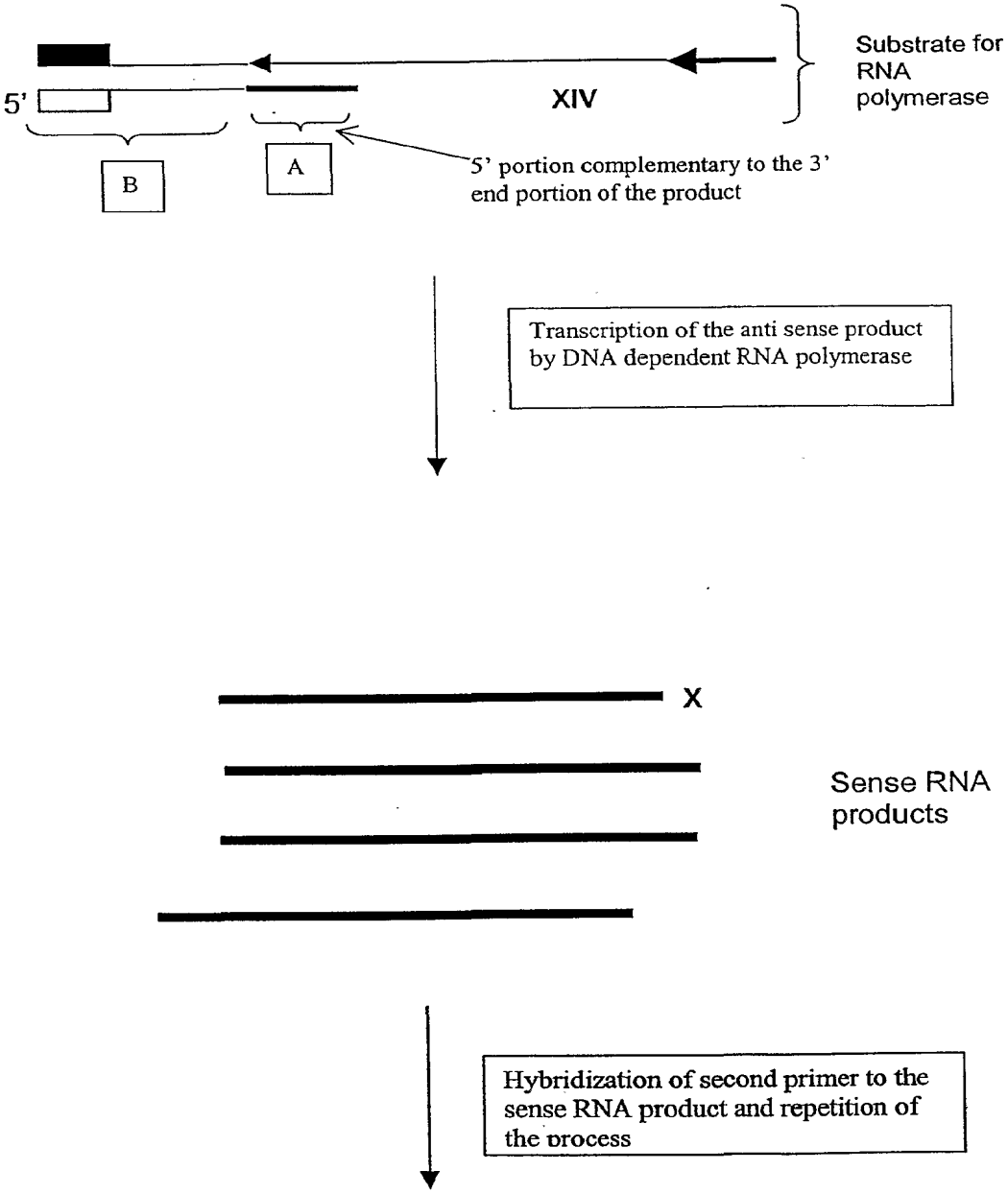

5'
XIV
Substrate for RNA polymerase
B
A
5' portion complementary to the 3' end portion of the product
Transcription of the anti sense product by DNA dependent RNA polymerase
X
Sense RNA products
Hybridization of second primer to the sense RNA product and repetition of the process

Figure 2A

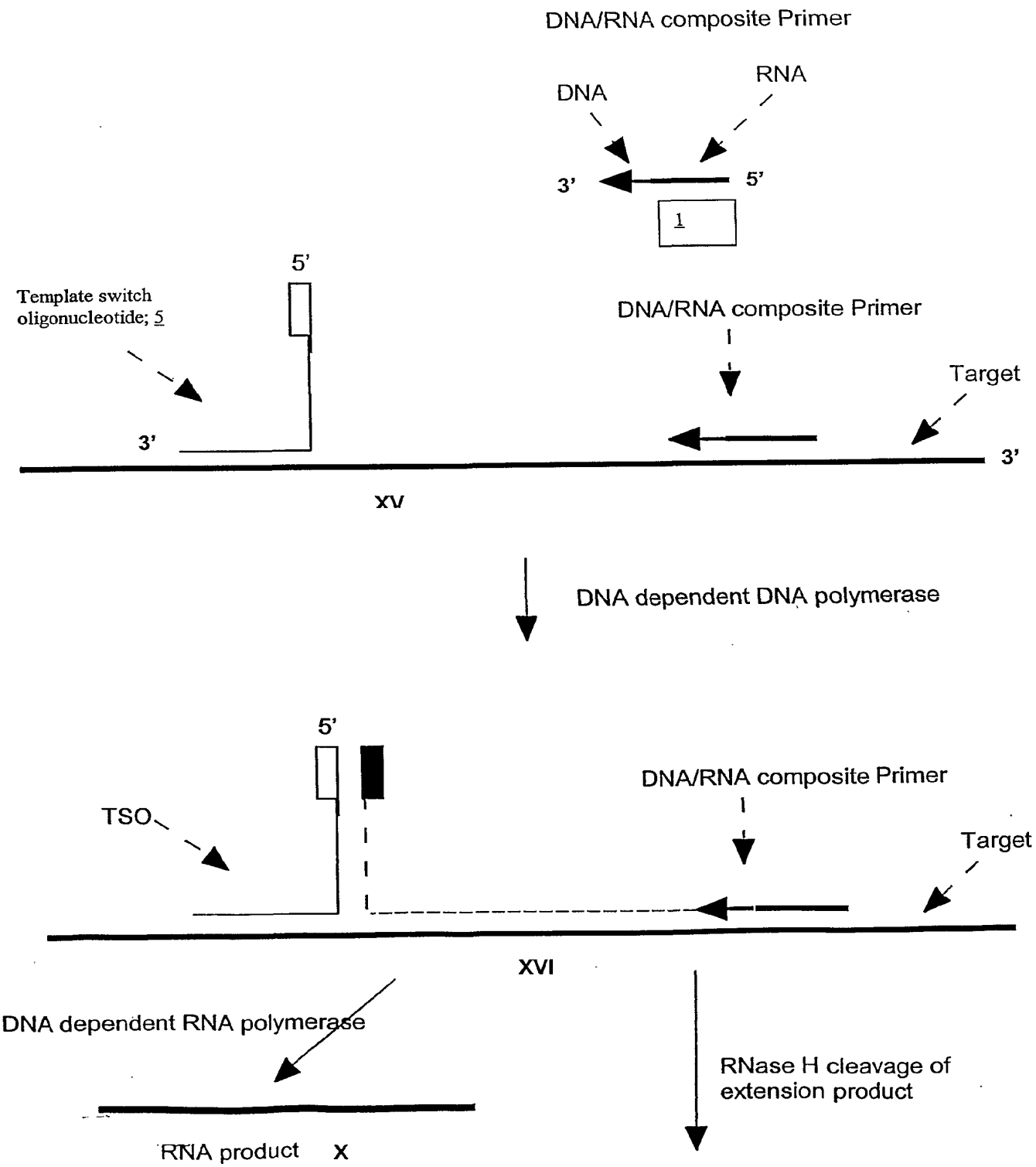

DNA/RNA composite Primer
DNA
RNA
3'
5'
1
5'
Template switch oligonucleotide; 5
DNA/RNA composite Primer
Target
3'
3'
XV
DNA dependent DNA polymerase
5'
TSO
DNA/RNA composite Primer
Target
XVI
DNA dependent RNA polymerase
RNase H cleavage of extension product
RNA product X

Figure 2B

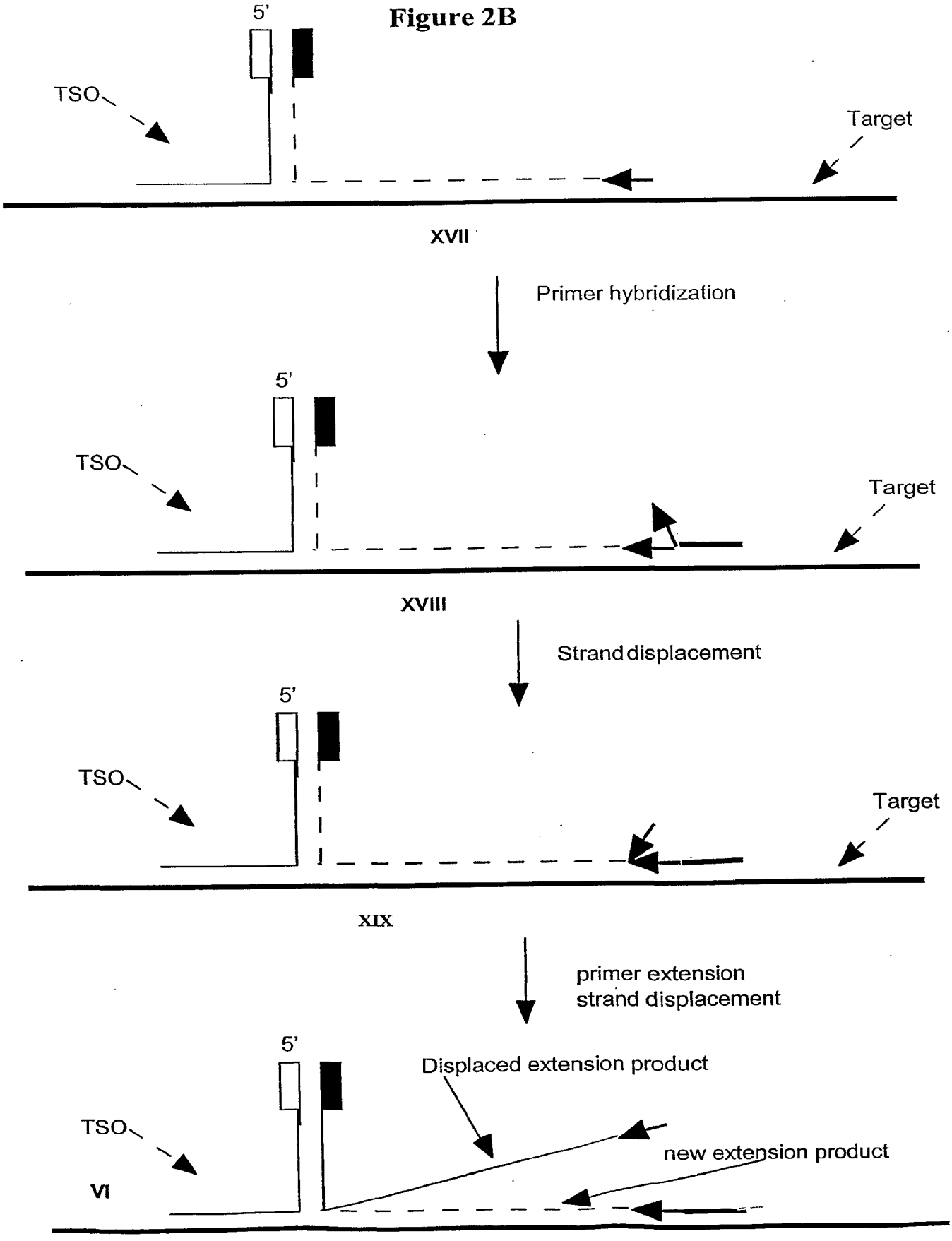

5'
TSO
Target
XVII
Primer hybridization
5'
TSO
Target
XVIII
Strand displacement
5'
TSO
Target
XIX
primer extension
strand displacement
5'
Displaced extension product
TSO
new extension product
VI

Figure 2C

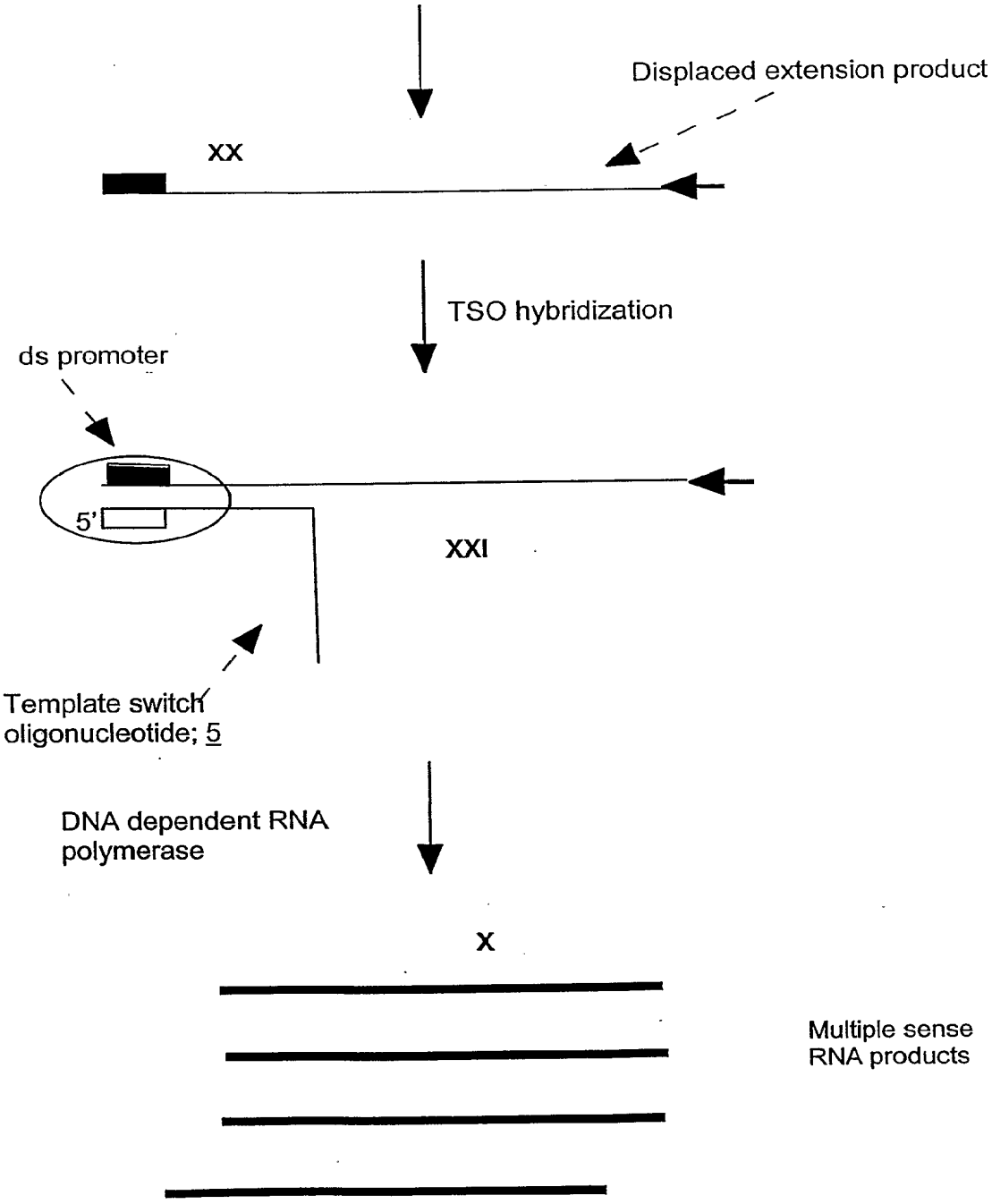

Displaced extension product
XX
TSO hybridization
ds promoter
5'
XXI
Template switch oligonucleotide; 5
DNA dependent RNA polymerase
X
Multiple sense RNA products

Figure 2D

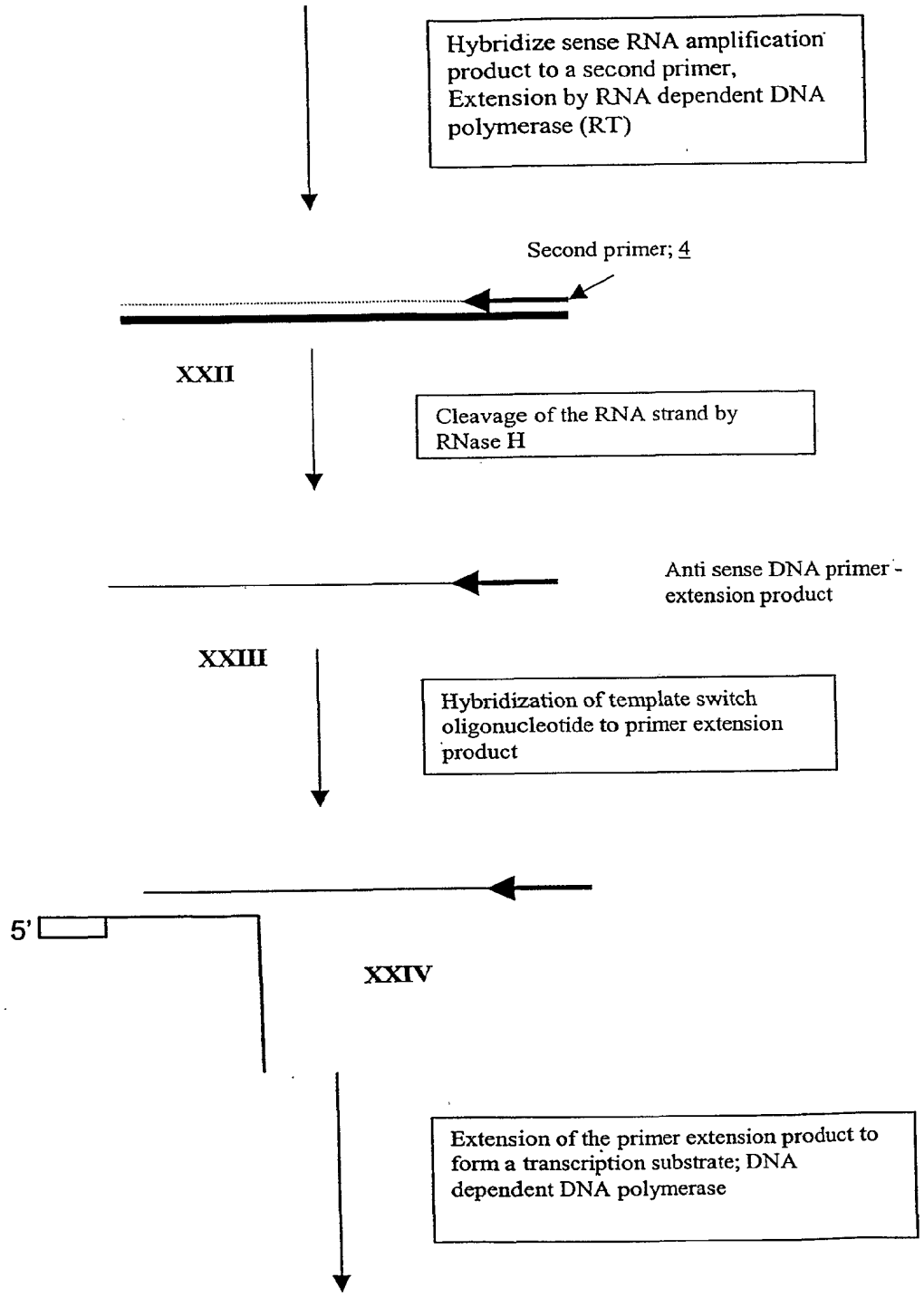

Hybridize sense RNA amplification product to a second primer, Extension by RNA dependent DNA polymerase (RT)
Second primer; 4
XXII
Cleavage of the RNA strand by RNase H
Anti sense DNA primer - extension product
XXIII
Hybridization of template switch oligonucleotide to primer extension product
5'
XXIV
Extension of the primer extension product to form a transcription substrate; DNA dependent DNA polymerase

Figure 2E

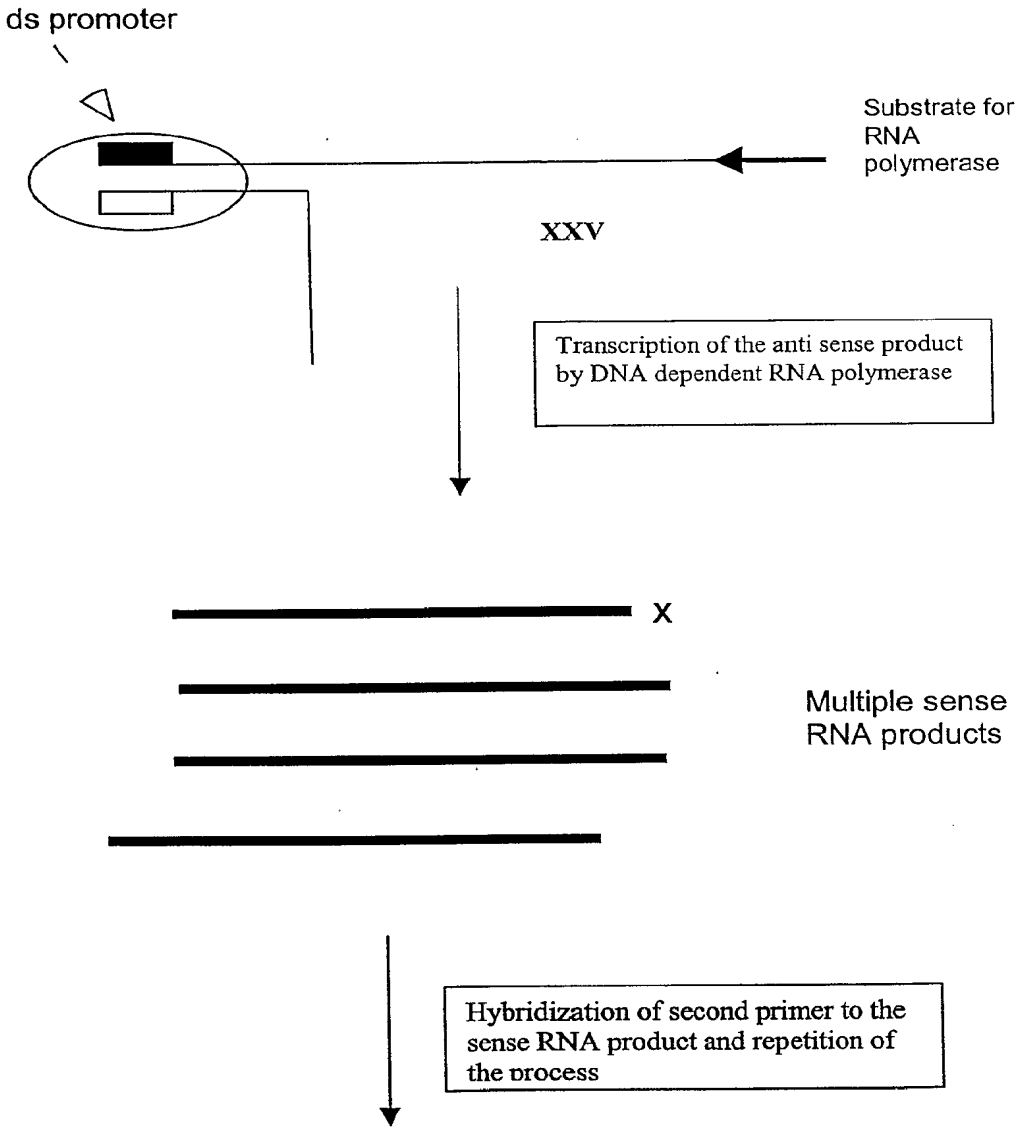

ds promoter
Substrate for RNA polymerase
XXV
Transcription of the anti sense product by DNA dependent RNA polymerase
X
Multiple sense RNA products
Hybridization of second primer to the sense RNA product and repetition of the process

Figure 3A

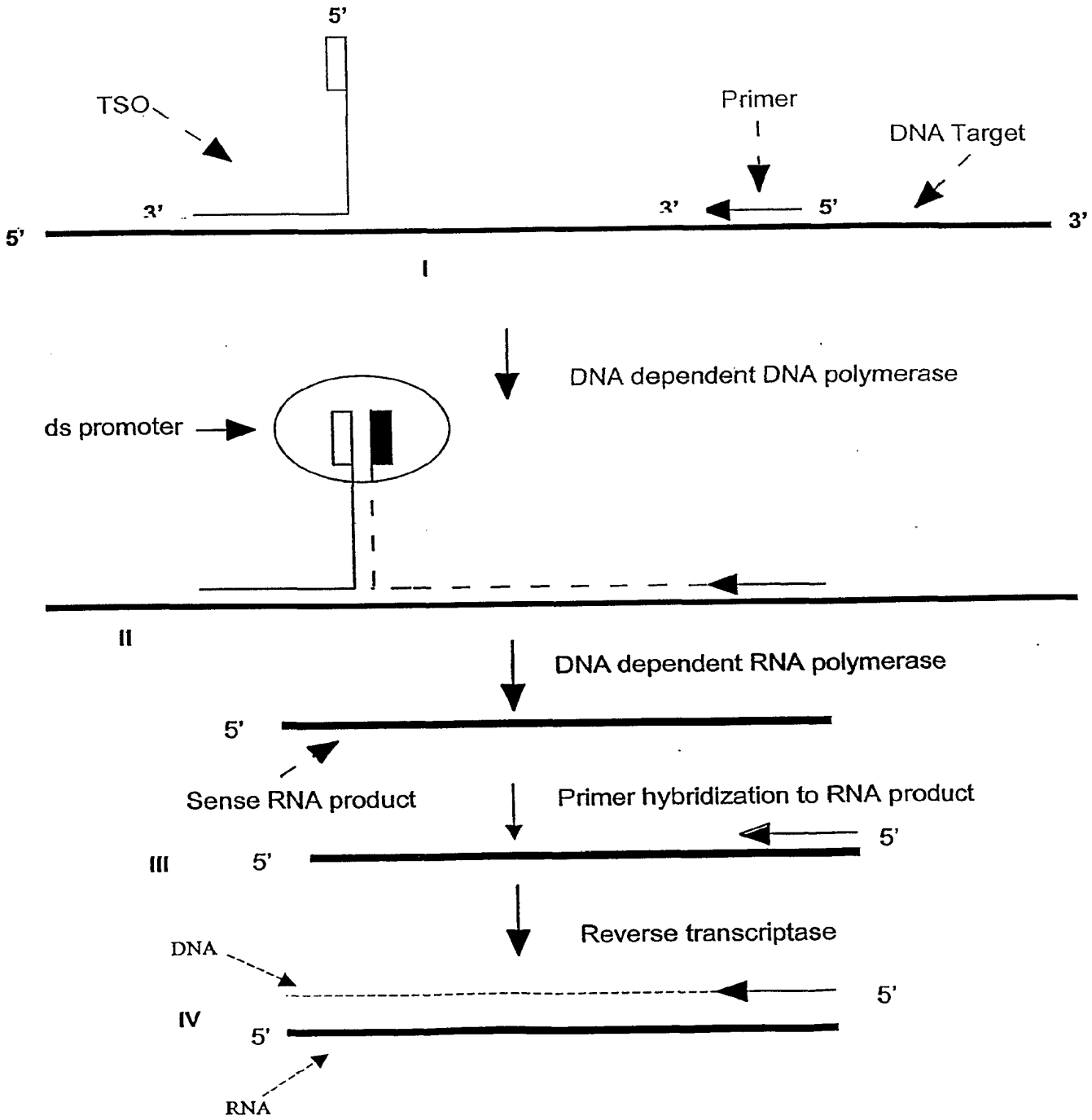

5'
TSO
3'
Primer
DNA Target
3'
5'
5'
3'
I
DNA dependent DNA polymerase
ds promoter
II
DNA dependent RNA polymerase
5'
Sense RNA product
Primer hybridization to RNA product
5'
III
5'
Reverse transcriptase
DNA
5'
IV
5'
RNA

Figure 3B

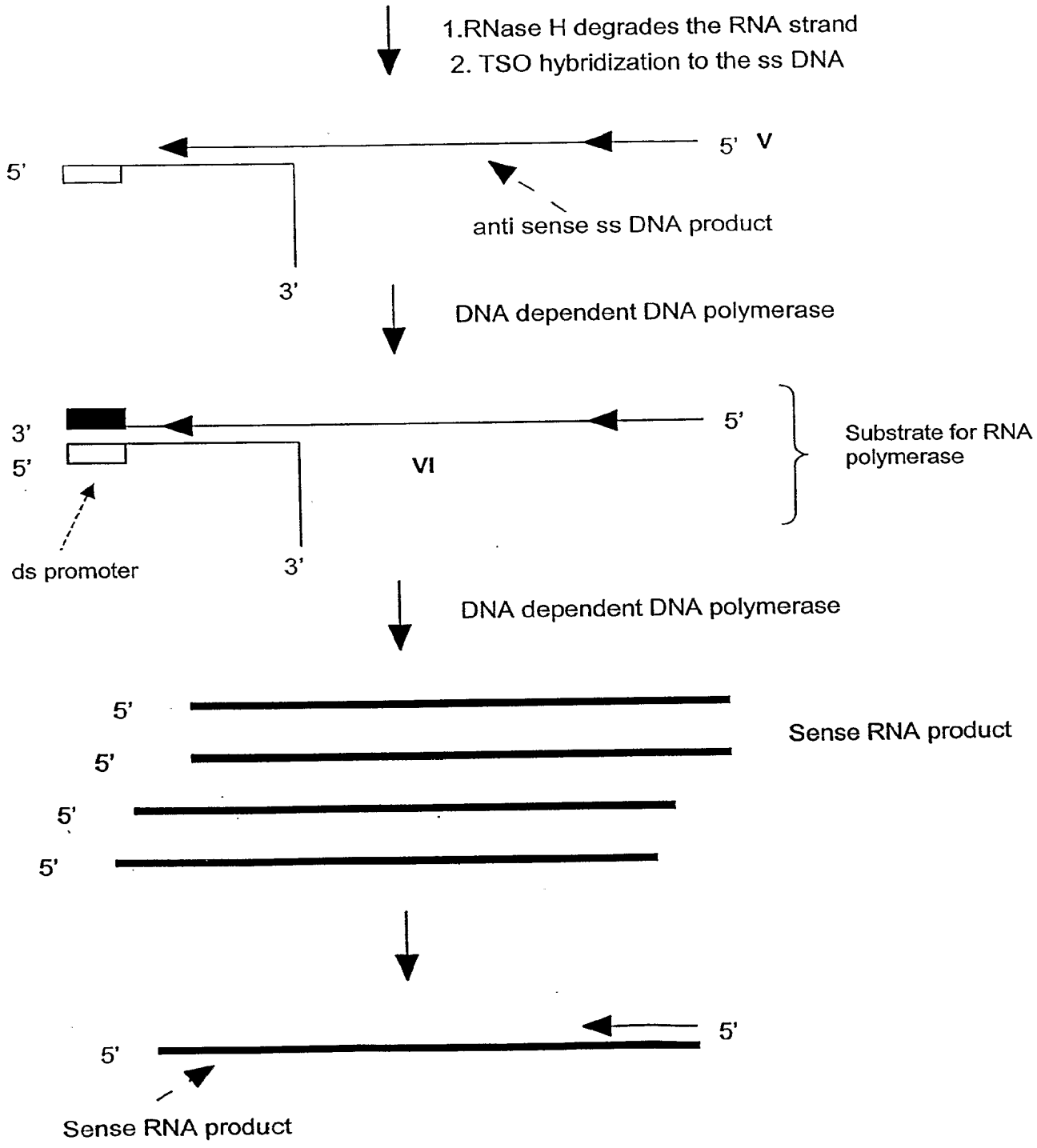

1.RNase H degrades the RNA strand
2. TSO hybridization to the ss DNA
5'
V
5'
anti sense ss DNA product
3'
DNA dependent DNA polymerase
3'
5'
VI
5'
Substrate for RNA polymerase
ds promoter
3'
DNA dependent DNA polymerase
5'
5'
5'
5'
Sense RNA product
5'
5'
Sense RNA product

Figure 4A

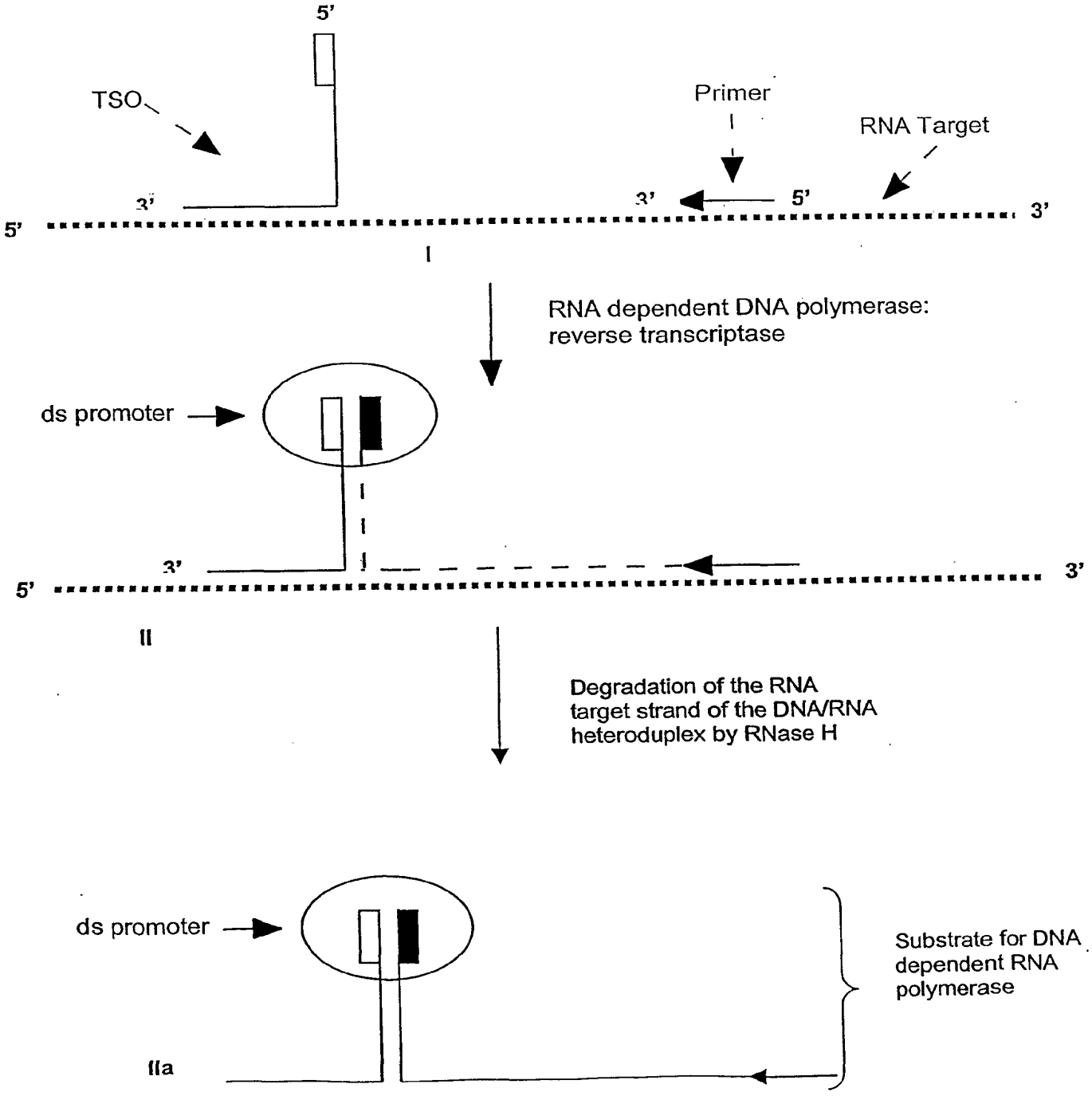
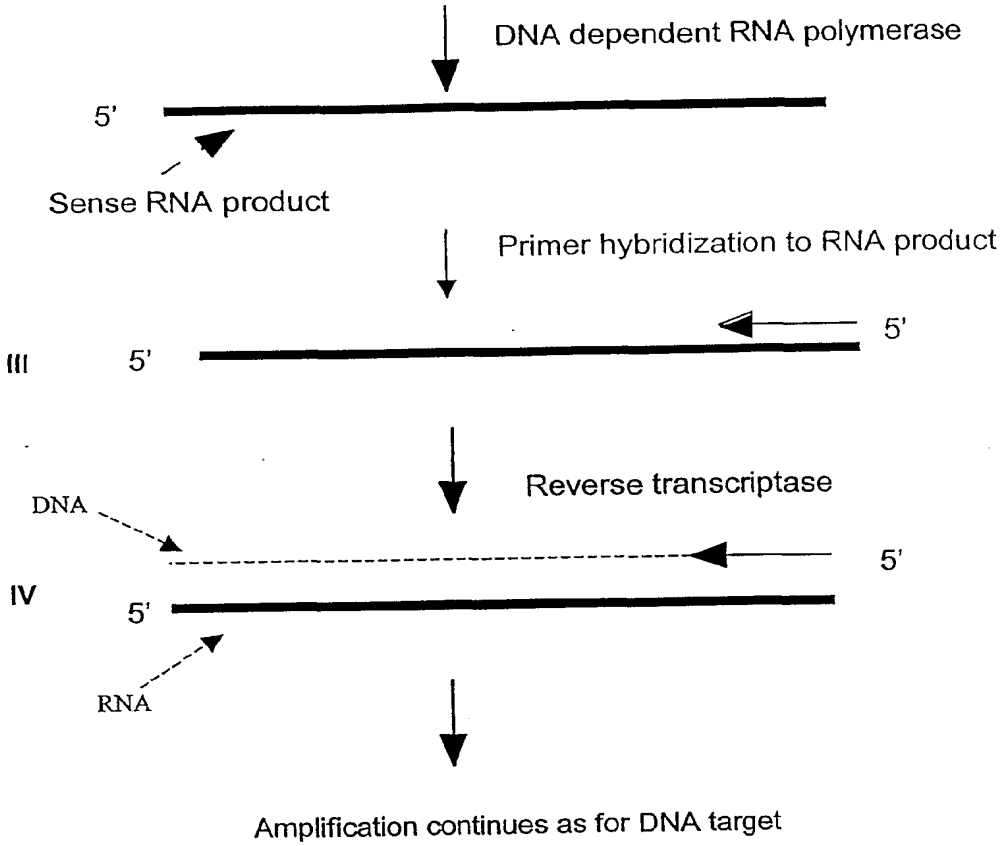

5'
TSO
Primer
RNA Target
3'
3'
5'
5'
3'
I
RNA dependent DNA polymerase:
reverse transcriptase
ds promoter
3'
5'
3'
II
Degradation of the RNA
target strand of the DNA/RNA
heteroduplex by RNase H
ds promoter
Substrate for DNA
dependent RNA
polymerase
IIa

Figure 4B


DNA dependent RNA polymerase
5'
Sense RNA product
Primer hybridization to RNA product
III
5'
5'
Reverse transcriptase
DNA
5'
IV
5'
RNA
Amplification continues as for DNA target

Espacenet

Bibliographic data: WO03027259 (A2) — 2003-04-03

ASSAYS FOR DNA METHYLATION CHANGES

Inventor(s): WANG ZUNDE [US] ± (WANG, ZUNDE)

Applicant(s): EPIGENX PHARMACEUTICALS INC [US]; WANG ZUNDE [US] ± (EPIGENX PHARMACEUTICALS, INC, ; WANG, ZUNDE)

Classification: **- international:** ***C12Q1/68;*** (IPC1-7): C12N
- cooperative: **C12Q1/683** more

Application number: WO2002US30977 20020926 Global Dossier

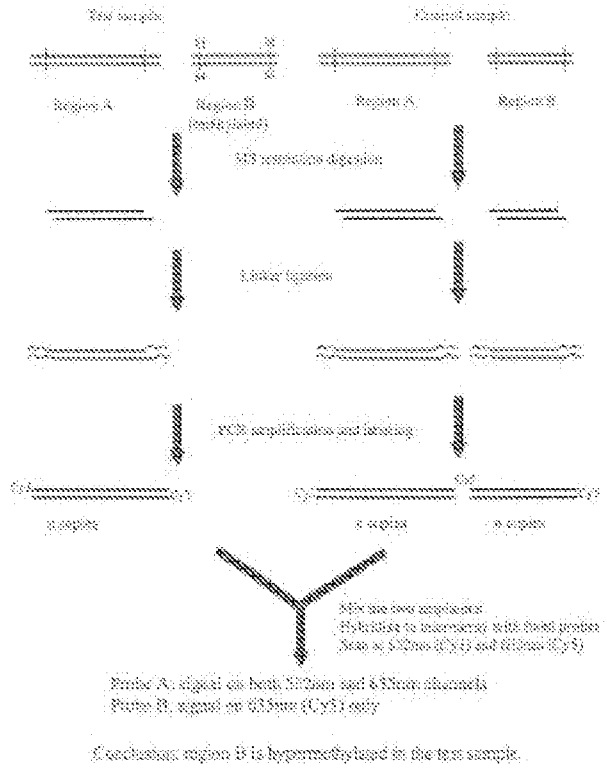
Priority number (s): US20010326045P 20010926

Also published as: WO03027259 (A3) AU2002341891 (A1) EP1438416 (A2) EP1438416 (A4) JP2006500901 (A) more

Abstract of WO03027259 (A2)

A new technique to determine the extent of DNA methylation entails generating DNA fragments of a test sample by cleaving at methylation sites that are not methylated while sparing methylation sites in the DNA that are methylated. This approach provides enhanced sensitivity to differences of even a single methylated cytosine, within or outside of a CpG island, and yet it can be employed to ascertain the methylation status of a region of DNA comprising a CpG island, a DNA region comprising one or more CpG-containing islands, or even a large variety of DNA regions.

IDT Exhibit 1020.8829



IDT Exhibit 1020.8830

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 April 2003 (03.04.2003)

PCT

(10) International Publication Number
WO 03/027259 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/30977

(22) International Filing Date:
26 September 2002 (26.09.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/326,045 26 September 2001 (26.09.2001) US

(71) Applicant *(for all designated States except US)*: **EPIGENX PHARMACEUTICALS, INC.** [US/US]; Pacific Technology Center, 5385 Hollister Avenue, Santa Barbara, CA 93105 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **WANG, Zunde** [CN/US]; 3458 Richland Drive, Apt# 12, Santa Barbara, CA 93105 (US).

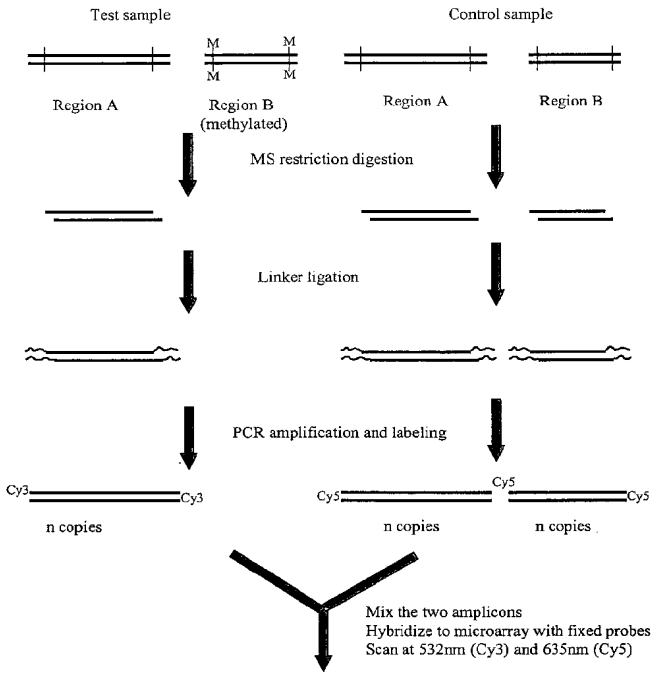
(74) Agents: **WILSON, Barry, S.** et al.; Foley & Lardner, P.O. Box 80278, San Diego, CA 92138-0278 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK,

[Continued on next page]

(54) Title: ASSAYS FOR DNA METHYLATION CHANGES



Probe A: signal on both 532nm and 635nm channels
Probe B: signal on 635nm (Cy5) only

Conclusion: region B is hypermethylated in the test sample.

(57) Abstract: A new technique to determine the extent of DNA methylation entails generating DNA fragments of a test sample by cleaving at methylation sites that are not methylated while sparing methylation sites in the DNA that are methylated. This approach provides enhanced sensitivity to differences of even a single methylated cytosine, within or outside of a CpG island, and yet it can be employed to ascertain the methylation status of a region of DNA comprising a CpG island, a DNA region comprising one or more CpG-containing islands, or even a large variety of DNA regions.

WO 03/027259 A2

IDT Exhibit 1020.8831

TR), OAPI patent (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.

ASSAYS FOR DNA METHYLATION CHANGES

FIELD OF THE INVENTION

[0001] The present invention relates to the detection of a modified nucleotide sequence and, more specifically, to the detection of nucleic acid base methylation of a nucleotide sequence.

BACKGROUND OF THE INVENTION

[0002] The methylation of DNA is thought to have important effects in the regulation of gene expression in eukaryotes, having epigenetic and mutagenic effects on various cellular activities such as differential gene expression, cell differentiation, chromatin inactivation, genomic imprinting, and carcinogenesis. Gonzalgo & Jones, *Mutat. Res.* 386:107-118 (1997). DNA methylation is also known to have a role in regulating gene expression during cellular development. Huang *et al.*, *Human Molecular Genetics* 8: 459-470 (1999). In mammals, DNA methylation usually occurs at cytosines located 5' of guanines, known as CpG dinucleotides. Many CpGs are clustered at/near regulatory regions of genes, which are called CpG islands. DNA methylation or hypermethylation within CpG islands is thought to be especially important in regulating gene expression. A significant number of genes are associated with CpG islands and CpG island hypermethylation has been observed in more than 90 gene promoters. CpG island hypermethylation is associated with the epigenetic inactivation of tumor suppressor or growth regulatory genes in cancer.

[0003] Because methylation of DNA often is reflective of disease status, accurate, sensitive, and efficient methods of detecting changes in methylation patterns would be highly desirable. The existing methods of detecting changes in methylation patterns frequently require laborious steps, such as subtractive hybridization and analysis using Southern blots. Existing methods that depend on reacting unmethylated cytosines with chemical reagents like sodium bisulfite can damage templates and alter base composition, causing problems with a subsequent PCR amplification. Existing methods have other weaknesses such as an inability to detect methylation changes for the entire genome with a

IDT Exhibit 1020.8833

high degree of sensitivity. Accordingly, improved methods with higher throughput and increased accuracy are needed for detecting methylation of DNA.

SUMMARY OF THE INVENTION

[0004] The present invention provides methods for detecting whether the extent of methylation of one or more regions of DNA in a test sample is different from that of a control.

[0005] In one aspect, the invention comprises generating DNA fragments from at least one test sample of DNA by cleaving methylation sites that are not methylated while sparing methylation sites in the DNA that are methylated, ligating an oligonucleotide-linker to the ends of the DNA fragments, amplifying the DNA by initiating amplification from the linker, hybridizing the amplified DNA to one or more polynucleotides immobilized on a solid support, the polynucleotides complementary to one or more regions of DNA in the test sample, and comparing the amplified DNA from the test sample that hybridizes to the immobilized polynucleotides versus that of a control. In an exemplary embodiment the DNA fragments may be labeled with a detectable moiety.

[0006] In another aspect, the invention comprises generating DNA fragments from at least one test sample of DNA by cleaving methylation sites that are not methylated while sparing methylation sites in the DNA that are methylated, separating the fragments of appropriate sizes, ligating an oligonucleotide-linker that is labeled with a detectable moiety to the ends of the DNA fragments, hybridizing the amplified DNA to one or more polynucleotides immobilized on a solid support, the polynucleotides complementary to one or more regions of DNA in the test sample, and comparing the amount of the detectable moiety associated with the polynucleotide for the test sample versus a control.

[0007] In another aspect, the present invention uses the methods described above to determine if an individual has a state of disease associated with an abnormal extent of methylation of one or more regions of DNA in the individual. The above aspects of the invention have many different embodiments which will be addressed in detail below.

IDT Exhibit 1020.8834

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] Figure 1 is a schematic illustration of one embodiment of the inventive methodology. Region A and B represent segments of genomic DNA present in DNA from a test sample and a control sample. Vertical lines indicate the location of methylation sites in the DNA. Vertical lines topped with an "M" indicate that the methylation site is methylated. Thus, as depicted, both DNA regions are not methylated in the control sample while region B is methylated in the test sample. Selective amplification of regions bordered by non-methylated methylation sites is achieved by digesting the DNA into fragments with a methylation sensitive restriction enzyme, ligating the fragments with an oligonucleotide linker, and amplifying the linker ligated fragments with a linker primer. The amplified product from the test and control samples are differentially labeled as shown with Cy3 and Cy5 cyanine dye. Labeled amplicons from the two samples are combined and hybridized to an array with fixed probes for region A and B. The microarray is scanned at the excitation wavelength of 532nm and 635nm. Dual signals for probe A indicate that region A in the test sample is not methylated while a single signal for probe B indicates that region B in the test sample is methylated.

[0009] Figure 2 illustrates the use of unphosphorylated linkers in a scheme for amplifying fragments generated by cleavage with a methylation sensitive restriction enzyme. Vertical bars represent matched base pairs. Panel A shows a double stranded HpaII (C`CGG) linker with a top strand (SEQ ID NO: 1) and a short bottom strand with no 5'-phosphate (SEQ ID NO: 2). Panel B shows a DNA fragment following HpaII restriction. Panel C shows the DNA formed following ligation of the unphosphorylated linker to the HpaII fragment. As seen, the short bottom strand of the linker lacking a 5' phosphate (SEQ ID NO: 2) is not ligated, leaving a nick, while the top strand is ligated to the 5' phosphate of the fragment (resulting in SEQ ID NO: 3). Addition of amplification reactants (polymerase and dNTPs) to the ligated fragments followed by a short incubation at 72°C causes the un-ligated lower strand of the linker to be replaced by a newly synthesized region (see bold font) (SEQ ID NO: 4). Subsequent temperature cycling leads to amplification by priming with unused unphosphorylated linker top strand.

[0010] Figure 3 illustrates the principle of amplicon preparation for interrogating rare methylation sites by cleaving with both a methylation sensitive ("MS") and a

IDT Exhibit 1020.8835

methylation insensitive ("MIS") restriction enzyme. As amplification is desired only for fragments involving an MS restriction (MS/MS and MS/MIS). Panel A indicates that two types of linkers are used: a phosphorylated linker for MS cleavage sites and an unphosphorylated linker for MIS cleavage sites. Panel B shows that following restriction with an MS and MIS enzyme, the lower strand of the MIS linker cannot be ligated because of the missing 5' phosphate. Thus, a nick is present between the short bottom strand of the linker and the fragment. This non-covalently associated lower strand of the linker DNA is unstable and cannot support primer directed amplification. Thus, MIS/MIS fragments with non-covalently attached linker DNA at both ends of the fragments are not amplifiable while fragments generated by MS only (MS/MS) or by MS at one end and MIS at the other end (MS/MIS) can be amplified.

[0011] Figure 4 illustrates how a linker primer can be designed to prevent amplification of linker dimers. Vertical bars represent matched base pairs. Panel A shows a double stranded (SEQ ID NO: 1 and 5) HpaII (C`CGG) linker with a 5'-phosphate. Panel B shows a DNA fragment following HpaII enzyme restriction. Panel C shows a linker primer (SEQ ID NO: 6) for avoiding amplification of linker dimers. Panel D1 depicts a linker dimer (SEQ ID NO:7) formed from the linker in panel A. Panel D2 shows a mismatched base (guanine) at the 3' end of the linker primer (SEQ ID NO: 6) when hybridized to the bottom strand of the linker dimer (SEQ ID NO:7). The presence of a mismatched 3' end nucleotide prohibits the primer from directing amplification. Panel E1 shows linker dimer ligated to the HpaII fragment (resulting in SEQ ID NO: 8 and 9). Panel E2 shows that there is no mismatched base at the 3' end of the linker primer (SEQ ID NO:6) when hybridized to the bottom strand of the ligated fragment (SEQ ID NO:9). The absence of a mismatched 3' base in the linker primer provides for primer directed amplification. Thus, amplification of linker dimers is avoided by extending the linker primer at its 3' end until a 3' base incompatibility is obtained between the primer and the primer dimer (in this example only a one base extension is required) while maintaining 3' end compatibility between the linker primer and the DNA fragment.

[0012] Figure 5 demonstrates that the inventive methylation detection method can identify strain specific methylation of a transgene. Panel A is a Southern blot of mouse DNA from transgenic strains D and B hybridized to a labeled probe for the transgene.

IDT Exhibit 1020.8836

The blot shows that the transgene is methylated in strain B but not in strain D (the increased molecular weight of the methylated ("meth") transgene versus the unmethylated ("um") transgene results from preparing the Southern using a methylation sensitive restriction enzyme). Panel B shows that when amplicons are prepared from strain D and B as described in Example 1, amplified transgene can be detected only from strain D (the strain that does not methylate the transgene). Gel shown panel B was prepared by electrophoresis of transgene specific PCR of the amplicons. Arrow indicates the position of amplified transgene. Panel C shows the amounts of transgene and an unmethylated control gene present in amplicons prepared from strain D and B as described in panel B. The amplicons from strain D were labeled with Cy5 while the amplicons from strain B were labeled with Cy3. The two amplicons were combined and hybridized to an oligonucleotide microarray containing an oligonucleotide for the transgene and one for an unmethylated control gene (both in triplicate). The table shows the amount of Cy3 and Cy5 dye hybridized to the transgene oligonucleotide and the control gene oligonucleotide. The observed reduction in relative Cy3 signal indicates that the transgene in strain B (labeled with Cy3) is methylated as compared to that in strain D. In contrast, the relative similarity in Cy3 and Cy5 signal for the control gene indicates that this gene is similarly methylated in the two strains of mice.

DETAILED DESCRIPTION OF THE INVENTION

[0013] As noted, methylation of DNA in mammals occurs at cytosines, including cytosines in CpG sites, by addition of a methyl group at the pyrimidine ring of the cytosine. A new approach developed herein for determining the extent of methylation change of one or more regions of DNA is provided. The inventive methodology can be used to determine whether the extent of methylation of a DNA region in one sample differs from that of a control.

[0014] In one aspect, the approach of the present invention provides an enhanced sensitivity to differences of even a single methylated cytosine, within or outside of a CpG island, and yet it can be employed to ascertain the methylation status of a region of DNA comprising one or more CpG islands, one or more regions of DNA comprising one or more CpG containing islands, or even a large variety of DNA regions. The latter may entail determining the methylation status of between about 10,000 to 50,000 or more

IDT Exhibit 1020.8837

DNA segments of the genome of a single individual. For example, one could use the method to interrogate the extent of methylation present at all known CpG rich sites (about 45,000 presently known). In another aspect, hundreds or more samples can be analyzed everyday in a standard molecular biology laboratory and a microarray facility. Thus, the present invention allows for rapid and relatively inexpensive interrogation of the methylation status, genome-wide, for many individuals.

[0015] In one aspect, the present invention comprises generating DNA fragments from a sample of DNA by cleaving methylation sites in the DNA that are not methylated while sparing methylation sites in the DNA that are methylated. Oligonucleotide-linkers are ligated to the ends of the DNA fragments. In some embodiments, the linkers are labeled with a detectable moiety. In other embodiments, the DNA is amplified by initiating amplification from the oligonucleotide-linkers. The DNA fragments are analyzed by hybridization to one or more polynucleotides immobilized on a solid support. Finally, the amount of DNA associated with the solid support for the test sample DNA is compared to that of a control to determine whether the extent of methylation of the regions of DNA sought to be evaluated is different between the test sample and the control. In some embodiments, an actual control sample of DNA is used and the test and DNA fragments may be mixed together before hybridization to the immobilized polynucleotides. A discussion of the various steps of these methods as well as additional embodiments based on modifications and/or additions to the procedure follows.

DNA Fragmentation

[0016] DNA fragments can be generated from DNA purified from a cell, tissue or body fluid. Techniques for isolating DNA from such sources are well known to those skilled in the art. See, for example, Sambrook *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL, Vols. 1-3 (Cold Spring Harbor Laboratory Press, 1989). Highly pure DNA is easy to prepare and is preferred, but lower degrees of purity may be used provided the DNA can be cleaved and processed as required in subsequent steps.

[0017] DNA fragments are generated by cleavage at methylation sites that are not methylated while sparing methylation sites in the DNA that are methylated. In this description, the phrase "methylation sites" refers to sites in the DNA that potentially

IDT Exhibit 1020.8838

could be methylated in a cell. Such sites contain a cytosine, which is found often (but not exclusively) in a CpG context. A preferred manner for cleaving DNA fragments at methylation sites that are not methylated, while sparing sites in the DNA that are methylated, is to cleave the DNA with one or more methylation-sensitive agents. A "methylation-sensitive" agent is one that normally cleaves DNA at a methylation site when the site is unmethylated, but does not cleave the site, or cleaves it at low efficiency, when the site is methylated. A methylation-sensitive agent preferably recognizes and cleaves only at methylation sites in the DNA when they are not methylated. Methylation sensitive agents that recognize methylation sites involving methylation of adenine or cytosine nucleotides in bacterial DNA or cytosine nucleotides in CpG, CpNpG and CCWGG sites in plant or mammal DNA can be used.

[0018] A preferred methylation sensitive agent is a methylation sensitive restriction enzyme. Such enzymes and details for their use are commercially available from, for example, New England BioLabs, ProMega Biochem and Boehringer-Mannheim, among others. Techniques for identifying other methylation-sensitive restriction enzymes are known to the skilled artisan. Sambrook *et al., supra*, (1989) provides a general description of methods for using restriction enzymes and other enzymes, for instance.

[0019] Illustrative methylation-sensitive restriction enzymes, suitable for use in the present invention for analyzing methylation changes in CpG sites, include HpaII (C`CGG), AciI (C`CGC), HinP1I (G`CGC), HpyCH4IV (A`CGT), EagI (C`GGCCG), NgoMIV (G`CCGGC), KasI (G`GCGCC), SmaI (CCC`GGG), BstUI (CG`CG), BspDI (AT`CGTA), BstBI (TT`CGAA), SalI (G`TCGAG), and XhoI (C`TCGAG). In accordance with the present invention, cleavage may be achieved by more than one agent, used separately or in combination. Enzymes such as EagI, NgoMIV, KasI and SmaI are especially useful because they are located mainly in CpG islands and are relatively rare in other areas of the genome and the amplicons generated are of lower complexity, meaning that the hybridization signals will be stronger and more specific than in the case where enzymes with a higher frequency of restriction sites are used. Enzymes such as HpaII (C`CGG), AciI (C`CGC), HinP1I (G`CGC), and HpyCH4IV (A`CGT) may be especially convenient because they produce the same "sticky" ends, allowing subsequent

IDT Exhibit 1020.8839

linkage of the cleaved end of the DNA fragment to a single oligonucleotide-linker (*i.e.*, a universal linker). A universal linker which can be ligated to all cleavage ends generated in a particular step of the method is preferred.

[0020] DNA fragments may be generated, at methylation sites that are non-methylated, by using a chemical agent other than a restriction enzyme. For example, DNA can be treated with an agent that specifically protects methylated DNA sites from cleavage (*e.g.*, a methylation binding protein), and then the treated DNA is cleaved via an agent that cleaves the unmethylated methylation sites. Such an agent may be a methylation-insensitive restriction enzyme or a methyl-DNA binding protein that can be fused to or coupled with an endonuclease which can then specifically cut at the methylated CpG sites.

[0021] A "methylation insensitive agent" in this context is one that cleaves DNA at a cleavage site, regardless of whether the site is methylated or not. In a preferred embodiment, the methylation insensitive agent employed is a methylation-insensitive restriction enzyme. Exemplary of methylation-insensitive enzymes for use in the present invention are EcoRI (G`AATTC), ApoI (R`AATTY), Tsp509I (AATT), MseI (TTAA), BfaI (C'TAG), Csp6I (G'TAC), NlaIII (CATG'), DpnII ('GATC), CviJI (RG'CY), Sau3A ('GATC), RsaI (GT'AC), Tsp509I ('AATT), MaeI (C'TAG), NlaIII (CATG'), and DpnI (GA'TC). Other methylation insensitive agents may be chemicals, such as formic acid, hydrochloric acid, and performic acid. For example, an engineered agent such as a fusion protein comprising a methylation-specific DNA-binding protein and an endonuclease may be used for cleaving DNA at methylated and non-methylated methylation sites.

Ligating to Oligonucleotide-linkers and Amplification

[0022] After generating DNA fragments by cleaving at non-methylated sites in the DNA, the ends of the fragments are ligated to an oligonucleotide-linker. An oligonucleotide-linker may be double- or single-stranded and generally is from about 12 to about 24 nucleotides in length, although shorter or longer oligonucleotide-linkers may be used. Double stranded oligonucleotide linkers are preferred. A double-stranded

IDT Exhibit 1020.8840

oligonucleotide-linker may be blunt-ended or have sticky ends. A blunt-ended oligonucleotide-linker may be linked to a blunt-ended DNA fragment. Preferable oligonucleotide-linkers of the invention have a sticky end that is complementary to a sticky end of a DNA fragment for which ligation is desired. Oligonucleotide-linkers may be prepared synthetically and, if double-stranded, may be designed with a sticky end by synthesizing top and bottom DNA stands of different length. In another approach, the linker may be synthetically prepared with a blunt end and then treated with an agent to create the appropriate sticky end. For example, a blunt ended oligonucleotide-linker may be treated with the same methylation sensitive restriction enzyme used to cleave the test or control DNA. In this case, a sticky end that is complementary will be generated on both the DNA fragment and the oligonucleotide-linker to be ligated to the fragment.

[0023] An exemplary double stranded oligonucleotide-linker for ligation to a HpaII-digested DNA fragment can be formed by the combination of SEQ ID NO:1 and SEQ ID NO: 2, shown below (the linker is shown at top of Figure 2). This oligonucleotide linker is partially double stranded and not phosphorylated. An alternative oligonucleotide linker can be made by combining SEQ ID NO:1 with SEQ ID NO: 5 shown below (the linker is shown at top of Figure 4). This oligonucleotide linker is fully double stranded except for a two base segment reflecting the HapII site. This second oligonucleotide linker also is phosphorylated at the 5' end of the bottom strand as shown (see 5' end of SEQ ID NO:5).

5'-CTGCTGACGATGAGTCCTGAGT-3'	(SEQ ID NO:1; top strand)
5'-CGACTCAGGA-3'	(SEQ ID NO:2; bottom strand)
5'-**p**CGACTCAGGACTCATCGTCAGCAG-3'	(SEQ ID NO:5; bottom strand)

With a phosphorylated double stranded oligonucleotide linker such as the combination of SEQ ID NO:1 and 5, both strands of the linker are ligated to the cut DNA (see Figure 4, panel E1). With a non-phosphorylated double stranded oligonucleotide linker such as the combination of SEQ ID NO:1 and 2, only one strand of the linker is ligated to the cut DNA (e.g., the bottom strand as seen in Figure 2, panel C). The other strand is held in place by Watson-Crick base pairing and may be removed by denaturing conditions.

IDT Exhibit 1020.8841

[0024] The characteristics of the oligonucleotide-linker depend on the particular embodiment. For example, the oligonucleotide-linker may be labeled with a detectable moiety so that the DNA fragments may be detected in later steps of the method. Labeling may be accomplished, for example, by conjugating the detectable moiety directly to nucleotides in the linker or the fragment, or by designing the linker to have a unique site for labeling, or by post amplification random labeling. A unique site in the linker may be a chemical linker or a capture sequence of nucleotides. As used herein, a "capture sequence" of nucleotides is a sequence designed to anneal by base pairing with a target oligonucleotide sequence. For example, a capture sequence can be an oligonucleotide with an abiotic and GC-rich sequence that strongly base-pairs with a complementary sequence on another nucleic acid such as a dendrimer. An "abiotic" sequence is one which is not present in the genome of the test DNA. A dendrimer is a branched molecule to which dozens of dye molecules can be attached (Stears et al., *Physiol Genomics* 3: 93-99, 2000). A capture sequence can be designed into the oligonucleotide-linker to support labeling with a detectable moiety such as a labeled dendrimer, which is described in detail below. A post-amplification random labeling can be achieved using normal and fluorescent deoxynucleotides, a random primer, and a DNA polymerase as is well known in the art.

[0025] An oligonucleotide-linker also may be designed to provide a target for initiating synthesis of DNA during DNA amplification. Hence, after ligation to the linker, the ends of the DNA fragments can have a unique sequence which is of sufficient length to provide a unique recognition site for primers selected for in vitro DNA amplification, for example, by the polymerase chain reaction (PCR). Preferably, the oligonucleotide-linker contains a unique sequence sufficient for targeting by the amplification primer. The unique sequence may include a portion of the linker or may include all of the linker. An oligonucleotide linker may thus function as the linker primer. The amplification primer also may include some of the sequence of the adjoining DNA fragment.

[0026] The length of the amplification primers for use in the present invention depends on several factors including the nucleotide sequence and the temperature at which these nucleic acids are hybridized or used during in vitro nucleic acid amplification. The

IDT Exhibit 1020.8842

considerations necessary to determine a preferred length for an amplification primer are well known to the person of ordinary skill in the art. For example, the length of a short nucleic acid or oligonucleotide can relate to its hybridization specificity or selectivity. Because the digested fragments of DNA may contain a complex mixture of nucleic acids, primers which are shorter than about 12 nucleotides may hybridize to more than the linker, and accordingly would not have sufficient hybridization selectivity for amplifying only the DNA fragments cut with a methylation-sensitive agent. However, a 12- to 15-nucleotide sequence generally is represented only once in a mammalian genome. Sambrook *et al.* (1989), *supra*, at pages 11.7-11.8. Accordingly, to eliminate amplification of DNA fragments that are not ligated to oligonucleotide-linkers, amplification primers are chosen which are generally at least about 14 or 15 nucleotides long. Preferably, the primers are at least about 16-17 nucleotides long, and may even be 20-25 nucleotides long or longer. (Sambrook et al., pp. 11.7-11.8). An example of an amplification primer for a HapII oligonucleotide linker is the following:

5'-CTGACGATGAGTCCTGAGTC-3' (SEQ ID NO. 6, linker primer)

This primer is suitable for hybridizing and priming at the oligonucleotide-linker shown above (e.g., the double stranded linker formed using SEQ ID NOs: 1 and 5).

[0027] Ligation of oligonucleotide-linkers to DNA is accomplished using methods well known in the art. For example, see Sambrook *et al.* (1989), *supra*, at 1.53. This typically involves annealing of the oligonucleotide-linker to the DNA fragment, by gradual cooling from 50°C to 25°C, followed by ligation using 400 U of T4 DNA ligase at 16°C. The step of gradual cooling can be eliminated if blunt-ended ligation is contemplated.

[0028] By ligating an oligonucleotide-linker to both ends of a DNA fragment generated by cleavage with a methylation sensitive agent, the ligated DNA fragment will be replicated from both ends when such ends are targeted by the amplification primer, resulting in efficient amplification, for example, by PCR (Figure 1). Since only relatively short fragments (e.g., those less than 1000 base pairs) can be amplified under certain

IDT Exhibit 1020.8843

experimental conditions, the amplified fragments represent regions that are not methylated. Conversely, when a region of the DNA is methylated or has no CpG dinucleotides, it will not be cleaved by a methylation-sensitive agent, no oligonucleotide-linker will be ligated, and no amplification will result (e.g., Figure 1). Therefore, the relative amount of amplified fragments from specific region of test and control samples will reflect their relative methylation level difference.

[0029] The polymerase chain reaction is a preferred method for DNA amplification. PCR synthesis of DNA fragments occurs by repeated cycles of heat denaturation of DNA fragments (*i.e.*, heating to at least about 95°C), incubation at a temperature permitting hybridization of the amplification primers to all or part of the primer adapter ends of the methylation-sensitive cut DNA fragment, and primer extension. These cycles can be performed manually or, preferably, automatically. Thermal cyclers such as the Perkin-Elmer Cetus® cycler (Perkin-Elmer Corp., Boston, MA) are specifically designed for automating the PCR process and are preferred. The number of cycles per round of synthesis generally varies from 2 to more than 50, the optimum number being readily determined by considering the source and amount of the nucleic acid template, the desired yield, and the procedure to be used for detection of the DNA. Heat stable amplification enzymes such as *pwo, Thermus aquaticus* or *Thermococcus litoralis* DNA polymerases are commercially available and eliminate the need to add enzyme after each denaturation cycle. Basic PCR techniques are described in U.S. patents by Saiki *et al.*, see also 1988 *Science* 239: 487, and by U.S. patents No. 4,683,195, No. 4,683,202 and No. 4,800,159, and other PCR variations well-known to the field.

[0030] The precise conditions for PCR and hybridization have a much reduced effect on the outcome of present invention as compared to that in other PCR-based assays in cases where both amplification and hybridization reactions are competitive: the DNA from the test and control sample is combined and amplified simultaneously, thus, competing equally for amplification resources like the polymerase and nucleotide triphosphates. Therefore, more accurate assays are attainable with the present invention, and more variation or error in performing the assay is tolerated without adversely affecting the results. Nevertheless, the variables generally affecting PCR including

IDT Exhibit 1020.8844

temperature, salt and cation concentration, pH, and related conditions may be optimized for each situation.

[0031] A variety of enzymes suitable for DNA amplification such as in PCR are well known in the art and include, for example, Pwo, Escherichia *coli* DNA polymerase I, Klenow fragment of *E. coli* DNA polymerase I, T4 DNA polymerase, T7 DNA polymerase, *Thermus aquaticus* (Taq) DNA polymerase, *Thermococcus litoralis* DNA polymerase, SP6 RNA polymerase, T7 RNA polymerase, T3 RNA polymerase, T4 polynucleotide kinase, Avian Myeloblastosis Virus reverse transcriptase, Moloney Murine Leukemia Virus reverse transcriptase, T4 DNA ligase, E. coli DNA ligase or Qβreplicase, and the like. Mixtures of enzymes also can be used (*e.g.*, Pwo and Taq). Preferred amplification enzymes are the Pwo and Taq polymerases because of their high fidelity and high processibility. Other DNA-amplification enzymes, known and yet to be discovered, may be utilized in the present invention.

[0032] It is desirable to use conditions for amplification such that the DNA fragments differing in GC content can be amplified to a similar extent. Addition of betaine (1.2 M) to a PCR reaction can be used to assist in the amplification of GC-rich regions. Under such conditions, a region of 42% GC and region of 73% GC can be amplified about equally well. Other approaches to achieve balanced amplification of GC rich and GC poor fragments are known in the art.

[0033] In situations where the site to be analyzed can be cleaved as a fragment (suitable for amplification) with a methylation sensitive agent, the oligonucleotide linker can be phosphorylated or unphosphorylated. A scheme for use of a phosphorylated oligonucleotide linker is shown in Figure 1. When a phosphorylated oligonucleotide-linker with sticky ends is used, the possibility of forming dimers of the oligonucleotide-linkers can occur and such dimers can be ligated during the ligation step. Linker dimers may hybridize to amplification primers and be amplified, thus depleting the resources of the amplification reaction, thereby reducing the amount of amplification of ligated DNA fragments. Different strategies for minimizing this problem are available and include, for example: (1) Using column chromatography, such as with a silica-gel-membrane type spin column, to purify the ligated DNA fragments from the linker dimers. In this approach, further elimination of linker dimers may be achieved by washing the column with 30%

IDT Exhibit 1020.8845

guanidine hydrochloride prior to the regular washing step; and (2) Using a linker primer that is comprised of a 3'-mismatched nucleotide when hybridized to the linker dimer but a 3' matched nucleotide when hybridized to the linker-ligated genomic fragment. An example of the second approach is shown in figure 4.

[0034] A scheme for amplification when a unphosphorylated oligonucleotide linker is used is shown in Figure 2. In this embodiment, the linker is designed as shown with a short lower strand that is not phosphorylated. Thus, following the step of ligation, a nick or space will be present between the lower strand of the linker and the fragment. Primer amplification may be achieved following ligation by a short chain extension reaction step in which polymerase and NTPs are added to the linked fragment and the mixture incubated (e.g., 72°C) for several minutes. During the polymerization step, the unligated bottom strand is denatured and removed allowing the polymerase to synthesize a bottom strand flush to the end of the top strand. Exponential amplification may be achieved by application of cycling temperatures. In this case, there is no need to add a linker primer since excess top strand from the oligonucleotide linker that remains from the ligation step can serve as the PCR primer.

[0035] In the case when a particular methylation site to be interrogated is located too far from the next nearest methylation site, making it difficult to amplify and thus detect a fragment of such length, the DNA can be cleaved with both a methylation sensitive and a methylation insensitive agent to generate amplifiable fragments. The methylation sensitive and insensitive cleavage agents may be used separately or at the same time if compatible. A particular methylation insensitive agent can be chosen (preferably a methylation-insensitive restriction enzyme) to provide a second cleavage site nearer to the site cleaved by the methylation-sensitive agent, yielding a fragment size more suited to amplification. Generally, fragments lengths of 50-1000 base pairs will be suitable for amplification.

[0036] A preferred methylation insensitive agent is a methylation insensitive restriction enzyme. The methylation insensitive restriction enzyme preferably is chosen to produce a different sticky end on the DNA than the end generated by the methylation-sensitive agent. In this case, oligonucleotide-linkers that are distinct from each other and which ligate only to one of the types of sticky ends generated should be used.

IDT Exhibit 1020.8846

[0037] To avoid amplifying fragments generated solely by cleavage with the methylation insensitive agent, one may use a phosphorylated oligonucleotide linker for ligation to the methylation sensitive site and a unphosphorylated oligonucleotide linker for ligation to the methylation insensitive site. This approach is illustrated further in Figure 3. As is seen in the figure, following ligation to the methylation insensitive site, a gap or "nick" remains between one strand of the linker and the fragment. The gap prevents amplification of the strand by a linker primer chosen to hybridize to that strand. Under such circumstance, exponential amplification will occur for DNA fragments resulting from cleavage at both ends by a methylation-sensitive agent and for DNA fragments resulting from cleavage at one end by a methylation-sensitive agent and at the other end by a methylation-insensitive agent. DNA fragments resulting from cleavage at both ends by a methylation insensitive enzyme will not be amplified.

[0038] When both a methylation-sensitive agent and a methylation-insensitive agent are used to cleave the DNA, oligonucleotide-linkers used preferably will selectively ligate only to the ends generated by one or the other agent. Selective ligation may not be needed when each of the DNA cleavage agents are used in a separate reaction and the oligonucleotide-linker for the first reaction is added before the DNA is exposed to the second cleavage agent.

[0039] In an alternative embodiment, the step of DNA amplification may be dispensed with entirely. In this case, the ends of the DNA fragments generated by cleavage with the methylation sensitive agent are ligated to an oligonucleotide-linker, the DNA fragments are size selected to isolate a particular size range, the fragments are labeled with a detectable moiety, and then hybridized and analyzed as described further ahead. The DNA fragments are size selected, preferably to a size range of about 2,500 base pairs or less in length, more preferably about 1,500 base pairs in length or less and even more preferably about 1,000 base pairs in length or less. DNA fragments may be size separated before or after ligation to oligonucleotide-linkers. Sizing may be achieved by any method known in the art, including, gel electrophoresis, column chromatography, density gradient centrifugation or ultracentrifugation (see, e.g. Steensel et al., (2001) *Nature Genetics* 27:304-308), filtration through membranes of controlled pore size, and

IDT Exhibit 1020.8847

the like. Filtration through membranes of controlled pore size may be achieved by applying an external pressure or by centrifugation.

[0040] In the embodiment without amplification, the detectable moiety is preferably attached to the oligonucleotide-linker and the moiety is preferably of suitably high specific activity. A labeled dendrimer or a fluorescent resonant energy transfer dye are examples of high specific activity detectable moieties. The oligonucleotide linker can have a capture sequence designed for binding to a detectable moiety such as a labeled dendrimer. The oligonucleotide-linker may be labeled before or after it is ligated to the DNA fragments.

[0041] The amount of DNA digested may be increased to improve detection sensitivity in embodiments that do not use amplification. DNA in the range of 0.5-1.5 mg may be suitable for this purpose, although persons of ordinary skill in the art realize that the quantity of DNA sufficient to allow omitting the DNA amplification step depends on other features of the assay and can be readily determined without resort to undue experimentation.

[0042] In yet another embodiment, the step of DNA amplification and the step of ligating the DNA fragments to an oligonucleotide-linker may be dispensed by using a sufficiently large amount of the DNA, size-fractionating the digested fragments, optionally labeling the fragments through random priming (*e.g.*, with Cy3 and Cy5 dye-labeled dCTP), and hybridizing them to a solid support such as a microarray. For example, see Steensel *et al.*, *Nature Genet.* 27: 304 (2001).

Labeling DNA

[0043] DNA can be labeled with a detectable moiety to improve detection in later steps. Labeling can occur at any of a variety of steps in the method. For example, the detectable moiety may be attached to the oligonucleotide-linker, which is then ligated to the DNA fragment. In another approach, the detectable moiety may be attached to the amplification primer, which is incorporated into the amplified DNA. In yet another approach, the DNA fragment may be directly labeled with the detectable moiety. In some situations, it may be preferable to label with a detectable moiety at more than one step

IDT Exhibit 1020.8848

(*e.g.*, use a labeled oligonucleotide-linker and direct labeling). This may be useful when increased sensitivity is desired. Various combinations of detectable moieties also may be used to label the same molecule to increase its specific activity.

[0044] The phrase "detectable moiety" is used here to denote any molecule (or combinations of molecules) that may be attached or otherwise associated with a molecule so that the molecule can be detected indirectly by detecting the detectable moiety. A detectable moiety can be a radioisotope (*e.g.*, iodine, indium, sulfur, hydrogen etc.) a dye or fluorophor (*e.g.*, cyanine, fluorescein, rhodamine), protein (*e.g.*, avidin, antibody), enzyme (peroxidase, phosphatase, etc.), or any other agent that can be detected directly or indirectly. An enzyme is an example of a detectable moiety detected by indirect means. In this case, the enzyme is attached to the target nucleic acid and the presence of the enzyme is detected by adding an appropriate substrate that when acted upon by the enzyme, causes the substrate to change in color or to release a cleavage product that provides a different color from the original substrate.

[0045] A fluorescent detectable moiety can be stimulated by a laser with the emitted light captured by a detector. The detector can be a charge-coupled device (CCD) or a confocal microscope, which records its intensity. In the case of an array, the intensities provided by the array image can be quantified by measuring the average or integrated intensities of the spots. Interpreting the data from a microarray experiment can be assisted using special software, such as Dapple; see (http://www.cs.washington.edu/homes/jbuhler/research/dapple/).

[0046] A preferred detectable moiety is a cyanine dye such as Cy-5 and Cy-3. For example, test DNA fragments can be labeled with Cy-5, while control DNA fragments can be labeled with Cy-3 (or vice versa). In this way, the two DNA fragment preparations may be mixed and then hybridized together to the solid support, scanned with the above devices at the corresponding wavelength for each dye, with the measured densities providing the ability to distinguish the relative amounts of each DNA fragment that hybridized from the two samples.

[0047] A detectable moiety may include more than one chemical entity such as in fluorescent resonance energy transfer (FRET). In FRET based assays, interaction

IDT Exhibit 1020.8849

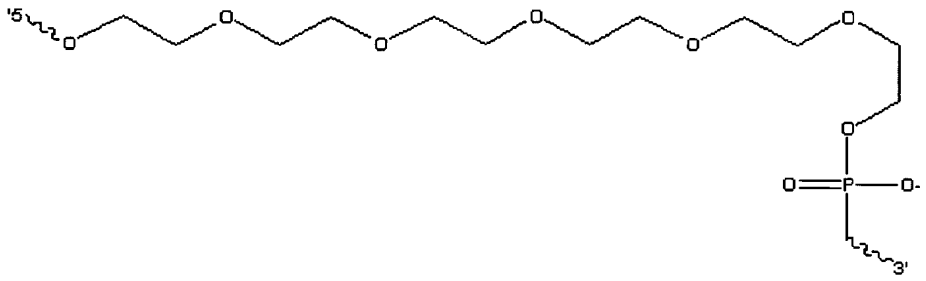
between biomolecules is measured indirectly by conjugating one of a pair of carefully selected fluorescent dyes to each of the molecules of interest. The absorption spectrum of the acceptor must overlap fluorescence emission spectrum of the donor, and donor and acceptor transition dipole orientations must be approximately parallel. In FRET, when these fluorescent dyes are held in close proximity (typically 10–100 Å) due to binding of the biomolecules, a unique fluorescence signal is developed that specifically confirms the proximity and thus the binding reaction. Resonance transfer results an overall enhancement of the emission intensity. For instance, see Ju *et. al.* (1995) *Proc. Nat'l Acad. Sci. (USA)* 92: 4347. To achieve resonance energy transfer, the first fluorescent molecule (the "donor" fluor) absorbs light and transfers it through the resonance of excited electrons to the second fluorescent molecule (the "acceptor" fluor).

[0048] In most cases, when donor and acceptor dyes are different, FRET can be detected by the appearance of sensitized fluorescence of the acceptor or by quenching of donor fluorescence. When the donor and acceptor are the same, FRET can be detected by the resulting fluorescence. Donor/acceptor pairs of dyes that can be used include, for example, fluorescein/tetramethylrohdamine, IAEDANS/fluroescein, EDANS/DABCYL, fluorescein/fluorescein, BODIPY FL/BODIPY FL, and Fluorescein/ QSY 7 dye. See, *e.g.*, U.S. patent No. 5,945,526 to Lee *et al.* Many of these dyes also are commercially available, for instance, from Molecular Probes Inc. (Eugene, Oregon).

[0049] In another embodiment, signal amplification may be achieved using labeled dendrimers as the detectable moiety. A dendrimer is a bulky three dimensional molecule that can be labeled to very high specific activity, providing signal amplification (see, *e.g.*, Physiol Genomics 3:93-99, 2000). Fluorescently labeled dendrimers are available from Genisphere (Montvale, NJ). These may be chemically conjugated to the DNA fragments by methods known in the art. The labeled dendrimer also can be attached to the amplified DNA by designing a capture sequence in the amplification primer (*e.g.*, a long abionic, GC-rich nucleotide sequence), which can be targeted by a complementary sequence in the dendrimer. In the event the capture sequence in the primer creates problems during amplification, a polymerization "stop signal" can be added 3' of the capture sequence (between the capture sequence and the primer sequence) so that the capture sequence in

IDT Exhibit 1020.8850

the primer will not be amplified. The C18 spacer below is an example of a useful "stop signal."



Persons of ordinary skill in the art will realize that this is just one example, and many other stop signals are available and known in the art. For example several may be obtained from Integrated DNA Technologies, Inc. (Coralville, Iowa) and other sources.

Hybridization to Immobilized Polynucleotides

[0050] A solid support with polynucleotides immobilized thereon is useful for capturing DNA fragments prepared as described above to determine methylation status for any region of the genome. The fragments are preferably labeled with a detectable moiety as already discussed to assist in detecting the amount of the DNA that hybridizes to the immobilized polynucleotides.

[0051] The immobilized polynucleotides are designed to specifically hybridize to a region of the DNA for which determination of its extent of methylation is desired. Thus, if one is interested in determining the extent of methylation at a single site in the genome, the solid support may have a polynucleotide complementary to a fragment that would arise in the method if that site were not methylated. A solid support may have multiple different immobilized polynucleotides, each designed to anneal to a different DNA fragment that may be generated in the assay depending on whether the methylation site at one end of the fragment is methylated. As described further below, multiple polynucleotides may be immobilized to a single support in the form of an array. The solid phase may have from hundreds to tens of thousands of polynucleotides immobilized

IDT Exhibit 1020.8851

as an array, depending on the size of the array and the technology used for immobilization.

[0052] Quantitation of the extent of methylation of one or more regions of DNA in a test sample versus a control sample is achieved through competitive amplification of fragments of DNA with single pairs of oligonucleotide-linkers (or one oligonucleotide-linker in the case of single enzyme digestion) and/or competitive hybridization to the solid support. This may be appreciated from the following example where genomic DNA from the test and control are digested with at least one methylation sensitive enzyme. If 100,000 fragments are amplified using oligonucleotide-linkers, and 100 of the fragments are differentially methylated between the two samples, 99,900 fragments will be amplified to very similar extent in both samples after the reactions are saturated, and the 100 fragments will be amplified differentially. After mixing the two amplified products (each labeled with a distinct dye), hybridizing to polynucleotides immobilized on the support (*e.g.*, a microarray), and scanning at the specified wavelength, 99,900 wells or spots on the support will show neutral color (i.e., similar levels of hybridization signal to the two DNA sources) while 100 wells or spots will show a red (Cy-5) or green (Cy-3) color (i.e., differential levels of hybridization signal to the two DNA sources), the latter thereby indicating a difference in methylation level at those sites (*e.g.*, CpG sites).

[0053] Using the combination of Cy3 labeled test sample and Cy5 labeled control sample, the resulting hybridization signal strength on the array following hybridization can be interpreted as follows:

Red: Means differential methylation between the test sample and control sample. The control sample dominates, so the control sample is less methylated than the test sample.

Green: Means differential methylation between the test sample and the control sample. The test sample dominates, so the test sample is less methylated than the control sample.

Yellow: Means about equal methylation for the test sample and the control sample.

No color: Means either heavy methylation in both test and control samples or the immobilized polynucleotide sequence is not present in the genomic DNA of the test or control.

IDT Exhibit 1020.8852

Color intensity: Although less predictive than color, a high intensity indicates a low overall extent of methylation for the region in both the test and control, while a low intensity indicates a high overall extent of methylation for the region in both the test and control.

A. Immobilized polynucleotides

[0054] The sequence of polynucleotides immobilized to a support can be determined by resort to DNA sequence readily available in publicly available sequence databases, including that gleaned from the Human Genome Project. As used herein, a polynucleotide can be a single or a double stranded nucleic acid and may be DNA or RNA. A polynucleotide includes at least 2 nucleotides and may be prepared synthetically or by cleaving natural nucleic acid. The term polynucleotide encompasses the term oligonucleotide, which refers to a shorter polynucleotide of between two to about 100 nucleotides in length.

[0055] Polynucleotide sequences can be designed to hybridize to virtually any region of the genome where methylation is suspected. A virtual (i.e., "computer based") restriction digestion may be performed on human genomic DNA sequence using the recognition sequence for the methylation-sensitive agents of interest, and the identity of sequence for the immobilized polynucleotides determined. Immobilized polynucleotides can be selected to interrogate particular sequences in genes, such as the upstream regulatory region of a gene, the protein encoding portion of the gene, or downstream sequence associated with the gene. In such cases, the immobilized polynucleotides may be designed to hybridize to a sequence in the gene or near to the gene.

[0056] The immobilized polynucleotide may be a short segment of DNA such as an oligonucleotide or the immobilized polynucleotide can be a longer stretch of DNA such as a cDNA or portion thereof or genomic fragment of DNA. Polynucleotides also may be chosen to interrogate intergenic regions, which are referred to herein as sequence upstream of a promoter and downstream of a poly A site. In either case, the sequence of immobilized polynucleotides is preferably selected to hybridize to the region of interest, but not to other regions. This can be determined with the aid of computer sequence programs such as BLAST.

IDT Exhibit 1020.8853

[0057] Polynucleotides for immobilization should be designed such that they will not hybridize to other fragments that may be generated by the particular embodiment of the method chosen. In the case where amplification is used, the immobilized polynucleotides preferably have sequence complementary to a segment of DNA that have potential cleavage sites for a methylation-sensitive agent. The sites should be close enough to each other so that the fragment can be amplified under the specified PCR conditions if the sites are not methylated, and, therefore, cleaved. Polynucleotides for immobilization are preferably oligonucleotides which comprise a hybridizing sequence preferably between about 40 to 70 nucleotides in length and more preferably between about 50 to 60 nucleotides in length. As already discussed, immobilized polynucleotides can be longer than 100 nucleotides in length and may even constitute a cDNA or genomic DNA fragment thereof.

B. Solid Support

[0058] Polynucleotides immobilized to a solid support are used for detecting hybridization to the amplified and detectably-labeled target nucleic acid. The polynucleotides may be immobilized on any solid support, organic or inorganic, or a combination of any of these - in the form of particles, strands, precipitates, gels, sheets, tubing, spheres, containers, capillaries, pads, slices, films, plates, slides, and the like. Typical supports are made of glass, plastic, or nylon.

[0059] With respect to implementing the present invention, a solid support for immobilizing polynucleotides preferably is flat but may take on alternative surface configurations. For example, the solid support may contain raised or depressed regions on which polynucleotide synthesis takes place or where polynucleotides are attached. In some embodiments, the solid support will be chosen to provide appropriate light-absorbing characteristics. Thus, the support may be a polymerized Langmuir Blodgett film, functionalized glass, Si, Ge, GaAs, GaP, SiO_2, SiN_4, modified silicon, or any one of a variety of gels or polymers such as (poly)tetrafluoroethylene, (poly)vinylidendifluoride, polystyrene, polycarbonate, or combinations thereof. The solid support may be a glass microscope slide.

IDT Exhibit 1020.8854

[0060] The surface of the solid support can contain reactive groups, which could be carboxyl, amino, hydroxyl, thiol, or the like suitable for conjugating to a reactive group associated with the immobilized polynucleotides. Polynucleotides can be attached to the solid support by chemical or physical means such as through ionic, covalent or other forces well known in the art. Polynucleotides also can be attached to a solid support by means of a spacer molecule, essentially as described in U.S. patent No. 5,556,752 to Lockhart *et al.* A spacer molecule typically comprises between 6-50 atoms in length although larger and shorter spacers are possible and includes a surface attaching portion that attaches to the solid support.

[0061] Attachment to the support can be accomplished by carbon-carbon bonds, using supports having, for instance, (poly)trifluorochloroethylene surfaces or, preferably, by siloxane bonds, employing glass or silicon oxide as the solid support, for example. Siloxane bonding can be formed by reacting the support with trichlorosilyl or trialkoxysilyl groups of the spacer. Aminoalkylsilanes and hydroxyalkylsilanes, bis(2-hydroxyethyl)-aminopropyltriethoxysilane, 2-hydroxyethylaminopropyltriethoxysilane, aminopropyltriethoxysilane or hydroxypropyltriethoxysilane are useful surface attaching groups. Additionally, for use in synthesis of polynucleotides, the spacer can have a protecting group, attached to a functional group (*i.e.*, hydroxyl, amino or carboxylic acid) on the distal or terminal end of the spacer (opposite the solid support). After deprotection and coupling, the distal end can be covalently bound to an oligomer.

[0062] Another solid support is a microelectronic chip as described, for example, in U.S. patents No. 6,051,380 to Sosnowski *et al.*, No. 6,129,828 to Sheldon, III *et al.*, and No. 6,017,696 to Heller. This type of chip exploits electronically accelerated hybridization conducted under very low salt conditions, which avoids problems, with DNA conformation and secondary structure, associated with other hybridization methods. Immobilized polynucleotides can be moved electronically to specific sites on the microchip, and then hybridization is determined. In the electronic mediated approach, use of an electronically mediated, active hybridization process to move and concentrate target DNA molecules reduces the time to detect hybridization from hours (as in conventional, passive methodology) to minutes.

IDT Exhibit 1020.8855

[0063] The polynucleotides can be attached, via conventional technology, to a solid support in the form of a microarray, also known as a "DNA chip," a "DNA microarray," a "gene array", a "gene chip," and a "genome chip," or in the form of a macroarray. In this context, an array is an orderly arrangement of samples that enables the matching of known and unknown DNA samples, founded on base-pairing rules, and the automation of identifying the unknowns.

[0064] "Microarray" and "macroarray" are relative terms, distinguished from each other in terms of the diameter of the sample spots involved. Thus, microarray sample spots typically are about 200 microns in diameter or less, making it possible to array thousands of sample spots on a single chip. These arrays can be prepared by hand but, preferably, are made using specialized robotics and read by means of specialized imaging equipment, particularly with spots with diameters in the lower-end range. By contrast, macroarray sample spots typically are greater than 200 microns in diameter, making them suitable for imaging by gel and blot scanners. Macroarrays may be prepared by hand using standard microplates or standard blotting membranes.

[0065] A microarray containing immobilized polynucleotides can be prepared by a number of well-known approaches including, for example, light-directed methods, such as VLSIPS™ (see U.S. patent No. 5,143,854), mechanical methods such as described in PCT No. 92/10183 or U.S. patent No. 5,384,261, bead-based methods described, for example, in PCT US/93/04145, and pin-based methods as detailed in U.S. patent No. 5,288,514, *inter alia*. Patent No. 5,556,752 to Lockhart describes the preparation of a library of different, double-stranded polynucleotides as a microarray, using the VLSIPS,™ and this approach also is suitable for preparing a library of polynucleotides in a microarray. Flow channel techniques, as described U.S. patents No. 5,677,195 and No. 5,384,261, also can be employed to prepare a microarray chip that has a variety of different immobilized polynucleotides. In this case, certain activated regions of the substrate are mechanically separated from other regions when the polynucleotides are delivered through a flow channel to the support. As noted, the Lockhart '752 patent describes flow channel methodology in some detail, including the use of protective coating-wetting facilitators to enhance the directed channeling of liquids though designated flow paths.

IDT Exhibit 1020.8856

[0066] Spotting methods also can be used to prepare a microarray biochip with a variety of immobilized polynucleotides. In this case, reactants are delivered by directly depositing relatively small quantities in selected regions of the support. In some steps, of course, the entire support surface can be sprayed or otherwise coated with a particular solution. In particular formats, a dispenser moves from region to region, depositing only as much polynucleotides or other reagent as necessary at each stop. Typical dispensers include a micropipette, nanopippette, ink-jet type cartridge or pin to deliver the polynucleotide containing solution or other fluid to the support and, optionally, a robotic system to control the position of these delivery devices with respect to the support. In other formats, the dispenser includes a series of tubes or multiple well trays, a manifold, and an array of delivery devices so that various reagents can be delivered to the reaction regions simultaneously.

[0067] Spotting methods are well known and include, for example, those described in U.S. patents No. 5,288,514, No. 5,312,233 and No. 6,024,138. In some cases, a combination of flowing channel and "spotting" on predefined regions of the support also can be used to prepare microarray biochips with immobilized polynucleotides.

[0068] Information and examples of constructing microarrays are readily available in the art, such as in Cheung et al., *Nature Genetics,* Vol. 21 (suppl.), pp. 15-19 (1999); Lipshutz et al., *Nature Genetics,* Vol. 21 (suppl.), pp. 20-24 (1999); Brown & Botstein, *Nature Genetics,* Vol. 21 (suppl.), pp. 33-37 (1999); Shoemaker et al., *Nature,* Vol. 409, pp. 922-27 (2001); Schena et al., *Science* (1995) 467-470; Schena et al., P.N.A.S. U.S.A. (1996) 93: 10614-10616; Pietu et al., *Genome Res.* (June 1996) 6: 492-503; Zhao et al., *Gene* (Apr. 24, 1995) 156: 207-213; Soares, *Curr. Opin. Biotechnol.* (October 1997) 8: 542-546; Raval, J. Pharmacol Toxicol Methods (November 1994) 32: 125-127; Chalifour et al., *Anal. Biochem* (Feb. 1, 1994) 216: 299-304; Stolz & Tuan, *Mol. Biotechnol.* (December 1996) 6: 225-230; Hong et al., *Bioscience Reports* (1982) 2: 907; and McGraw, *Anal. Biochem.* (1984) 143: 298.. Numerous patents related to microarray development include U.S. patents No. 5,082,830, No. 6,110,426, No. 6,004,755, No. 5,445,934, No. 5,532,128, No. 5,556,752, No. 5,242,974, No. 5,384,261, No. 5,405,783, No. 5,412,087, No. 5,424,186, No. 5,429,807, No. 5,436,327, No.

IDT Exhibit 1020.8857

5,472,672, No. 5,527,681, No. 5,529,756, No. 5,545,531, No. 5,554,501, No. 5,561,071, No. 5,571,639, No. 5,593,839, No. 5,599,695, No. 5,624,711, No. 5,658,734 and No. 5,700,637, and PCT application WO 97/27317.

C. Hybridization and Analysis

[0069] Hybridization conditions (either initial hybridization without washing or a combination of initial hybridization and washing) can be selected that allow the DNA fragments to hybridize the complementary immobilized polynucleotides of the solid support and avoid unwanted cross hybridization. See *e.g.*, Sambrook (1989), Chapter 9.47-9.62. In general, DNA annealing is carried out in high ionic strength to maximize the annealing rate and at temperatures about 20-25°C below the melting temperature of the DNA (T_m). Various agents may be used to reduce non-specific binding (*e.g.* Denhardt's solution, BLOTTO, heparin, and denatured salmon sperm DNA). Washing conditions should be as stringent as possible, generally sufficient salt and at a temperature 12-20°C below the T_m. Washing conditions may be determined empirically in preliminary experiments. Other approaches may be used and may readily be determined by one of ordinary skill in the art.

[0070] The amount of amplified DNA that hybridizes to the solid support is compared to a control to determine whether the extent of methylation of one or more regions of DNA in a test sample is different from the control sample. In one embodiment, the control can be an actual sample of DNA such as genomic DNA from a normal healthy tissue that is processed in parallel with the test sample. Such control sample may be obtained from the same type of cell, tissue or organ as that from which the test sample was obtained. Alternatively, a control sample may contain one or more selected DNA fragments, which may be prepared so that it is fully methylated, fully unmethylated, or partially methylated as desired. The control also can be a theoretical control which represents a historical value derived from prior experimentation.

[0071] In a preferred embodiment, test and control DNA fragments are separately labeled with Cy-3 or Cy-5 dye. The labeled test and control fragments are then mixed together and hybridized to the solid support. After washing, the slide is scanned with an array scanner either successively or simultaneously at the excitation wavelength of 635 nm

IDT Exhibit 1020.8858

(Cy5) and 532 nm (Cy3). The images for each wavelength are recorded, pseudo-colors are applied to each image (red for Cy5 and green for Cy3) and the images are overlaid to show the relative density differences. The densities at both wavelength for each spot are also quantified and analyzed. The extent of methylation ("methylation status") of DNA in a test sample relative to a control sample is then determined either visually by the color of each spot or through the measured density for each scanned channel.

[0072] Due to the qualitative nature of the present method achieved through the competitive amplification and hybridization, a very high degree of sensitivity, enabling the detection of subtle changes in methylation level for each region is attainable. This is highly desirable because tissue samples often contain mixed cell populations.

Applications of the Method

[0073] The present invention provides a high throughput methodology for detecting methylation changes at specific regions throughout the genome. High throughput can be accomplished using well known systems involving robotic control, automated liquid handling, automatic spotting, plate and slide loaders, pin arrays (96, 384, 1536 etc.), and the like (see e.g., automated array systems from Cartesian Technologies Inc.). This will allow tens of thousands of such regions, both CpG islands and non-island regions, to be tested simultaneously so that any methylation change in the genome can be detected. Also, hundreds or more samples from different individuals can be analyzed, and changes in methylation that are linked with a particular disease can later be exploited as a diagnostic screen for the disease. Other applications for the present invention are detailed below.

[0074] Diagnosis: The information thereby gleaned from using the methods herein informs diagnoses and prognoses that reflects, in whole or in part, the gene-silencing pattern characterizing a particular disease. That is, the methylation determined via the present invention offers a disease-state indicator of exceptional selectivity, specificity, and sensitivity. Furthermore, aberrant CpG island-hypermethylation, which occurs at high frequency in tumors, can yield diagnostic information as well. Determining patient's genome methylation by means of the present invention opens the way, in a cost-effective manner, for an unprecedented early warning diagnosis.

IDT Exhibit 1020.8859

Inappropriate methylation changes in CpG islands is one of the earliest known markers in the development of many cancers and direct detection of these molecular aberrations using the methods disclosed herein provides an extraordinary opportunity for unprecedented early stage molecular diagnosis of cancer. In addition, epigenetic changes have been implicated in several other important diseases. These include atherosclerosis (Post *et al.*, 1999), Angelman syndrome (Lalande *et al.*, 1999), Duchenne muscular dystrophy (Yoshioka *et al.*, 1998) and ICF syndrome (Kondo *et al.*, 2000), to name a few.

[0075] Enabling technology for improved clinical trials: The methods disclosed herein can be used to determine which individuals are afflicted with methylation dependent cancers. This can increase the success of efficacy studies in clinical trials of drugs targeting the basis for the methylation difference in the cancer. Drug candidates that effect methylation status represent the next generation of non-cytotoxic cancer therapies. In addition, since methylation status of a gene often correlates with its expression, methylation profile of test subjects can also be used in patient selection in clinical trials of other drugs, including cancer drugs.

[0076] Personalized Medicine: With the high cost of cancer therapies and the wide variety of cancer types, it is important that tests be developed to determine which patients will respond to which therapies. As the new generation of methylation dependent cancer therapies advances, the assays of the present invention will be important to determine which cancer patients are afflicted with methylation defects. Such data can help the oncologist's therapy decision process, determining patient suitability for a methylation-based drug regimen.

[0077] Discovery: Detection of inappropriate methylation in CpG islands acts as an indicator of which genes are involved in the development of cancer. Furthermore, only a small fraction (<3%) of all CpG islands have been investigated as to their role in cancer. Discoveries of new gene silencing events in cancer using the methods described herein will provide critical information for the initiation of new drug discovery and optimization strategies.

[0078] Toxicology: The cost of bringing drug candidates through clinical trials that eventually fail due to toxicological problems is enormous. Thus, there is a great need

IDT Exhibit 1020.8860

for methylation detection methods to "weed-out" drugs with toxicology problems at an early (pre-clinical) stage. Applying the methods of the present invention in a high throughput screening format will be helpful in determining if a particular drug impacts the methylation status of cells or tissues. Such screening would lower the likelihood that candidate drugs with mutagentic or epigenetic-based toxicity will proceed inappropriately to clinical trials.

EXAMPLES

Example 1: Direct Microarray Screening of Differentially Methylated Sites ("DMS^2") for a Mouse Transgene

[0079] This example demonstrates the use of an embodiment of the present invention for detecting the extent of methylation of a transgene in mouse DNA. The method involves a DNA amplification step and is referred to as DMS^2 (direct microarray-based screening of differentially methylated sites).

[0080] Oligonucleotides of 50-60 nucleotides in length were synthesized for hybridizing to selected CpG islands in mouse genes and for hybridizing to a mouse transgene known as HRD (Engler et al., (1991) *Cell* 65(6):939-47). HRD provides a useful control DNA in the assay because the transgene is highly methylated in one mouse strain (C57BL/6J, abbreviated as B) and unmethylated in another (DBA/2J, abbreviated as D), and HRD is integrated in the mouse genome in amounts comparable with other mouse genes. Figure 5A is a Southern blot that demonstrates that the transgene is methylated in strain B but not in strain D.

[0081] DNA (1 μg) from each HRD transgenic mouse strain was separately digested with five units of HpaII in 50 μl for 6 hours at 37°C. The digestion mixture was heat inactivated at 65°C for 15 minutes, precipitated with ethanol, and resuspended in a volume of 20.5 μl. For ligation, two microliters of 100 μM of oligonucleotide-linker, 2.3 μl of 10x ligation buffer and 0.2 μl T4 DNA ligase (both from New England Biolabs, Beverly, MA) were added to the samples and the mixture incubated at room temperature for 6-16 hours. The oligonucleotide-linker is shown in Figure 4A.

IDT Exhibit 1020.8861

[0082] The ligated DNA fragments were purified using a Qiagen MinElute PCR purification Kit based on silica gel-membrane spin columns (Qiagen, Valencia, CA). Samples were processed in accordance with the manufacturer's instructions except that a 30% guanidine hydrochloride wash step was added prior to the regular wash. The product was eluted in 25 μl (15 μl + 10 μl) elution buffer provided in the kit.

[0083] For PCR, 2.5 μl of the eluted DNA was amplified with AmpliTaq Gold® in 50 μl reaction with Gold buffer (Applied Biosystems, Foster City, CA), 100 μM dNTP, 1.2 M betaine, 1 μM Cyanine dye-labeled amplification primer (Cy-3 for the methylated mouse strain DNA and Cy-5 for the unmethylated mouse strain DNA) and 1 Unit of DNA polymerase. The primer sequence was as follows:

Dye-5'-CTGACGATGAGTCCTGAGTCGG-3' (SEQ ID NO:6)

The cycle condition was 95°C for 6.5 minutes followed by 35 cycles of 94°C for 30 sec, 55°C for 20 seconds, and 72 °C for 1 minute. The amplified products (*i.e.*, amplicons) were purified with the Qiagen PCR Purification kit and eluted in 10 μl of elution buffer.

[0084] A second PCR reaction using transgene primers, followed by agarose gel electrophoresis, was performed to determine if the amplicon contained the expected fragment predicted by the assay design. The result in Figure 5 B shows the expected strong amplification of transgene DNA fragment (see arrow) from the processed DNA of the unmethylated transgene mouse strain (lane D) while no amplification of a transgene DNA fragment was seen from the processed DNA of the methylated transgene mouse strain (lane B). The presence of a Cy-3 or Cy-5 label on the DNA amplicon did not affect the results.

[0085] Microarrays were prepared by spotting polynucleotides in triplicate onto a glass slide using a mechanical arrayer (Affymetrix, Santa Clara, CA). The Cy-3 and Cy-5 labeled DNA amplicons were mixed together and hybridized to the array in 6xSSC at 55°C for several hours. The washing conditions were 2xSSC and 0.03% SDS; 2x

IDT Exhibit 1020.8862

SSC;1x SSC; 0.2x SSC, all at room temperature. Arrays were scanned at 635 nm for the Cy-5 and 532 nm for Cy-3.

[0086] As shown in Figure 5C, the oligonucleotide probe for the transgene detected higher Cy5 (from strain D) hybridization signal than Cy3 (from strain B) signal consistently in the triplet, indicating the successful detection of the methylation difference for the transgene. The control probe, as expected, detected a similar level of hybridization signal in both Cy5 and Cy3 channels consistent with the lack of methylation in the control gene in the two strains.

Example 2: Screening of Differentially Methylated Sites Using a cDNA Microarray

[0087] This example demonstrates the use of an embodiment of the present invention for detecting the extent of methylation of mouse genes using a high-density cDNA microarray. The mixture of Cy-3 and Cy-5 labeled DNA amplicons were prepared as described in Example 1 and hybridized to the M9K mouse cDNA microarray, which contains approximately 9000 sequence-verified cDNA clones spotted on a microscope slide. The hybridized array was scanned at 532nm and 635 nm as described in Example 1.

[0088] An overlay of the two scans was achieved by pseudo-coloring (green for Cy-3 and red for Cy-5). Four independent hybridizations were performed and the results indicate a high degree of reproducibility. Most of the spots show a yellow color of weak intensity, indicating that the genomic regions for the clones are not differentially methylated (they have similar methylation level) and likely are unmethylated in both samples. Red or green spots were observed for about 50 clones which have a Cy5/Cy3 ratio larger than 3 or smaller than 0.3, indicating a different level of methylation in the genomic regions they represent. Many spots show very weak or no signal, indicating that the regions detected by the immobilized cDNA are either highly methylated in both samples or are not represented in the amplicon (the latter may be due to the fact that the cDNA microarray contained many clones from the 3' part of the genes that may not have CpG dinucleotides).

IDT Exhibit 1020.8863

Example 3: Screening of differentially methylated sites in human breast cancers using a 762-feature CpG island fragment microarray

[0089] In this example, 2 μg of genomic DNA from three human breast tumor cell lines, T47D, MB435, and SKBR3, were digested with HpaII overnight at 37°C. One fourth of the digested DNA was ligated to 10 μl of 100 μM of the unphosphorylated linker shown in Figure 2 in 30 μl reaction at 16°C for 5-16 hours. A 0.2ml PCR tube contained the following mixture: 10 μl of the ligation product, 10 μl of 10x reaction buffer, 10 μl of 2 mM dNTPs, 20 μl of 5M betaine (Sigma-Aldrich, St Louis, MO), 5 μl DMSO, 44.5 μl H2O, and 0.5 μl (2.5 units) of Taq polymerase (Applied Biosystems, Foster City, CA). PCR conditions were as follows: 72°C, 5 minutes, then 94°C, 3 minutes followed by 25 cycles of 94°C, 30 seconds; 60°C, 30 seconds; 72°C, 3 minutes. The amplified samples were stored at 4°C. Amplified products (amplicons) were purified with Qiagen PCR purification kit (Qiagen, Valencia, CA). 2 μg of each amplicon was labeled either with Cy3 (for T47D) or Cy5 (for MB435 and SKBR3) dCTP to generate labeled probes as follows: 2 μg of amplicon (diluted in 21 μl with H_2O) and 20 μl of 2.5x random primer/reaction buffer (125 mM Tris pH 6.8, 12.5 mM MgCl2, 25 mM 2-mercaptoethanol, 750 μg/ml random octamers) are combined in a 0.5 ml eppendorf tube, incubated in boiling water for 5 minutes, placed on ice for 5 minutes. The samples were centrifuged briefly and the liquid fraction placed on ice, and the following was added: 5 μl 10x dNTP mix (1.2 mM each of dATP, dGTP and dTTP, 0.6 mM dCTP, 10 mM Tris pH8.0, 1 mM EDTA); 3μl Cy5 or Cy3 dCTP (Amersham Pharmacia, 1mM), and 1 μl klenow fragment (New England Biolabs). The mixture was incubated at 37°C for 1.5 hours. The labeled probes were purified with a Microcon 30 filter (Amicon/Millipore).

[0090] The purified probes were mixed to form two mixtures: T47D/Cy3 with MB435/Cy5 and T47D/Cy3 with SKBR3/Cy5. The microarray composing of 762 CpG island fragments was manufactured as described in Example 1. The dye-labeled mixtures

IDT Exhibit 1020.8864

were hybridized to the microarray in a 65°C water bath for 16 hours and washed and scanned as described (Example 1).

[0091] The results showed 25 CpG island fragments with significant differential hybridization signals (a ratio of Cy5/Cy3 greater than 3 or smaller than 0.3) indicating differential methylation between the cell lines. Bisulfite-based quantitative methylation assay COBRA (Xiong & Laird, Nucleic Acids Res. 1997 Jun 15;25(12):2532-2534) was used to verify the detected differences in methylation among the samples. The COBRA assays confirmed differential methylation for six of nine tested CpG islands fragments.

****

[0092] The invention thus has been disclosed broadly and illustrated in reference to representative embodiments described above. Those skilled in the art will recognize that various modifications can be made to the present invention without departing from the spirit and scope thereof.

IDT Exhibit 1020.8865

WHAT IS CLAIMED IS:

1. A method for detecting whether the extent of methylation of one or more regions of DNA in a test sample is different from that of a control, comprising:

generating DNA fragments from at least one test sample of DNA by cleaving methylation sites in the DNA that are not methylated while sparing methylation sites in the DNA that are methylated;

ligating an oligonucleotide-linker to the ends of the DNA fragments;

amplifying the DNA fragments by initiating DNA amplification at the oligonucleotide linker with a linker-primer;

hybridizing the amplified DNA to one or more polynucleotides immobilized on a solid support, said polynucleotides being complementary to one or more regions of DNA in the test sample; and

comparing the amount of amplified DNA from the test sample that hybridizes to the immobilized polynucleotides versus that of a control, thereby detecting whether the extent of methylation of the one or more regions of DNA is different between the test sample and the control.

2. The method of claim 1, wherein said control is one or more control samples of DNA processed the same as the one or more test samples of DNA.

3. The method of claim 1, wherein said DNA fragments are generated by cleaving with a methylation-sensitive agent.

4. The method of claim 3, wherein said methylation-sensitive agent is a methylation sensitive restriction enzyme.

5. The method of claim 4, wherein said restriction enzyme cleaves a recognition sequence selected from the group consisting of CCGG, CCGC, GCGC, ACGT, CGGCCG, GCCGGC, GGCGCC, CCCGGG, CGCG, ATCGTA, TTCGAA, GTCGAG, and CTCGAG.

IDT Exhibit 1020.8866

6. The method of claim 4, wherein said methylation-sensitive restriction enzyme cleaves a recognition sequence selected from the group consisting of: CGGCCG, GCCGGC, GGCGCC, CCCGGG, CCGG, CCGC, GCGC, and ACGT.

7. The method of claim 1 wherein said oligonucleotide linker is phosphorylated.

8. The method of claim 1 wherein said oligonucleotide linker is not phosphorylated.

9. The method of claim 8 wherein said non phosphorylated oligonucleotide linker is double stranded and wherein only one strand of the linker is ligated to the fragment.

10. The method of claim 9 wherein a polymerase chain mediated extension of the ligated fragment precedes the step of amplification and wherein one chain of the oligonucleotide linker is the linker primer.

11. The method of claim 1, further comprising the step of cleaving the DNA with at least one methylation-insensitive agent.

12. The method of claim 11, wherein said methylation-insensitive agent is a methylation-insensitive restriction enzyme.

13. The method of claim 12, wherein said methylation insensitive restriction enzyme is selected from the group consisting of EcoRI, ApoI, Tsp509I, MseI, BfaI, Csp6I, NlaIII, DpnII, and CviJI.

14. The method of claim 11, wherein said cleavage by the methylation-sensitive agent and the methylation-insensitive agent occurs simultaneously in the same mixture.

15. The method of claim 11, further comprising the step of ligating a non-phosphorylated oligonucleotide-linker to the end of each DNA fragment generated by the at least one methylation-insensitive agent.

IDT Exhibit 1020.8867

16. The method of claim 15, wherein said oligonucleotide-linkers that ligate to sites cleaved by the methylation-insensitive agent are distinct from the oligonucleotide-linkers that ligate to the sites cleaved by the methylation-sensitive agent.

17. The method of claim 1, wherein said oligonucleotide-linker is a universal oligonucleotide-linker.

18. The method of claim 1, wherein said step of comparing the amount of amplified DNA that hybridizes is performed by measuring a detectable moiety associated with the amplified DNA.

19. The method of claim 18, wherein said detectable moiety is a cyanine dye.

20. The method of claims 18, wherein said detectable moiety is a labeled dendrimer.

21. The method of claim 18, wherein said detectable moiety is associated with the amplification primer prior to its incorporation into amplified product.

22. The method of claims 18, wherein said detectable moiety is attached to the amplified DNA following amplification.

23. The method of claim 18, wherein said detectable moiety is incorporated into the synthesized DNA during amplification.

24. The method of claim 1, wherein said immobilized polynucleotides comprise sequences complementary to one or more CpG islands.

25. The method of claim 1, wherein said immobilized polynucleotides comprise sequences complementary to one or more gene regulatory sequences or an encoding gene sequence.

26. The method of claim 1, wherein said immobilized polynucleotides comprise at least two polynucleotides immobilized as an array.

27. The method of claim 26, wherein said array is a microarray.

IDT Exhibit 1020.8868

28. The method of claim 1, wherein said step of amplifying DNA occurs via the polymerase chain reaction.

29. The method of claim 2, wherein said DNA fragments from the test and control samples are labeled with different detectable moieties that are distinguishable from each other when mixed together.

30. The method of claim 29, wherein said different detectable moieties comprise a cyanine-3 dye and a cyanine-5 dye.

31. The method of claim 29, wherein said test and control samples are mixed together before the step of hybridization.

32. A method for detecting whether the extent of methylation of one or more regions of DNA in a test sample is different from that of a control, comprising:

generating DNA fragments from at least one test sample of DNA by cleaving methylation sites in the DNA that are not methylated while sparing methylation sites in the DNA that are methylated;

ligating an oligonucleotide-linker to the ends of the DNA fragments;

selecting a particular size range of the DNA fragments;

labeling the DNA fragments with a detectable moiety;

hybridizing the DNA fragments to one or more polynucleotides immobilized on a solid support, said polynucleotides being complementary to one or more regions of DNA in the test sample; and

comparing the amount of the detectable moiety associated with the immobilized polynucleotides for the test sample versus a control, thereby detecting whether the extent of methylation of the one or more regions of DNA is different between the test sample and the control.

33. The method of claim 32, wherein said control is one or more control samples of DNA processed the same as the one or more test samples of DNA.

IDT Exhibit 1020.8869

34. The method of claim 32, wherein said DNA fragments are generated by cleaving with a methylation-sensitive agent.

35. The method of claim 34, wherein said methylation-sensitive agent is a methylation sensitive restriction enzyme.

36. The method of claim 35, wherein said restriction enzyme cleaves a recognition sequence selected from the group consisting of CCGG, CCGC, GCGC, ACGT, CGGCCG, GCCGGC, GGCGCC, CCCGGG, CGCG, ATCGTA, TTCGAA, GTCGAG, and CTCGAG.

37. The method of claim 35, wherein said methylation-sensitive restriction enzyme cleaves a recognition sequence selected from the group consisting of: CGGCCG, GCCGGC, GGCGCC, CCCGGG, CCGG, CCGC, GCGC, and ACGT.

38. The method of claims 32, wherein said detectable moiety is a labeled dendrimer.

39. The method of claim 32, wherein said one or more immobilized polynucleotides comprise sequences complementary to a one or more CpG islands.

40. The method of claim 32, wherein said one or more immobilized polynucleotides comprise sequences complementary to one or more gene regulatory sequences or an encoding gene sequence.

41. The method of claim 32, wherein said one or more immobilized polynucleotides comprise at least two polynucleotides immobilized as an array.

42. The method of claim 41, wherein said array is a microarray.

43. The method of claim 33, wherein said DNA fragments from the test and control samples are labeled with different detectable moieties that are distinguishable from each other when mixed together.

44. The method of claim 43, wherein said different detectable moieties comprise a cyanine-3 dye and a cyanine-5 dye.

IDT Exhibit 1020.8870

45. The method of claim 43, wherein said test and control samples are mixed together before the step of hybridization.

46. The method of claim 32, wherein said DNA is labeled by labeling the oligonucleotide-linker.

47. The method of claim 46, wherein said oligonucleotide-linker comprises a capture sequence used for attaching the detectable label.

48. The method of claim 32, wherein said step of selecting a particular size range of the DNA fragments is conducted after the step of linker ligation.

49. The method of claim 32, wherein said step of selecting a particular size range of the DNA fragments step results in the separation of fragments about 1,500 base pairs in length or less.

50. A method of detecting whether an individual has a disease state associated with an abnormal extent of methylation of one or more regions of DNA in the individual, said method comprising:

providing a test sample of DNA derived from tissue of the individual;

determining the extent of methylation for one or more regions of DNA in the sample from the individual in accordance with the method of claim 1; and

determining abnormal methylation by relating the extent of methylation of the one or more regions of DNA in the test sample to that of a normal healthy control, thereby indicating if the individual has the state of disease.

51. A method of detecting whether an individual has a disease state associated with an abnormal extent of methylation of one or more regions of DNA in the individual, said method comprising:

providing a test sample of DNA derived from tissue of the individual;

determining the extent of methylation for one or more regions of DNA in the DNA of sample from the individual in accordance with the method of claim 32; and

IDT Exhibit 1020.8871

determining abnormal methylation by relating the extent of methylation of the one or more regions of DNA in the test sample to that of a normal healthy control, thereby indicating if the individual has the state of disease.

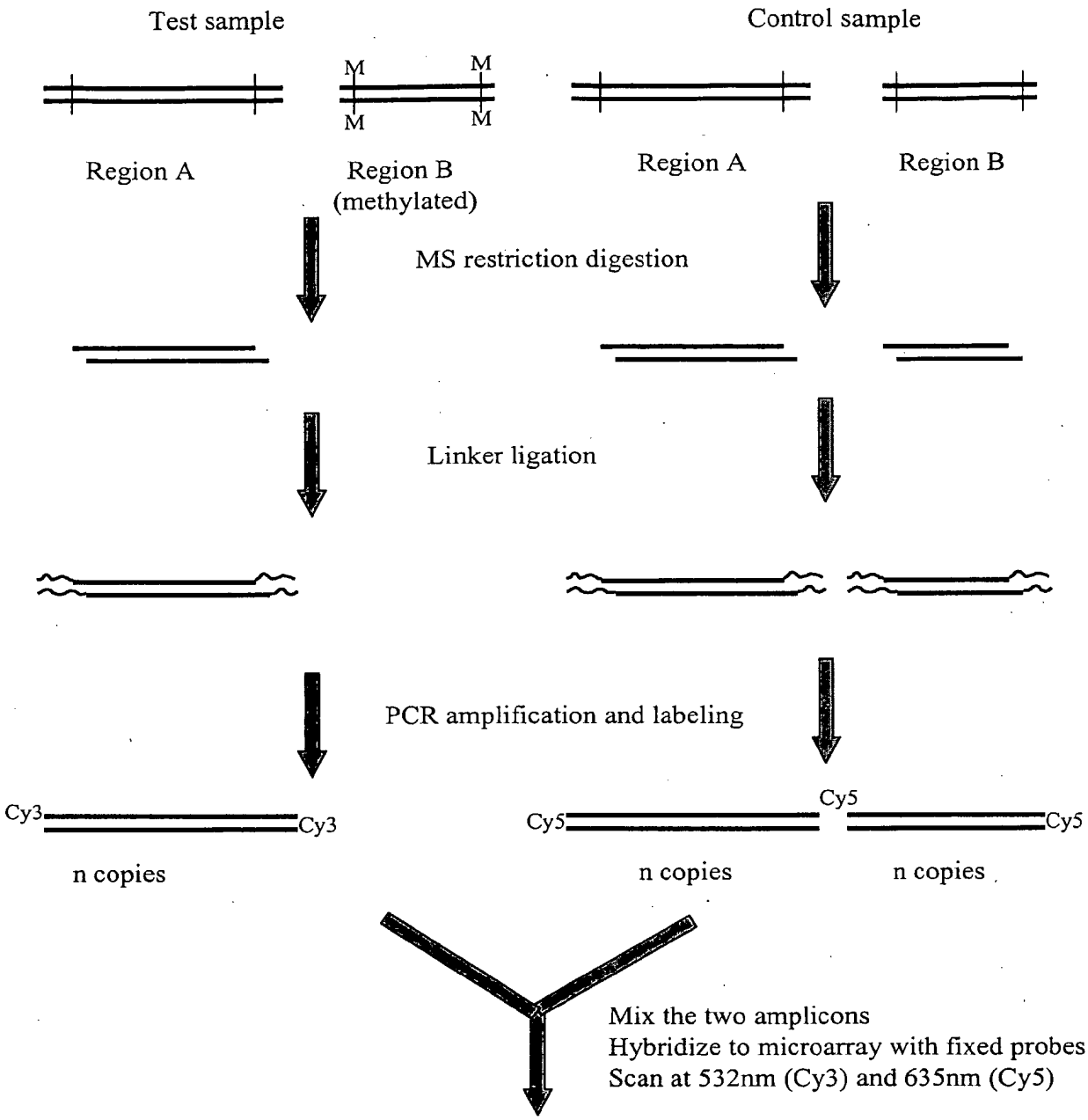
52. A method of establishing whether the extent of methylation in a DNA region of an individual correlates with a disease state, said method comprising:

determining the extent of methylation for said region in the DNA for various individuals with the disease state and individuals without the disease state in accordance with the method of claim 1 and then determining if the extent of methylation of the region correlates with the disease state.

53. A method of establishing whether the extent of methylation in a DNA region of an individual correlates with a disease state, said method comprising:

determining the extent of methylation for said region in the DNA for various individuals with the disease state and individuals without the disease state in accordance with the method of claim 32 and then determining if the extent of methylation of the region correlates with the disease state.

IDT Exhibit 1020.8872



Probe A: signal on both 532nm and 635nm channels
Probe B: signal on 635nm (Cy5) only

Conclusion: region B is hypermethylated in the test sample.

Figure 1

A. Unphosphorylated linker

5′-CTGCTGACGATGAGTCCTGAGT-3′ (SEQ ID No. 1)

3′-CAGGACTCAGC-5′ (SEQ ID No. 2)

B. HpaII (methylation sensitive) digested genomic fragment

5′-pCGGNNNNNNNNNNNNNNNNNN.....

3′-CNNNNNNNNNNNNNNNNNNN.....

C. Linker-ligated fragment and its amplification

5′-CTGCTGACGATGAGTCCTGAGTCGGNNNNNNNNNNNNN..... (SEQ ID No. 3)

(SEQ ID No. 2) 3′-CAGGACTCAGC CNNNNNNNNNNNN.....

nick

72°C, 5 minutes in amplification mixture

5′-CTGCTGACGATGAGTCCTGAGTCGGNNNNNNNNNNNNN..... (SEQ ID No. 3)

3′-**GACGACTGCTACTCAGGACTCAGC**CNNNNNNNNNNNN..... (SEQ ID No. 4)

Newly synthesized region

Top linker strand serves as amplification primer

5′-CTGCTGACGATGAGTCCTGAGT (SEQ ID No. 1)

3′-**GACGACTGCTACTCAGGACTCAGC**CNNNNNNNNNNNN..... (SEQ ID No. 4)

PCR

Exponential amplification

Figure 2

A. Two types of linkers

MS linker with 5'-phosphate at ligation site, allowing both strands to be ligated

MIS linker with no 5'-phosphate, allowing one strand of the linker to be ligated

B. Selective amplification after ligation of the two linkers

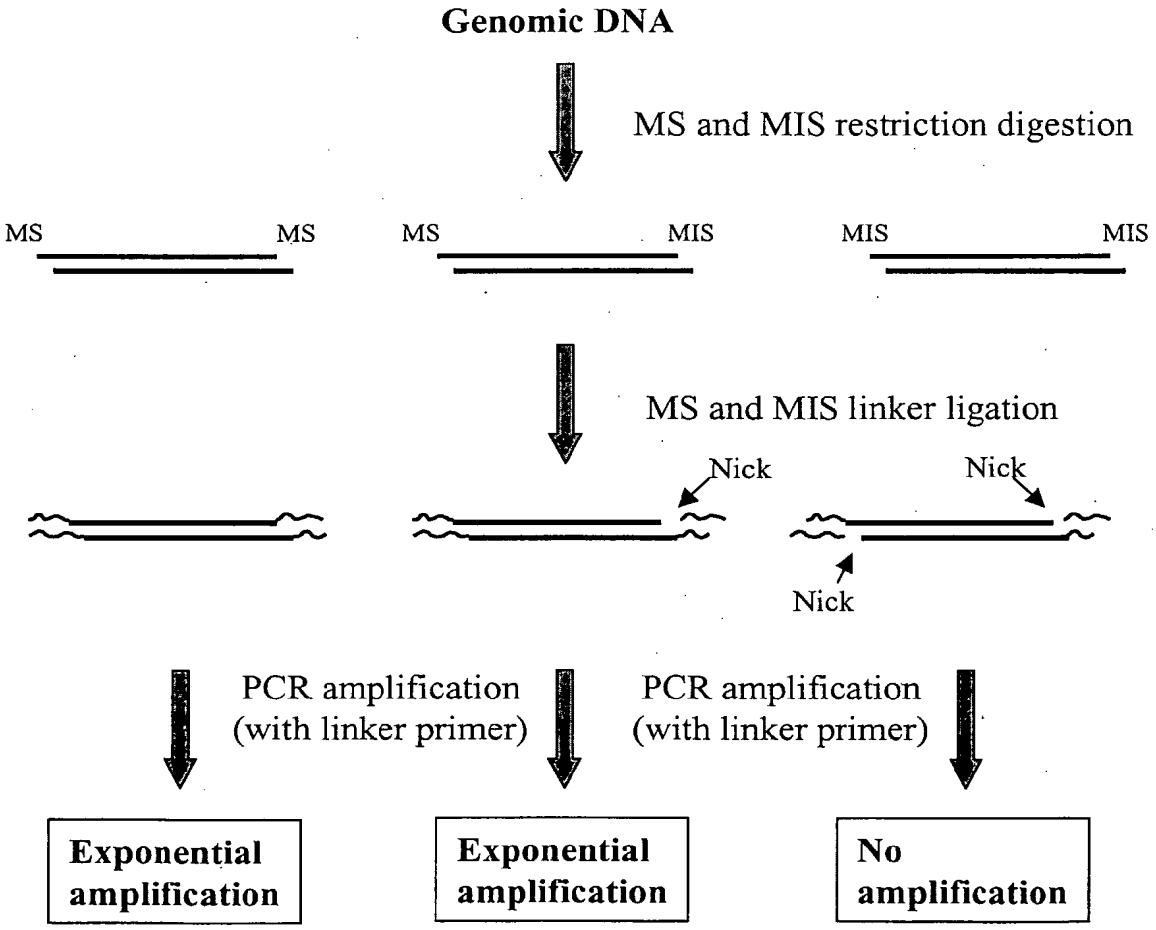


Figure 3

A. Linker (5'-phosphorylated)

```
5'-CTGCTGACGATGAGTCCTGAGT-3'                    (SEQ ID No. 1)
   ||||||||||||||||||||||
3'-GACGACTGCTACTCAGGACTCAGCp-5'                 (SEQ ID No. 5)
```

B. HpaII (methylation sensitive) digested genomic fragment

```
5'-pCGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN...
     ||||||||||||||||||||||||||||||||||||||||||||||||
   3'-CNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN...
```

C. Linker primer

```
5'-CTGCTGACGATGAGTCCTGAGTCGG-3'                  (SEQ ID No. 6)
```

D. Linker dimer (not amplifiable with the linker primer)

```
1. 5'-CTGCTGACGATGAGTCCTGAGTCGACTCAGGACTCATCGTCAGCAG-3'  (SEQ ID No.7)
      ||||||||||||||||||||||||||||||||||||||||||||||||
   3'-GACGACTGCTACTCAGGACTCAGCTGAGTCCTGAGTAGCAGTCGTC-5'  (SEQ ID No.7)

2. 5'-CTGCTGACGATGAGTCCTGAGTCGG                         (SEQ ID No. 6)
      ||||||||||||||||||||||||  <------ mismatched base
   3'-GACGACTGCTACTCAGGACTCAGCTGAGTCCTGAGTAGCAGTCGTC-5'  (SEQ ID No.7)
```

E. Ligated genomic fragment (amplifiable with the linker primer)

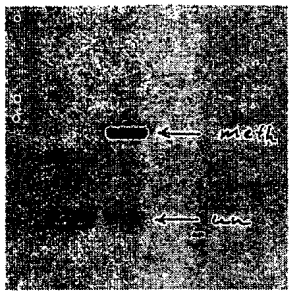
```
1. 5'-CTGCTGACGATGAGTCCTGAGTCGGNNNNNNNNNNNNNNNNNNNNN...  (SEQ ID No.8)
      ||||||||||||||||||||||||||||||||||||||||||||||||
   3'-GACGACTGCTACTCAGGACTCAGCCNNNNNNNNNNNNNNNNNNNNN...  (SEQ ID No.9)


2. 5'-CTGCTGACGATGAGTCCTGAGTCGG                         (SEQ ID No. 6)
      ||||||||||||||||||||||||  <------ matched base
   3'-GACGACTGCTACTCAGGACTCAGCCNNNNNNNNNNNNNNNNNNNNN...  (SEQ ID No.9)
```

Figure 4

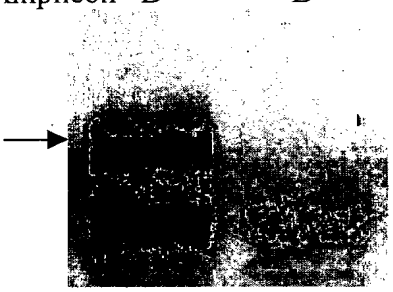
A. Southern blot hybridization

DNA sample from mouse D B



B. Amplified transgene in the amplicon

Amplicon D B



C. Measured Densities of the Hybridization Signals

Oligo I (Transgene)		Oligo II (Control gene)	
Cy3(B)	Cy5(D)	Cy3(B)	Cy5(D)
470	1102	1592	1402
541	1288	1750	1406
500	1082	2401	1808
Differential		No change	

Figure 5

Patent Translate
Powered by EPO and Google

Notice

This translation is machine-generated. It cannot be guaranteed that it is intelligible, accurate, complete, reliable or fit for specific purposes. Critical decisions, such as commercially relevant or financial decisions, should not be based on machine-translation output.

DESCRIPTION CN101565746

The DNA linkage sequencing method encoded by the signal combination with parity check of the present invention relates to a DNA sequencing method of signal combination encoding and belongs to the field of biotechnology. The invention is characterized in that a parity check is added to the coding of connection sequencing of signal combination coding, the parity of the detected coding result is effectively verified, and both the full-space signal and no connection reaction are successfully resolved. The accuracy of sequencing. The present invention adds a parity check marker outside the signal combination coding marker and utilizes both "signal" and "no signal" states of the parity checker and the "signaled" state of the obtained encoded marker state. The type of the marker is subjected to parity check, and after successful verification, a linked DNA sequencing reaction in which signals are combined and coded by different markers "signal" and "no signal" when different coded markers are detected, is not Reduce the sequencing throughput and increase the accuracy of sequencing reactions.

Combinatorial signal encoding DNA linkage sequencing method with parity

Technical field

The DNA linkage sequencing method with signal combination coding and parity check of the present invention relates to a DNA linkage sequencing method using signal combination coding, and combined with a parity check scheme, is a method for implementing DNA sequence analysis and belongs to the field of biotechnology.

IDT Exhibit 1020.8878

technical background

With the deepening of genome research, especially the completion of the Human Genome Project and various models of the biological genome project, great changes have taken place in the way of biological research and medical research. Understanding the differences in life at the genetic level, the occurrence and development of diseases, the interaction between drugs and living organisms, the genetic differences between different species, and the genetic differences among different individuals within the same species are possible. Despite numerous factors that contribute to the development of the disease, genetic sequence differences (including single nucleotide polymorphism, DNA methylation, etc.) are widely considered to be an important intrinsic factor. The occurrence and development of most complex diseases, such as cancer, diabetes, cardiovascular diseases, and mental illnesses, are the result of a combination of many genes and environments. Through the large-scale detection of gene mutations in a large number of genome samples of a particular disease and the comparison with non-affected control genome samples, genotype information related to the disease can be obtained through subsequent mutation of the drug sensitive gene locus. Screening can provide information that is instructive for clinical treatment and medication. The genetic differences between species that are close relatives, or the genetic differences between individuals within the same species, are mainly reflected in genetic sequence differences. Therefore, how to rapidly screen for gene sequence differences in the genome has become one of the main topics in the post-genome era.

Existing methods for detecting gene sequence differences include: traditional Sanger DNA sequencing methods, restriction enzyme length polymorphisms, single strand conformation polymorphisms, and gene chip-based oligonucleotide probe hybridization methods. Among these methods, only the traditional Sanger DNA sequencing method can complete the omni-directional sequence determination of target fragments, and the rest of the methods can only determine a small part of the sequence information. However, the traditional Sanger DNA sequencing method has the disadvantages of low flux, high cost, and long time. The cost of the first human genomic sequence determination is approximately $1 billion. Although this cost has now been reduced to 20 million U.S. dollars, it is still a bottleneck restricting the study of functional genomics. Significantly reducing the cost of DNA sequencing will greatly promote the development of life sciences. To this end, the United States Venter Foundation in 2003 proposed a $1,000 human genome sequencing study. 2004 At the beginning of the year, the U.S. National Institutes of Health invested more than 40 million U.S. dollars in support of research projects for new DNA sequencing technologies, accumulating more than 100 million U.S. dollars. The goal is to develop a $100,000 human genome-wide DNA sequencing technology and eventually reduce it to $1,000.

At present, in addition to the improvement of the existing electrophoresis-based sequencing technology, the sequencing technology being developed can be generally divided into four categories. The first type is the extended sequencing method. The signal-labeled bases are added to the DNA chain that is undergoing polymerization. The second type is the ligation and sequencing method. The signal-labeled oligonucleotide

IDT Exhibit 1020.8879

fragment is added to the ligation reaction that is occurring. The third type is the hybridization sequencing method, which uses the hybridization signal of a set of high-density oligonucleotide microarray chips to identify the sequence of the target gene; the fourth type is the molecular imaging and a series of Sequencing at the single-molecule level; the last kind of technology is to induce DNA molecules to pass through very tiny holes, and in the process to read the bases electronically or optically, it also becomes nanopore sequencing. technology. At present, only extended sequencing methods and ligation sequencing methods are used for whole genome sequencing, and commercialized instruments and reagents have been developed, which greatly increase the efficiency of DNA sequencing and drastically reduce the cost of DNA sequencing. However, the new generation of DNA sequencing still cannot meet the needs of life science research in terms of sequencing cost, throughput, and speed. One of the main reasons is that the signal labeling method for detecting nucleic acid is simple and inefficient. For example, in the DNA ligation sequencing method, limited by the type of the label (eg, fluorophore), only one base can be determined for each ligation reaction, and two or more bases can be determined in one ligation reaction. Base information requires two-dimensional or multidimensional coding of the markers. However, existing coding methods cannot effectively distinguish between the "no signal" state of multiple labels and the "null signal" of unsuccessful ligation reactions, and there is a certain error rate in sequencing reactions.

Summary of the Invention

The purpose of the present invention is to address the existing DNA linkage sequencing technology in the detection of nucleic acid signal labeling method is simple, the efficiency is not high, trying to code through the combination of different label states for two-dimensional coding, synthesizing signal encoding DNA sequencing probes, and Through the introduction of parity markers, "no signal" and "null signal" can be resolved, and the accuracy of the link-sequencing coding scheme can be improved, and the purpose of using the same number of labels to detect more bases or combinations of bases can be established, and rapid establishment can be achieved. Accurate, low-cost and high-throughput sequencing methods.

DNA sequencing is a DNA sequencing method. The basic steps of DNA ligation sequencing are: firstly hybridize a sequencing primer that is complementary to a specific region on the DNA template to be tested; then add one or a group of oligonucleotides to the reactor containing the template to be tested and the sequencing primer. Sequencing probes, each oligonucleotide probe consists of three parts: degenerate bases or non-strictly matched bases, labels, and sequenced bases; degenerate bases or non-strictly matched bases are a group of non-specific The bases that can hybridize to the sequencing template are combined, the labels are used to label the oligonucleotide probes to facilitate detection after the DNA ligation reaction occurs, and the sequenced bases are one or more defined bases for sequencing. The probe selectively reacts with a part of the sequencing template-sequencing primer complex, the corresponding base of the reacting sequencing template and the sequenced base of the sequencing probe are complementary paired, and at the same time, there is some sort of base sequence and label. Correspondence relationship; After the DNA sequencing reaction is performed, the sequencing template-

IDT Exhibit 1020.8880

primer-probe complexes are detected using the corresponding detection means to obtain the corresponding signal information, and then the pre-use signal is used. Correspondence relation setting markers and sequencing bases interpretation information base sequencing templates in the corresponding position.

After a ligation sequencing reaction is completed, the label on the sequencing template-primer-probe complex is removed, the sequencing probe is used as a new ligation starting point for the next ligation sequencing reaction, and the ligation reaction is continued until the desired sequencing is complete. length. The sequencing primer is denaturalized and separated from the template to be tested, and the sequencing primers are then hybridized with the first primer sequence to one or more bases on the template to be tested for a new round of multiple ligation reactions; the denaturation, hybridization, and ligation are repeated after completion. The process completes the determination of base information at all positions.

The technical solution of the present invention: DNA linking sequencing method with signal combination coding of parity check is characterized in that a certain DNA sequencing probe is used for a combination of a set of coding markers and a signal coding marker. Labeling, for a batch of DNA sequencing probes, using a set of different combinations of coding markers that encode signal-bound markers to prepare a set of DNA sequencing probes with a combination of parity-check signals to achieve detection When distinguishing and identifying different DNA sequencing probes:

A. Preparation of a Set of DNA Sequencing Probes Encoded with a Combination of Signals with Parity: First, unlabeled DNA sequencing probes are prepared, and each unlabeled DNA sequencing probe consists of one or more sequenced bases, one or Multiple degenerate bases N or non-strictly matched bases. Sequencing bases are used to determine the base information at corresponding positions in the test DNA by the base pairing pairing rule. The degenerate base N is any of the four bases of A, T, C, and G. After the completion of the preparation of the unlabeled DNA sequencing probes, each unlabeled DNA sequencing probe is labeled with a combination of the states of a plurality of encoding markers, each of which is on the DNA sequencing probe. Whether or not the mark is determined by the state corresponding to the state of the coded mark is combined with the parity mark, and when the number of coded marks is an odd number, the mark is not marked. When the number of marked coded tags is even, the parity tag is marked. A different set of state combinations was used in combination with the parity markers to label different unlabeled DNA sequencing probes to complete the preparation of a set of DNA sequencing probes encoded with a combination of parity signals.

B The sequence of sequencing of the DNA sequencing probe encoded using the above set of signals with parity check is as follows:

a. A sequencing primer was selected and the DNA template to be tested was hybridized according to the

IDT Exhibit 1020.8881

principle of complementary base pairing.

b. A DNA ligation reaction is performed by adding a set of the DNA sequencing probe encoded by the combination of the above-mentioned signals with a parity check, DNA ligase and its reaction system to the reaction system;

c. After the completion of the DNA ligation reaction, the signals of all the markers and the parity markers participating in the state combination are detected, the states of the encoded marker signals are interpreted, the number of coded markers obtaining the signals is calculated, and the parity check markers are used. The signal is subjected to parity check on the value. If the check passes, the type of the DNA sequencing probe that has undergone the ligation reaction is determined based on the combination of all the states of the encoded marker, thereby determining the type of the sequenced base on the sequencing probe and Arrange and finally determine the base or base sequence information at the corresponding position on the tested DNA template;

d. After the detection of the marker signal is completed, the marker on the DNA sequencing probe is removed;

e. Repeat steps b-d above until the DNA sequencing probe reaches or exceeds the end of the DNA template to be tested.

f. The sequencing primer was denatured and separated from the test DNA template and the second sequencing primer was selected:

g. The above steps a-f are repeated until the entire sequence information of the region to be tested of all the unknown DNA templates is determined.

The plurality of labels are combined for labeling unlabeled DNA sequencing probes by simultaneously labeling one or more labels on the same DNA sequencing probe molecule, or labeling the same DNA with different labels respectively The different molecules of the probe are then mixed with the different sequencing probe molecules described above.

The sequencing primer refers to a set of oligonucleotide fragments that can only hybridize to a universal sequence of all DNA templates. A set of sequencing primers has one or several bases that are different from each other in the hybridization region of the DNA template and can be in several rounds. Sequencing reaction completes the

IDT Exhibit 1020.8882

sequencing of the full length of the DNA template; the universal sequence on the sequencing template is added during the preparation of the sequencing template by a ligation reaction or during the amplification process by primers.

The DNA template is used to increase the amount of a target segment of interest in the genome through a DNA amplification technique through a DNA fragment to be measured. The amplification is single-fold, ie, one target fragment is amplified at a time, or multiplex, ie, once. Multiple target fragments are amplified; the immobilization in the DNA template to be measured is fixed on a flat substrate by chemical or physical methods, or is fixed on "96-well plate", "384-well plate" and various modifications. Bead carrier.

The ligation reaction may be a ligation reaction to the 3' end from the 5' end of the sequencing primer, or a ligation reaction of the 3' end to the 5' end.

The label is a fluorescent group that is used, or a change in the chemical and physical properties of the sequencing probe, such as a change in resistance or a change in current.

The detection refers to the properties of the marker. When the laser-excited fluorescent labeling group is used, the laser excitation intensity may be 95%, 70%, 45%, or other different intensities, and the photomultiplier may be separately 95%, 70%, 45% and other detection intervals.

The label is removed by separating the labeling group from the DNA sequencing probe by a chemical or physical method, or restoring the chemical, physical properties of the DNA sequencing probe to its original state. The removal of the marker removes only the marker itself, or at the same time removes one or more bases and related groups that are or are not linked to the marker.

The second sequencing primer is one of a group of sequencing primers except for the sequencing primer that has been used, and does not necessarily correspond to the closest hybridization position on the sequencing template corresponding to the used sequencing primer.

Technical Principle: The purpose of the present invention is to provide a DNA ligation sequencing method with parity-check signal coding, which detects multiple bases simultaneously in one ligation reaction, and improves the accuracy of the experiment through the introduction of parity check marks. The principle of this method is to decide whether or not to connect the conditional parity markers in the sequencing reaction according to the odd and even number of the signal encoding markers in the sequencing reaction, and to encode the markers by

IDT Exhibit 1020.8883

the state check signal of the parity markers at the time of detection. Realizes the resolution of the signalless state and the successful connection reaction, and improves the reliability of the DNA sequencing method of the signal coding cycle.

It is an object of the present invention to provide a DNA ligation sequencing method with a combination of background verification signals that simultaneously detects multiple bases in a ligation reaction and reduces the error rate of the sequencing reaction by the introduction of background markers. The principle of this method is to add a specific background verification marker to the sequencing reaction of binding, effectively resolve the "unsignaled" unconnected two states, and improve the accuracy of the linkage sequencing reaction encoded by the signal combination.

The linked sequencing method of signal combination coding utilizes different labels for binary encoding or multi-pass encoding on the presence or absence (two-dimensional) and intensity difference (multidimensional) of the detected signal. Here we take the presence or absence of a signal as an example. When there are two kinds of markers, two kinds of markers "signal" and "no signal" are arranged and combined, and 2<2>=4 compound states are obtained, namely "markers". 1 with signal, marker 2 with signal, "marker 1 with signal, marker 2 with no signal", "marker 1 no signal, marker 2 with signal", "marker 1 no signal, marker 2 without signal". When the number of markers is n (n is an integer greater than or equal to 2), the composite state of the recording marker is all n-bit binary numbers, and each bit of the binary number records the state of the detected one marker, and "1" is the detection. To the signal, "0" is an undetected signal, and all 2<n> n-bit binary numbers correspond to the 2<n> composite states of the n signal markers. Since there are four common bases that make up DNA, namely adenine (A), guanine (G), thymine (T), and cytosine (C), the need to accurately detect four types of one base requires two There are four kinds of complex states of "signal" and "no signal" for the markers; 4 bases of 4 bases are combined and 16 types are needed, and 4 markers are needed for a total of 2<4>=16 combinations. State to detect; and so on, n base combinations (n is an integer greater than or equal to 2) need a total of 2n kinds of labels, a total of 2<2n> composite signal states to detect, each composite state through a The 2n-bit binary number is uniquely expressed and corresponds to the type of sequencing probe.

However, the simple sequencing method for signal combination coding has the following problems: When the complementary sequencing probe corresponding to the base or combination of bases to be tested does not have a corresponding modification marker, the sequencing probes added to the reaction system are all Without a modified label, we have obtained a set of signal combinations that are all "no signal" by detecting all the labels; at the same time, when the sequencing probes added to the reaction system failed due to a ligation reaction or other reasons, the sequencing reaction failed. By testing we can also get a set of all "no signal" signal combinations. At this point, a set of signal combinations with "no signal" corresponds to two possibilities. The traditional combined encoding sequencing method cannot effectively resolve the two possibilities, and thus there is a certain error rate in the sequencing process. At the same time, sequencing of the linkage is a biochemical reaction, which itself has a certain error rate.

IDT Exhibit 1020.8884

In the present invention, a marker is added as a parity check marker on the basis of the existing signal coding (see Table 1), and the modification of the sequencing probe is involved. The combination of parity markers and sequencing probe types proceeds according to the following rules: When a sequencing probe type is simultaneously combined with zero or even several encoding markers, then the addition of a parity marker to the sequencing probe type Modifications; when a sequencing probe type is combined with an odd number of coding markers simultaneously, the modification of the sequencing probe type by the parity markers is not increased;

The main function of the parity checker is: 1. Verify that the number of odd-even numbers of the "signal" number of the encoded marker after the sequencing reaction is correct in order to improve the accuracy of sequencing; 2. Resolve all encoded markers after the sequencing reaction." No signal" status and connection response failed.

Table 1 Encoding Method with Parity 2n Markers 2<2n> Complex State Detection n Base Oligonucleotide Recipe List

("1" in the table indicates that the probe mixture contains the oligonucleotide, and "0" indicates that the probe mixture does not contain the oligonucleotide probe.)

The beneficial effects of the present invention:

1.

The greatest advantage of the present invention is to perform parity check on the state of the signal-encoded marker by the introduction of a parity check marker, while effectively distinguishing the fact that the encoded marker is all "no signal" and unsuccessful connection reaction occurs. Improve the accuracy of the sequencing reaction, making the signal-encoding loop sequencing method more practical. The combined signal coding by the coded markers allows the simultaneous detection of more bases or combinations of bases in a ligation reaction.

2.

IDT Exhibit 1020.8885

Since the types of labels required to detect a plurality of bases in a single ligation reaction are exponentially decreased compared to the conventional method, it becomes possible to simultaneously detect the types of bases in a single ligation reaction. The fold shortens the time required for the determination of the entire sequence.

3.

With signal combination coding, the number of ligation reactions required to detect sequences of the same length is reduced, and the cost of sequencing is also reduced by a factor of two. At the same time, by signal combination coding, the types of labels required to detect the same number of base types are substantially reduced, reducing the need for detection of marker devices, thereby reducing costs.

Description of the drawings

The present invention will be further described below with reference to the accompanying drawings.

Fig. 1 shows that when the state of the marker is two-dimensional, that is, when the signal is "presence" or "absence", the combination of the two-dimensional states of the two markers is used to detect A, T, C, and G in the sequencing reaction of one connection. The schematic diagram of bases. The solid icon in the figure indicates "signal" and the hollow icon indicates "no signal." 1, When the base to be detected is A, the combined state of the encoded markers corresponding to the sequenced probes whose sequencing bases are U is "Signal 1 has a signal, Marker 2 has a signal", ie, "1, 1". Then get the parity mark (blocks in the figure, solid signal, hollow no signal) signal is "with signal", that is "1", according to the preset on the verification principle, the coded marker state has two 1, even, the corresponding parity mark should be 1, consistent with the status of the actual parity mark; when the base to be measured is C, the encoded marker corresponding to the sequenced probe with the sequenced base G The combined state is "Signal 1 has signal, Marker 2 has no signal", ie, "1,0", and then a parity checkmark is obtained (block in the figure, solid signal, hollow signal) signal " "No signal", ie "0". According to the preset verification principle, the state of the encoded marker has a 1 and is an odd number. The corresponding parity marker should be 0, which is consistent with the state of the actual parity marker. When the base to be detected is G, the sequenced base is U corresponding to the sequenced probe The combined status of the code markers is "Marker 1 no signal, Marker 2 has signal", ie "0, 1", then the parity checker is obtained (block in the figure, solid signal, hollow signal) The signal is "no signal", ie "0". According to the preset verification principle, the state of the encoded marker has a 1 and is an odd number. The corresponding parity marker should be 0, and the actual parity marker. The state is consistent; when the base to be detected is T, the combined state of the encoded markers corresponding to the sequenced probes whose sequencing bases are A is "marker 1 no signal, marker 2 no signal", ie "0," 0", then the parity checkmark (block in the figure, solid signal, no signal in the open signal) is "signal", ie "1". According to the preset verification principle, the

IDT Exhibit 1020.8886

status of the coded flag is 0 of 1, 0, the corresponding parity check flag should be 1, consistent with the status of the actual parity check marker.

FIG. 2 is a workflow diagram of Example 2 and Example 3. FIG. 1, The PCR products amplified by the P1 and P2 primers are immobilized on the surface of the slide; 2. The first sequencing primer I1 is hybridized with the PCR product immobilized on the surface of the slide; 3. The sequencing primers I1 are linked to the sequencing bases and the PCR is to be performed. Oligonucleotide probes with complementary positions; 4, removal of fluorescently modified groups after completion of the assay; 5, ligation of ligation for ligation of the second sequencing probe; 6, removal of fluorescently modified groups after detection is complete; 7, I1 completes all four ligation reactions; 8, the sequencing primer I1 is denatured and separated from the PCR product; 9, the second sequencing primer I2 hybridizes with the PCR product and begins a new round of sequencing reaction.

Figure 3 is a detailed illustration of the first two sequencing ligation reactions performed on the sequencing primer I1 of Example 2. 1, The sequencing primer I1 hybridizes to the PCR product immobilized on the surface of the slide. All 15 bases of I1 are complementary paired with 15 bases of the 5′ -end of the complementary strand of the PCR primer P2. 2. The first ligation reaction is performed, and 40 types of oligonucleotides are obtained. A sequence of oligonucleotides with a base sequence of "5'-AA-3'" in a mixture consisting of a nucleotide probe (32 encoding oligonucleotide probes and 8 parity-check oligonucleotide probes) The body (the remaining position of the oligonucleotide probe is degenerate base N) is connected to I1 hybridized to the PCR product. By detecting all the markers, only the signal of CY7 is obtained, and the state of the encoded marker is all "no signal", and the parity is The check flag is "signaled", conforms to a preset parity rule, and the parity check passes. The oligonucleotide probe formulation table shows that when the coding markers are all "no signal", the 3'th and 4th bases of the 3' end of the test DNA are "5'-AA-3'"; In addition to the fluorescently modified group at the end of the oligonucleotide probe, 4, a second ligation was performed to detect bases 11 and 12 at the 3' end of the DNA to be tested.

Specific implementation

The implementation method of the present invention is divided into two types:

The first is for a certain sequencing probe, using several encoding markers and one parity marker to simultaneously label the same sequencing probe molecule. For a ligation reaction, each label has two states of "signal" and "no signal" respectively when detected, determined by the specific sequencing probe added to the sequencing linkage reaction system, and the preparation contains different coding labels and Background marker sequencing probes are key to the present invention. The sequencing probe added to the reaction system consists of a set of oligonucleotide probes labeled with different labels. Each oligonucleotide probe consists of

IDT Exhibit 1020.8887

one or more sequenced bases (base pairing rule for determination of the corresponding base in the test DNA), one or more degenerate bases N (N is A, Any of the four bases of T, C, and G) or a non-strictly-paired base (such as hypoxanthine, etc.) and a marker having the ability to be detected. The combination of the sequenced bases and the labels was prepared according to the conditions listed in Table 1. When the "sequencing probe type i" determined in Table 1 corresponds to the state of the k labels being "1", the sequencing base was prepared. When the base is the "base-combination type i" sequencing probe molecule, all the labels of the "sequencing probe type i" in the table are all "1"; the 2<2n species listed in Table 1 are synthesized row by row. An oligonucleotide probe, which is a mixture of oligonucleotide probes to be prepared.

The second is to divide the sequencing probe into several parts, each labeled with a label (the label contains a coding label and a parity label). For a single ligation reaction, each of the two labels exhibits two states of "signal" and "no signal" when detected, respectively, determined by the specific sequencing probe added to the sequencing linkage reaction system. Therefore, it is critical to the present invention to prepare sequencing probes containing different coding markers and background markers. The sequencing probes added to the reaction system consist of a set of oligonucleotide probes labeled with different labels. Each oligonucleotide probe consists of one or more sequenced bases (base pairing rule for determination of the corresponding base in the test DNA), one or more degenerate bases N (N is A, Any of the four bases of T, C, and G) or a non-strictly-paired base (such as hypoxanthine, etc.) and a marker having the ability to be detected. The combination of the sequenced bases and the markers was prepared according to the state listed in Table 1, and when the "base combination type i" and the "marker j" identified in Table 1 were "1", the sequencing was performed. Oligonucleotide probes whose bases are "base-combination type i" and whose tags are "marker j"; if the listed state is "0", they are not synthesized, and all the states listed in the synthesis table 1 are "1". The oligonucleotide probe, which is a mixture of oligonucleotide probes to be prepared.

Example 1: DNA Linkage Sequencing Method with Parity-Coupled Signal Combinations, Characterized by Marking a Combination of a Group of Encoded Labels and a Signal-Encoding Label for a Specific DNA Sequencing Probe For a batch of DNA sequencing probes, a set of different combinations of encoding markers that encode signal-bound markers are used for labeling to prepare a set of DNA sequencing probes with a combination of parity-check signals to achieve detection Differentiate and identify different DNA sequencing probes:

A. Preparation of a set of DNA sequencing probes encoded with a combination of parity signals: First, unlabeled DNA sequencing probes are prepared, and each unlabeled DNA sequencing probe consists of one or more sequenced bases, one or more Multiple degenerate bases N or non-strictly matched bases. Sequencing bases are used to determine the base information at corresponding positions in the test DNA by base pairing rules. The degenerate base N is any one of the four bases of A, T, C, and G. After completion of the preparation of the unlabeled DNA sequencing probes, each unlabeled DNA sequencing probe is labeled with a combination of the states of a plurality of encoding markers, each of which is on the DNA sequencing probe. Whether or not the

IDT Exhibit 1020.8888

mark is determined by the state corresponding to the state of the coded mark is combined with the parity mark, and when the number of coded marks is an odd number, the mark is not marked. When the number of marked coded tags is even, the parity tag is marked. A different set of state combinations was used in combination with the parity markers to label different unlabeled DNA sequencing probes to complete the preparation of a set of DNA sequencing probes encoded with a combination of parity signals.

B. The sequencing procedure for the DNA sequencing probe encoded using the above set of signals with parity check is as follows:

a. A sequencing primer was selected and the DNA template to be tested was hybridized according to the principle of complementary base pairing.

b. A DNA ligation reaction is performed by adding a set of the DNA sequencing probe encoded by the combination of the above-mentioned signals with a parity check, DNA ligase and its reaction system to the reaction system;

c. After the completion of the DNA ligation reaction, the signals of all the markers and the parity markers participating in the state combination are detected, the states of the encoded marker signals are interpreted, the number of coded markers obtaining the signals is calculated, and the parity check markers are used. The signal is subjected to parity check on the value. If the check passes, the type of the DNA sequencing probe that has undergone the ligation reaction is determined based on the combination of all the states of the encoded marker, thereby determining the type of the sequenced base on the sequencing probe and Arrange and finally determine the base or base sequence information at the corresponding position on the tested DNA template;

d. After the detection of the marker signal is completed, the marker on the DNA sequencing probe is removed;

e. Repeat steps b-d above until the DNA sequencing probe reaches or exceeds the end of the DNA template to be tested.

f. The sequencing primer was denatured and separated from the DNA template to be tested, and the second sequencing primer was selected.

g. The above steps a-f are repeated until all the sequence information of the region to be tested of all the unknown

IDT Exhibit 1020.8889

DNA templates is determined.

The plurality of labels are combined for labeling unlabeled DNA sequencing probes by simultaneously labeling one or more labels on the same DNA sequencing probe molecule or individually labeling the same DNA with different labels. The different molecules of the probe are then mixed with the different sequencing probe molecules described above.

The sequencing primer refers to a group of oligonucleotide fragments that can only hybridize to a universal sequence of all DNA templates. A set of sequencing primers has one or several bases that are different from each other in the hybridization region of the DNA template, and can be in several rounds. Sequencing reaction completes the sequencing of the full length of the DNA template; the universal sequence on the sequencing template is added during the preparation of the sequencing template by a ligation reaction or during the amplification process by primers.

The DNA template is used to increase the amount of a target segment of interest in the genome through a DNA amplification technique through a DNA fragment to be measured. The amplification is single-fold, ie, one target fragment is amplified at a time, or multiplex, ie, once. Multiple target fragments are amplified; the immobilization in the DNA template to be measured is fixed on a flat substrate by chemical or physical methods, or is fixed on "96-well plate", "384-well plate" and various modifications. Bead carrier.

The ligation reaction may be a ligation reaction to the 3' end from the 5' end of the sequencing primer, or a ligation reaction of the 3' end to the 5' end.

The label is a fluorescent group that is used, or a change in the chemical and physical properties of the sequencing probe, such as resistance changes, current changes.

The detection refers to the properties of the marker. When the laser-excited fluorescent labeling group is used, the laser excitation intensity may be 95%, 70%, 45%, or other different intensities, and the photomultiplier may be separately 95%, 70%, 45% and other detection intervals.

The label is removed by separating the labeling group from the DNA sequencing probe by a chemical or physical method, or by restoring the chemical, physical properties of the DNA sequencing probe to its original state. The removal of the marker removes only the marker itself, or at the same time removes one or more bases and related groups that are or are not linked to the marker.

IDT Exhibit 1020.8890

The second sequencing primer is one of a group of sequencing primers except for the sequencing primer that has been used, and does not necessarily correspond to the closest hybridization position on the sequencing template corresponding to the used sequencing primer.

Example 2 Four-Color Fluorescent Complex Coding Method with Parity (Respective Labeling Scheme) A total of 32 base pair sequences of human RP11-354P11 clone 45332-45363 were detected on human chromosome 17.

Human whole genome DNA was extracted and the target fragment was amplified by PCR. The primers were P1 and P2 (see Table 2 for sequence). The 5' end of primer P1 was modified with hydroxyl. After the amplified PCR product was immobilized on the surface of the aldehyde-modified slide, the slide was heated to 95° C to detach one strand of the double-stranded PCR product from the immobilization strand, and the sequencing reaction was then started.

Table 2 Related Oligonucleotide Sequence Information in Example 2

The hybridization buffer and the first sequencing primer I1 (sequences are shown in Table 2) were added to the surface of the slides, and hybridization was repeated for 20 minutes at 37° C. Add a mixture of all the 40 sequencing probes listed in Table 3 to the reaction cell, where CY7 is the parity checker, and the remaining four markers are coding markers. The oligonucleotide consisting of the marker CY7 is probed. The needle only appears when the corresponding sequenced base corresponds to 0 or even number of encoded oligonucleotide probes, but when the sequenced base corresponds to an odd number of encoding oligonucleotide probes, the sequencing base is not synthesized. Oligonucleotide probe consisting of base and CY7. The ligation reaction was performed using T4 nucleic acid ligase, and the reaction conditions were 25° C for 30 minutes. After the connection was completed, the slides were scanned using a laser confocal microscope. The scans were performed using wavelengths corresponding to CY7, CY3, CY5, TXR, and FTC, and the corresponding base information was read based on the obtained fluorescence signals (Fig. 2).

Table 3 Detailed list of sequencing probe mixtures

In the first ligation reaction, the I1 primer hybridizes to the "5'-AAGAACTCCTGGAACACTG-3'" sequence on

IDT Exhibit 1020.8891

the P2 complementary strand, and after the addition of the oligonucleotide probe mixture, the oligonucleotide probe mixture is sequenced in bases. The oligonucleotides and the test oligonucleotides are located on the 3rd and 4th bases downstream of the 3' end of the I1 primer (which is also the 3rd and 4th bases of the 3' end of the test DNA). These oligonucleotide probes are to be tested. The DNA hybridizes and undergoes a ligation reaction with the primer I1. The 3rd and 4th bases downstream of the 3' end of the I1 primer are "5'-AA-3'" and the reverse complementary base is "5'-TT-3'". Only one oligonucleotide probe based on "5'-TT-3'" is "5'-NNTTNNNN-3'-CY7", and a ligation reaction occurs with the DNA template to be tested and the primer I1 (Figure 3). -2). When performing signal detection, the composite state type of the coded marker is first detected as "CY3 no signal, CY5 no signal, TXR no signal, FTC no signal", there are 0 kinds of markers with signals, and the parity is detected at the same time The status of the marker is "signal". When the encoded marker has zero or even number of markers, "there is a signal", the parity marker is a "signal" parity check rule, and the parity passes. Then, through the look-up table, when the status of the encoded markers is "no signal", the inverted complementary strand information of the 3rd and 4th bases of the tested DNA is "5'-TT-3'", Thus, the 3rd and 4th bases at the 3' end of the test DNA can be obtained as "5'-AA-3'".

After the second round of detection was completed, the fluorescent group at the 3' end of the oligonucleotide probe on the linker and primer I1 was chemically excised to obtain a free hydroxyl group at the 3' end for a re-ligation reaction (Fig. 3-3). A second round of ligation reaction was performed by adding the entire acylic acid probe mixture listed in Table 3 to the reaction cell. The ligation reaction was to detect the 11th and 12th bases of the test DNA located downstream from the 3' end of the I1 primer. "5'-CC-3'" at the 3' end of the DNA at bases 11 and 12). At this time, the three sequenced bases were "5'-GG-3'" oligonucleotide probes "5"-NNGGNNNN-3'-CY7", "5'-NNGGNNNN-3'-CY5", "5'-NNGGNNNN-3'-FTC" occurred a ligation reaction, and fluorescence detection detected CY7, CY5, and FTC at the same time. The verification of the parity mark was successful, and the read and reverse complementation was then performed to obtain bases 11 and 12 at the 3' end of the DNA to be tested as "5'-CC-3'" (FIGS. 3-4). The 3rd and 4th round ligation reactions of the I1 primers were performed again after chemically excising the oligo-terminal fluorophores of the ligated DNA, and the 19'th and 20th bases of the 3' end of the DNA to be tested and 27, 28 were measured, respectively. No. Base information.

After completing the four rounds of ligation reaction, the sequencing primer I1 was separated from the test DNA at 95° C., a hybridization buffer and a second sequencing primer I2 (sequences are shown in Table 2) were added to the surface of the slide, and hybridization was repeated at 37° C. for 20 minutes.. Then four rounds of ligation reactions were started to detect the bases of the four groups of 5, 6, 13, 14, 14 and 21, and 29 and 30 at the 3' end of the DNA to be detected, respectively.

Similarly, four bases No. 7, 8, No. 15, No. 16, No. 23, No. 24, no No. 31 and No. 32 of the 3′ end of the DNA to be detected were detected by sequencing primer I3, and 3 of the test DNA was detected by sequencing primer I4. 'Terminal Nos. 1, 2, 9, 10, 17, 18, and 25, 26 four bases.

IDT Exhibit 1020.8892

By splicing the obtained sequence information, sequence information of all 32 bases of the test DNA can be obtained.

Example 3 Four-Color Fluorescence Complex Coding Method with Parity (Common Labeling Scheme) A total of 32 base pair sequences of human RP11-354P11 clone 45332-45363 were detected on chromosome 17;

Table 4 Related Oligonucleotide Sequence Information in Example 3

Human whole genome DNA was extracted and the target fragment was amplified by PCR. The primers were P1 and P2 (see Table 4 for sequence). The 5' end of primer P1 was modified by hydroxyl group. After the amplified PCR product was immobilized on the surface of the aldehyde-modified slide, the slide was heated to 95° C to detach one strand of the double-stranded PCR product from the immobilization strand, and the sequencing reaction was then started.

The hybridization buffer and the first sequencing primer I1 (sequences are shown in Table 4) were added to the surface of the slides, and hybridization was repeated for 20 minutes at 37° C. A mixture of all 16 oligonucleotide probes listed in Table 5 was added to the reaction cell. Each oligonucleotide probe in the table was an oligonucleotide probe, in which CY7 was a parity checker and the remaining four markers were It is a coded marker. One or more signal markers are simultaneously labeled on each oligonucleotide probe molecule. The marker CY7 only participates in labeling sequencing probes that have been labeled with zero or even several signal encoding markers, and sequencing probes The CY7 group is not labeled when the labeled signal marker is odd. The ligation reaction was performed using T4 nucleic acid ligase and the reaction conditions were 25° C for 30 minutes. After the connection was completed, the slides were scanned with a laser confocal microscope. Scans were performed using wavelengths corresponding to CY7, CY3, CY5, TXR, and FTC, respectively, and corresponding base information was read based on the obtained fluorescence signals (Fig. 2).

In the first ligation reaction, the I1 primer hybridizes to the "5'-AAGAACTCCTGGAACACTG-3'" sequence on the P2 complementary strand, and after the addition of the oligonucleotide probe mixture, the oligonucleotide probe mixture is sequenced in bases. The oligonucleotides and the test oligonucleotides are located on the 3rd and 4th bases downstream of the 3' end of the I1 primer (which is also the 3rd and 4th bases of the 3' end of the test DNA). These oligonucleotide probes are to be tested. The DNA hybridizes and undergoes a ligation reaction with the primer I1. The 3rd and 4th bases downstream of the 3' end of the I1 primer are "5'-AA-3'" and the reverse complementary base is "5'-TT-3'". The sequencing probe " The 5' -NNTTNNNN-3' ' tagged only one of the CY7 fluorophores and connected with the DNA template to be tested and the primer I1. When

IDT Exhibit 1020.8893

performing signal detection, the type of the composite state of the coded label is first detected as "CY3 no signal, CY5 no signal, TXR no signal, FTC no signal", there are 0 kinds of labels with signals, and the parity is detected at the same time The status of the marker is "Signal". When the encoded marker is 0 or the even number of markers is "Signaled", the parity marker is a "signaled" parity check rule, and the parity passes. Then, through the look-up table, when the status of the encoded markers is "no signal", the inverted complementary strand information of the 3rd and 4th bases of the tested DNA is "5'-TT-3'", As a result, the 3rd and 4th bases at the 3' end of the test DNA can be obtained as "5'-AA-3'".

After the second round of detection, the fluorescent group at the 3' end of the oligonucleotide probe on the linker and primer I1 was chemically excised, and a free hydroxyl group was obtained at the 3' end to perform a re-ligation reaction. A second round of ligation reaction was performed by adding the entire acylic acid probe mixture listed in Table 5 to the reaction cell. The ligation reaction was to detect the 11th and 12th bases of the test DNA located downstream from the 3' end of the I1 primer. "5'-CC-3'" at the 3' end of the DNA at bases 11 and 12. At this time, sequencing probe "5'-NNGGNNNN-3'" is labeled with three bases: CY7, CY5, and FTC. group. After the ligation reaction occurs, fluorescence detection is performed simultaneously with detection of CY7, CY5, and FTC. The verification of the parity mark is successful, and the read and reverse complementation are performed to obtain bases 11 and 12 at the 3′ end of the test DNA. "5'-CC-3'". The 3rd and 4th round ligation reactions of the I1 primer were performed again after chemically excising the oligo-terminal fluorophores of the ligated oligonucleotides, and the 19'- and 20-bases of the 3'-end of the DNA to be tested and 27, 28 were measured, respectively. No. Base information.

After completing the four rounds of ligation reaction, the sequencing primer I1 was separated from the test DNA at 95° C., a hybridization buffer and a second sequencing primer I2 (sequences are shown in Table 4) were added to the surface of the slide, and hybridization was repeated at 37° C. for 20 minutes. . Then four rounds of ligation reactions were started to detect the bases of the four groups of 5, 6, 13, 14, 14 and 21, and 29 and 30 at the 3' end of the DNA to be detected, respectively.

Table 5 Detailed list of oligonucleotide probe mixtures

Similarly, four bases No. 7, 8, No. 15, No. 16, No. 23, No. 24, no No. 31 and No. 32 of the 3′ end of the DNA to be tested were detected by sequencing primer I3, and 3 of the test DNA was detected by sequencing primer I4. 'Terminal Nos. 1, 2, 9, 10, 17, 18, and 25, 26 four bases.

IDT Exhibit 1020.8894

By splicing the obtained sequence information, sequence information of all 32 bases of the test DNA can be obtained.

IDT Exhibit 1020.8895

Espacenet

Bibliographic data: WO03012118 (A1) — 2003-02-13

COMPLEXITY MANAGEMENT OF GENOMIC DNA

Inventor(s): DONG SHOULIAN [US] ± (DONG, SHOULIAN)

Applicant(s): AFFYMETRIX INC [US]; DONG SHOULIAN [US] ± (AFFYMETRIX, INC, ; DONG, SHOULIAN)

Classification: - international: ***C07H21/00; C07H21/02; C07H21/04; C12M1/34; C12M3/00; C12P19/34; C12Q1/68; G01N33/00;*** (IPC1-7): C12P19/34
- cooperative: **C12N15/1093**; **C12Q1/6806**; **C12Q1/683**; **C12Q1/6855**; **C12Q1/686**; Y10T436/143333 more

Application number: WO2002US23568 20020723 Global Dossier

Priority number (s): US20010920491 20010731

Also published as: US2003096235 (A1) US2006035274 (A1) US2010292097 (A1) US2011287491 (A1) US2014080724 (A1) more

Abstract of WO03012118 (A1)

The presently claimed invention provides for novel methods and kits for reducing the complexity of a nucleic acid sample by providing non-gel based methods for amplification of a subset of the sequences in a sample. In a preferred embodiment, amplification of a subset can be accomplished by digesting a sample with two or more restriction enzymes and ligating adaptors to the fragments so that only a subset of the fragments can be amplified. The invention further provides for analysis of the above amplified sample by hybridization to an array, which may be specifically designed to interrogate the desired fragments for particular characteristics, such as, for example, the presence or absence of a polymorphism.

IDT Exhibit 1020.8896

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 February 2003 (13.02.2003)

PCT

(10) International Publication Number
WO 03/012118 A1

(51) International Patent Classification[7]: C12P 19/34

(21) International Application Number: PCT/US02/23568

(22) International Filing Date: 23 July 2002 (23.07.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/920,491 31 July 2001 (31.07.2001) US

(71) Applicant *(for all designated States except US)*: **AFFYMETRIX, INC.** [US/US]; 3380 Central Expressway, Santa Clara, CA 95051 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **DONG, Shoulian** [CN/US]; 3350 Mira Vista Circle, San Jose, CA 95132 (US).

(74) Agents: **LIEBESCHUETZ, Joe** et al.; TOWNSEND AND TOWNSEND AND CREW LLP, Two Embarcadero Center, Eighth Floor, San Francisco, CA 94111 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:
— *of inventorship (Rule 4.17(iv)) for US only*

Published:
— *with international search 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 03/012118 A1

(54) Title: COMPLEXITY MANAGEMENT OF GENOMIC DNA

(57) Abstract: The presently claimed invention provides for novel methods and kits for reducing the complexity of a nucleic acid sample by providing non-gel based methods for amplification of a subset of the sequences in a sample. In a preferred embodiment, amplification of a subset can be accomplished by digesting a sample with two or more restriction enzymes and ligating adaptors to the fragments so that only a subset of the fragments can be amplified. The invention further provides for analysis of the above amplified sample by hybridization to an array, which may be specifically designed to interrogate the desired fragments for particular characteristics, such as, for example, the presence or absence of a polymorphism.

IDT Exhibit 1020.8897

COMPLEXITY MANAGEMENT OF GENOMIC DNA

RELATED APPLICATIONS

[01] This application claims the benefit of U.S. Application No. 09/428,350 filed 10/27/99, the entire teachings of which are incorporated herein by reference.

FIELD OF THE INVENTION

[02] The invention relates to enrichment and amplification of sequences from a nucleic acid sample. In one embodiment, the invention relates to enrichment and amplification of nucleic acids for the purpose of further analysis. The present invention relates to the fields of molecular biology and genetics.

BACKGROUND

[03] The past years have seen a dynamic change in the ability of science to comprehend vast amounts of data. Pioneering technologies such as nucleic acid arrays allow scientists to delve into the world of genetics in far greater detail than ever before. Exploration of genomic DNA has long been a dream of the scientific community. Held within the complex structures of genomic DNA lies the potential to identify, diagnose, or treat diseases like cancer, Alzheimer disease or alcoholism. Exploitation of genomic information from plants and animals may also provide answers to the world's food distribution problems.

[04] Recent efforts in the scientific community, such as the publication of the draft sequence of the human genome in February 2001, have changed the dream of genome exploration into a reality. Genome-wide assays, however, must contend with the complexity of genomes; the human genome for example is estimated to have a complexity of $3x10^9$ base pairs. Novel methods of sample preparation and sample analysis that reduce complexity may provide for the fast and cost effective exploration of complex samples of nucleic acids, particularly genomic DNA.

[05] In the future disease discovery and diagnosis will be increasingly done at the genetic level. There are already many diseases that are known to be associated with specific genetic components. That number will rapidly increase as researchers

continue to acquire information about the genetic variation between individuals and to associate genotypic information with phenotypic characteristics.

SUMMARY OF THE INVENTION

[06] The present invention provides for novel methods of sample preparation and analysis comprising managing or reducing the complexity of a nucleic acid sample. The present invention eliminates the need for multiplex PCR and for many of the embodiments the step of complexity reduction may be performed entirely in a single tube. The invention further provides for analysis of the sample by hybridization to an array which may be specifically designed to interrogate fragments for particular characteristics, such as, for example, the presence or absence of a polymorphism. The invention further provides for methods of using a computer system to model enzymatic reactions in order to determine experimental conditions and to design arrays. In a preferred embodiment the invention discloses novel methods of genome-wide polymorphism discovery and genotyping.

[07] In one embodiment the present invention provides methods for reducing the complexity of a nucleic acid sample by fragmenting the nucleic acid sample using restriction enzymes; ligating adaptors to the fragments; and selectively amplifying the fragments that were cut on one end by one restriction enzyme and on the other end by a different restriction enzyme. The complexity of the resulting sample can be regulated by choosing enzymes that cut with a particular frequency.

[08] In one embodiment of the current invention selective amplification is accomplished by blocking ligation of one strand of each adaptor so that fragments that have the same adaptor ligated to both ends will have gaps in both strands but fragments that have one adaptor ligated to one end and a second adaptor ligated to the other end will have one complete strand for amplification. In this embodiment, ligation is blocked at the 5' end of one adaptor and at the 3' end of the other adaptor.

[09] In another embodiment of the current invention selective amplification is accomplished by designing one adaptor to have a 5' overhang and the second adaptor to have a 3' overhang. The overhangs comprise primer binding sites.

[10] Many of the embodiments of the current invention can further reduce complexity by designing the primers for amplification to bind to one or more nucleotides of the variable region of the restriction fragments.

IDT Exhibit 1020.8899

[11] In yet another embodiment, the invention relates to a kit comprising reagents and instructions for reducing the complexity of a nucleic acid sample. The kit may comprise buffers, restriction enzymes, adaptors, primers, a ligase, a polymerase, dNTPS and instructions necessary for use of the kit.

[12] The present invention also provides methods for genotyping an individual which may further comprise contacting the amplified sequences with a solid support comprising nucleic acid probes, and detecting the presence or absence of hybridization of the amplified sequences to the nucleic acid probes on the solid support. The immobilized probes in a preferred embodiment are capable of interrogating one or more polymorphic sites. The identity of the polymorphic base is determined from the hybridization information.

BRIEF DESCRIPTION OF THE DRAWINGS

[13] Figure 1 is a schematic illustration of a method of complexity reduction that uses targeted amplification of a subset of restriction fragments wherein the targeting is accomplished by selectively blocking ligation of one strand of each of two adaptors.

[14] Figure 2 is a schematic showing how blocked ligation can be used to selectively amplify the fragments that have been digested by two different restriction enzymes.

[15] Figure 3 shows an example of how an adaptor can be used to introduce a gap, resulting in blocked ligation and blocked extension at the gap.

[16] Figure 4 is a schematic illustration of a method of complexity reduction that uses targeted amplification of a subset of restriction fragments wherein the targeting is accomplished by adding overhangs to the adaptors, which comprise primer binding sites.

DETAILED DESCRIPTION

(A) General

[17] The present invention provides a flexible and scalable method for analyzing complex samples of nucleic acids, such as genomic DNA. These methods are not limited to any particular type of nucleic acid sample: plant, bacterial, animal (including human) total genome DNA, RNA, cDNA and the like may be analyzed using some or all of the methods disclosed in this invention. The word "DNA" may be used below as an example of a nucleic acid. It is understood that this term includes all

IDT Exhibit 1020.8900

nucleic acids, such as DNA and RNA, unless a use below requires a specific type of nucleic acid. This invention provides a powerful tool for analysis of complex nucleic acid samples. From experimental design to isolation of desired fragments and hybridization to an appropriate array, the invention provides for fast, efficient and inexpensive methods of complex nucleic acid analysis.

[18] The present invention relies on many patents, applications and other references for details known to those of the art. Therefore, when a patent, application, or other reference is cited or repeated below, it should be understood that it is incorporated by reference in its entirety for all purposes as well as for the proposition that is recited.

[19] As used in the specification and claims, the singular form "a," "an," and "the" include plural references unless the context clearly dictates otherwise. For example, the term "an agent" includes a plurality of agents, including mixtures thereof. An individual is not limited to a human being but may also be other organisms including but not limited to mammals, plants, bacteria, or cells derived from any of the above.

[20] Throughout this disclosure, various aspects of this invention are presented in a range format. It should be understood that the description in range format is merely for convenience and brevity and should not be construed as an inflexible limitation on the scope of the invention. Accordingly, the description of a range should be considered to have specifically disclosed all the possible subranges as well as common individual numerical values within that range. For example, description of a range such as from 1 to 6 should be considered to have specifically disclosed subranges such as from 1 to 3, from 1 to 4, from 1 to 5, from 2 to 4, from 2 to 6, from 3 to 6 etc., as well as individual numbers within that range, for example, 1, 2, 3, 4, 5, and 6. The same holds true for ranges in increments of 10^5, 10^4, 10^3, 10^2, 10, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, or 10^{-5}, for example. This applies regardless of the breadth of the range.

[21] The practice of the present invention may employ, unless otherwise indicated, conventional techniques of organic chemistry, polymer technology, molecular biology (including recombinant techniques), cell biology, biochemistry, and immunology, which are within the skill of the art. Such conventional techniques include polymer array synthesis, hybridization, ligation, and detection of hybridization using a label. Specific illustrations of suitable techniques can be had by reference to the example hereinbelow. However, other equivalent conventional procedures can, of course, also be used. Such conventional techniques can be found in standard laboratory manuals

IDT Exhibit 1020.8901

such as *Genome Analysis: A Laboratory Manual Series (Vols. I-IV)*, *Using Antibodies: A Laboratory Manual*, *Cells: A Laboratory Manual*, *PCR Primer: A Laboratory Manual*, and *Molecular Cloning: A Laboratory Manual* (all from Cold Spring Harbor Laboratory Press), all of which are herein incorporated in their entirety by reference for all purposes.

[22] Some aspects of the present invention make use of microarrays, also called arrays. Methods and techniques applicable to array synthesis have been described in U.S. Patents Nos. 5,143,854, 5,242,974, 5,252,743, 5,324,633, 5,384,261, 5,424,186, 5,451,683, 5,482,867, 5,491,074, 5,527,681, 5,550,215, 5,571,639, 5,578,832, 5,593,839, 5,599,695, 5,624,711, 5,631,734, 5,795,716, 5,831,070, 5,837,832, 5,856,101, 5,858,659, 5,936,324, 5,968,740, 5,974,164, 5,981,185, 5,981,956, 6,025,601, 6,033,860, 6,040,193, and 6,090,555. All of the above patents incorporated herein by reference in their entireties for all purposes.

(B) Definitions

[23] Nucleic acids according to the present invention may include any polymer or oligomer of pyrimidine and purine bases, preferably cytosine, thymine, and uracil, and adenine and guanine, respectively. (*See* Albert L. Lehninger, *Principles of Biochemistry*, at 793-800 (Worth Pub. 1982) which is herein incorporated by reference in its entirety for all purposes). Indeed, the present invention contemplates any deoxyribonucleotide, ribonucleotide or peptide nucleic acid component, and any chemical variants thereof, such as methylated, hydroxymethylated or glycosylated forms of these bases, and the like. The polymers or oligomers may be heterogeneous or homogeneous in composition, and may be isolated from naturally occurring sources or may be artificially or synthetically produced. In addition, the nucleic acids may be DNA or RNA, or a mixture thereof, and may exist permanently or transitionally in single-stranded or double-stranded form, including homoduplex, heteroduplex, and hybrid states.

[24] An "oligonucleotide" or "polynucleotide" is a nucleic acid ranging from at least 2, preferably at least 8, 15 or 20 nucleotides in length, but may be up to 50, 100, 1000, or 5000 nucleotides long or a compound that specifically hybridizes to a polynucleotide. Polynucleotides of the present invention include sequences of deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) or mimetics thereof which

IDT Exhibit 1020.8902

may be isolated from natural sources, recombinantly produced or artificially synthesized. A further example of a polynucleotide of the present invention may be a peptide nucleic acid (PNA). (*See* U.S. Patent No. 6,156,501 which is hereby incorporated by reference in its entirety for all purposes.) The invention also encompasses situations in which there is a nontraditional base pairing such as Hoogsteen base pairing which has been identified in certain tRNA molecules and postulated to exist in a triple helix. "Polynucleotide" and "oligonucleotide" are used interchangeably in this application.

[25] The term "fragment," "segment," or "DNA segment" refers to a portion of a larger DNA polynucleotide or DNA. A polynucleotide, for example, can be broken up, or fragmented into, a plurality of segments. Various methods of fragmenting nucleic acids are well known in the art. These methods may be, for example, either chemical or physical in nature. Chemical fragmentation may include partial degradation with a DNAse; partial depurination with acid; the use of restriction enzymes; intron-encoded endonucleases; DNA-based cleavage methods, such as triplex and hybrid formation methods, that rely on the specific hybridization of a nucleic acid segment to localize a cleavage agent to a specific location in the nucleic acid molecule; or other enzymes or compounds which cleave DNA at known or unknown locations. Physical fragmentation methods may involve subjecting the DNA to a high shear rate. High shear rates may be produced, for example, by moving DNA through a chamber or channel with pits or spikes, or forcing the DNA sample through a restricted size flow passage, *e.g.*, an aperture having a cross sectional dimension in the micron or submicron scale. Other physical methods include sonication and nebulization. Combinations of physical and chemical fragmentation methods may likewise be employed such as fragmentation by heat and ion-mediated hydrolysis. *See* for example, Sambrook *et al.*, "Molecular Cloning: A Laboratory Manual," 3rd Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (2001) ("Sambrook *et al.*") which is incorporated herein by reference in its entirety for all purposes. These methods can be optimized to digest a nucleic acid into fragments of a selected size range. Useful size ranges may be from 100, 200, 400, 700 or 1000 to 500, 800, 1500, 2000, 4000 or 10,000 base pairs. However, larger size ranges such as 4000, 10,000 or 20,000 to 10,000, 20,000 or 500,000 base pairs may also be useful.

IDT Exhibit 1020.8903

[26] A number of methods disclosed herein require the use of restriction enzymes to fragment the nucleic acid sample. In general, a restriction enzyme recognizes a specific nucleotide sequence of four to eight nucleotides and cuts the DNA at a site within or a specific distance from the recognition sequence. For example, the restriction enzyme *Eco*RI recognizes the sequence GAATTC and will cut a DNA molecule between the G and the first A. The length of the recognition sequence is roughly proportional to the frequency of occurrence of the site in the genome. A simplistic theoretical estimate is that a six base pair recognition sequence will occur once in every 4096 (4^6) base pairs and is considered to be a "rare cutter", while a four base pair recognition sequence will occur once every 256 (4^4) base pairs and is considered to be a "frequent cutter". An enzyme with an 8 base pair recognition sequence is also a "rare cutter" cutting once every 4^8 base pairs. *In silico* digestions of sequences from the Human Genome Project show that the actual occurrences are even more infrequent, depending on the sequence of the restriction site. Because the restriction sites are rare, the appearance of shorter restriction fragments, for example those less than 1000 base pairs, is much less frequent than the appearance of longer fragments. Many different restriction enzymes are known and appropriate restriction enzymes can be selected for a desired result. (For a description of many restriction enzymes, *see* New England BioLabs Catalog which is herein incorporated by reference in its entirety for all purposes).

[27] "Adaptor sequences" or "adaptors" are generally oligonucleotides of at least 5, 10, or 15 bases and preferably no more than 50 or 60 bases in length, however, they may be even longer, up to 100 or 200 bases. Adaptor sequences may be synthesized using any methods known to those of skill in the art. For the purposes of this invention they may, as options, comprise templates for PCR primers, restriction sites and promoters. The adaptor may be entirely or substantially double stranded. The adaptor may be phosphorylated or unphosphorylated on one or both strands. Adaptors are particularly useful in one embodiment of the current invention if they comprise short single stranded regions, which are complementary to the single stranded region, also known as an overhang, created by digestion with a restriction enzyme. For example, when DNA is digested with the restriction enzyme *Eco*RI the resulting double stranded fragments are flanked at either end by the single stranded overhang 5'-AATT-3', an adaptor that carries a single stranded overhang 5'-AATT-3' will hybridize to the fragment through complementarity between the overhanging

IDT Exhibit 1020.8904

regions. This "sticky end" hybridization of the adaptor to the fragment facilitates ligation of the adaptor to the fragment, although blunt ended ligation is also possible. Different restriction enzymes cleave DNA to produce different overhangs and adaptors can be designed for any overhang. For a list of many restriction enzymes, their recognition sites and the resulting cleavage products, *see*, for example, New England Biolabs Catalog, which is herein incorporated by reference in its entirety for all purposes, available at www.neb.com, last visited 7/25/01.

[28] Digestion with two or more enzymes can be used to selectively ligate separate adaptors to either end of a restriction fragment. Restriction fragments that have been generated by two different restriction enzymes will often have two distinct overhangs. A pair of adaptors can be designed to ligate specifically to one or the other overhang so that the fragment will ligate to one adaptor on one end and to the second adaptor on the other end. For example, if a fragment is the result of digestion with *Eco*RI at one end and *Bam*HI at the other end, the overhangs will be 5'-AATT-3' and 5'GATC-3', respectively. An adaptor with an overhang of AATT will be preferentially ligated to one end while an adaptor with an overhang of GATC will be preferentially ligated to the second end.

[29] A single adaptor can also be ligated to both ends of a fragment resulting from digestion with two different enzymes. For example, if the method of digestion generates blunt ended fragments, the same adaptor sequence can be ligated to both ends. Alternatively some pairs of enzymes leave identical overhanging sequences. For example, *Bgl*II recognizes the sequence 5'-AGATCT-3', cutting after the first A, and *Bam*HI recognizes the sequence 5'-GGATCC-3', cutting after the first G; both leave an overhang of 5'-GATC-3'. A single adaptor with an overhang of 5'-GATC-3' may be ligated to both digestion products. Adaptors can also be designed to recreate the original restriction site, create a new restriction site or to eliminate the original site.

[30] Methods of ligation will be known to those of skill in the art and are described, for example in Sambrook *et al.* and the New England BioLabs catalog, both of which are incorporated herein in their entireties by reference for all purposes. Methods include using T4 DNA ligase, which catalyzes the formation of a phosphodiester bond between juxtaposed 5′ phosphate and 3′ hydroxyl termini in duplex DNA or RNA with blunt or and sticky ends; *Taq* DNA ligase, which catalyzes the formation of a phosphodiester bond between juxtaposed 5′ phosphate and 3′ hydroxyl termini of two

IDT Exhibit 1020.8905

adjacent oligonucleotides that are hybridized to a complementary target DNA; *E.coli* DNA ligase, which catalyzes the formation of a phosphodiester bond between juxtaposed 5′ -phosphate and 3′ -hydroxyl termini in duplex DNA containing cohesive ends; and T4 RNA ligase which catalyzes ligation of a 5′ phosphoryl-terminated nucleic acid donor to a 3′ hydroxyl-terminated nucleic acid acceptor through the formation of a 3′ –>5′ phosphodiester bond, substrates include single-stranded RNA and DNA as well as dinucleoside pyrophosphates; or any other methods described in the art.

[31] "Genome" designates or denotes the complete, single-copy set of genetic instructions for an organism as coded into the DNA of the organism. A genome may be multi-chromosomal such that the DNA is distributed among a plurality of individual chromosomes. For example, in human there are 22 pairs of chromosomes plus a gender associated XX or XY pair.

[32] The term "chromosome" refers to the heredity-bearing gene carrier of a living cell which is derived from chromatin and which comprises DNA and protein components (especially histones). The conventional internationally recognized individual human genome chromosome numbering system is employed herein. The size of an individual chromosome can vary from one type to another with a given multi-chromosomal genome and from one genome to another. In the case of the human genome, the entire DNA mass of a given chromosome is usually greater than about 100,000,000 bp. For example, the size of the entire human genome is about 3 x 10^9 bp. The largest chromosome, chromosome no. 1, contains about 2.4 x 10^8 bp while the smallest chromosome, chromosome no. 22, contains about 5.3 x 10^7 bp.

[33] A "chromosomal region" is a portion of a chromosome. The actual physical size or extent of any individual chromosomal region can vary greatly. The term "region" is not necessarily definitive of a particular one or more genes because a region need not take into specific account the particular coding segments (exons) of an individual gene.

[34] The term subset or representative subset refers to a fraction of a genome. The subset may be, for example, 0.01, 0.1, 1, 3, 5, 10, 25, 50 or 75% of the genome. The partitioning of fragments into subsets may be done according to a variety of physical characteristics of individual fragments. For example, fragments may be divided into subsets according to size, according to the particular combination of restriction sites at

IDT Exhibit 1020.8906

the ends of the fragment, or based on the presence or absence of one or more particular sequences.

[35] An "array" comprises a support, preferably solid, with nucleic acid probes attached to the support. Arrays typically comprise a plurality of different nucleic acid probes that are coupled to a surface of a substrate in different, known locations. These arrays, also described as "microarrays" or colloquially "chips" have been generally described in the art, for example, U.S. Pat. Nos. 5,143,854, 5,445,934, 5,744,305, 5,677,195, 5,800,992, 6,040,193, 5,424,186 and Fodor *et al.*, *Science*, 251:767-777 (1991). Each of which is incorporated by reference in its entirety for all purposes.

[36] Arrays may generally be produced using a variety of techniques, such as mechanical synthesis methods or light directed synthesis methods that incorporate a combination of photolithographic methods and solid phase synthesis methods. Techniques for the synthesis of these arrays using mechanical synthesis methods are described in, *e.g.*, U.S. Pat. No. 5,384,261, and 6,040,193, which are incorporated herein by reference in their entirety for all purposes. Although a planar array surface is preferred, the array may be fabricated on a surface of virtually any shape or even a multiplicity of surfaces. Arrays may be nucleic acids on beads, gels, polymeric surfaces, fibers such as fiber optics, glass or any other appropriate substrate. (*See* U.S. Patent Nos. 5,770,358, 5,789,162, 5,708,153, 6,040,193 and 5,800,992, which are hereby incorporated by reference in their entirety for all purposes.)

[37] Arrays may be packaged in such a manner as to allow for diagnostic use or can be an all-inclusive device; *e.g.*, U.S. Patent Nos. 5,856,174 and 5,922,591 incorporated in their entirety by reference for all purposes.

[38] Preferred arrays are commercially available from Affymetrix under the brand name GeneChip® and are directed to a variety of purposes, including genotyping and gene expression monitoring for a variety of eukaryotic and prokaryotic species. (*See* Affymetrix Inc., Santa Clara and their website at www.affymetrix.com.)

[39] Polymorphism refers to the occurrence of two or more genetically determined alternative sequences or alleles in a population. A polymorphic marker or site is the locus at which divergence occurs. Preferred markers have at least two alleles, each occurring at a frequency of preferably greater than 1%, and more preferably greater than 10% or 20% of a selected population. A polymorphism may comprise one or more base changes, an insertion, a repeat, or a deletion. A polymorphic locus may be as small as one base pair. Polymorphic markers include restriction fragment length

IDT Exhibit 1020.8907

polymorphisms, variable number of tandem repeats (VNTR's), hypervariable regions, minisatellites, dinucleotide repeats, trinucleotide repeats, tetranucleotide repeats, simple sequence repeats, and insertion elements such as Alu. The first identified allelic form is arbitrarily designated as the reference form and other allelic forms are designated as alternative or variant alleles. The allelic form occurring most frequently in a selected population is sometimes referred to as the wildtype form. Diploid organisms may be homozygous or heterozygous for allelic forms. A diallelic polymorphism has two forms. A triallelic polymorphism has three forms. A polymorphism between two nucleic acids can occur naturally, or be caused by exposure to or contact with chemicals, enzymes, or other agents, or exposure to agents that cause damage to nucleic acids, for example, ultraviolet radiation, mutagens or carcinogens.

[40] Single nucleotide polymorphisms (SNPs) are positions at which two alternative bases occur at appreciable frequency (>1%) in the human population, and are the most common type of human genetic variation. The site is usually preceded by and followed by highly conserved sequences of the allele (*e.g.*, sequences that vary in less than 1/100 or 1/1000 members of the populations). It is estimated that there are as many as $3x10^6$ SNPs in the human genome.

[41] A single nucleotide polymorphism usually arises due to substitution of one nucleotide for another at the polymorphic site. A transition is the replacement of one purine by another purine or one pyrimidine by another pyrimidine. A transversion is the replacement of a purine by a pyrimidine or vice versa. Single nucleotide polymorphisms can also arise from a deletion of a nucleotide or an insertion of a nucleotide relative to a reference allele.

[42] The term genotyping refers to the determination of the genetic information an individual carries at one or more positions in the genome. For example, genotyping may comprise the determination of which allele or alleles an individual carries for a single polymorphism or the determination of which allele or alleles an individual carries for a plurality of polymorphisms.

[43] Hybridization probes are oligonucleotides capable of binding in a base-specific manner to a complementary strand of nucleic acid. Such probes include peptide nucleic acids, as described in Nielsen *et al.*, *Science* 254, 1497-1500 (1991), and other nucleic acid analogs and nucleic acid mimetics. *See* US Patent Application No. 6,156,501.

IDT Exhibit 1020.8908

[44] Hybridizations are usually performed under stringent conditions, for example, at a salt concentration of no more than 1 M and a temperature of at least 25°C. For example, conditions of 5X SSPE (750 mM NaCl, 50 mM NaPhosphate, 5 mM EDTA, pH 7.4) and a temperature of 25-30°C are suitable for allele-specific probe hybridizations. For stringent conditions, *see,* for example, Sambrook *et al.*, which is hereby incorporated by reference in its entirety for all purposes above.

[45] An individual is not limited to a human being, but may also include other organisms including but not limited to mammals, plants, bacteria or cells derived from any of the above.

[46] *In silico* digestion is a computer aided simulation of enzymatic digests accomplished by searching a sequence for restriction sites. *In silico* digestion provides for the use of a computer system to model enzymatic reactions in order to determine experimental conditions before conducting any actual experiments. An example of an experiment would be to model digestion of the human genome with specific restriction enzymes to predict the sizes of the resulting restriction fragments.

(C.) The Process

[47] The present invention provides for novel methods of sample preparation and analysis involving managing or reducing the complexity of a nucleic acid sample, such as genomic DNA, by amplifying a representative subset of the sample. The invention further provides for analysis of the above subset by hybridization to an array, which may be specifically designed to interrogate the desired fragments for particular characteristics, such as, for example, the presence or absence of a polymorphism. The invention is particularly useful when combined with other methods of genome analysis. As an example, the present techniques are useful to genotype individuals after polymorphisms have been identified.

[48] One method that has been used to isolate a subset of a genome is to separate fragments according to size by electrophoresis in a gel matrix. The region of the gel containing fragments in the desired size range is then excised and the fragments are purified away from the gel matrix. The SNP consortium (TSC) adopted this approach in their efforts to discover single nucleotide polymorphisms (SNPs) in the human genome. *See*, Altshuler *et al.*, *Science* 407: 513-516 (2000) and The International

IDT Exhibit 1020.8909

SNP Map Working Group, *Nature* 409: 928-933 (2001) both of which are herein incorporated by reference in their entirety for all purposes.

[49] The present invention provides methods of complexity management of nucleic acid samples, such as genomic DNA, that can be used as an alternative to separation of fragments by gel electrophoresis and purification of fragments from a gel matrix. Generally, the embodiments include the steps of: fragmenting the nucleic acid by digestion with two or more restriction enzymes; ligating adaptors to the ends of the fragments; and selectively amplifying a subset of the fragments that have a first adaptor ligated to one end and a second, different adaptor ligated to the other end. In a preferred embodiment the amplified sequences are then exposed to an array which may or may not have been specifically designed and manufactured to interrogate the amplified sequences. Design of both the complexity management steps and the arrays may be aided by computer modeling techniques. Generally, the steps of the present invention involve reducing the complexity of a nucleic acid sample using the disclosed techniques alone or in combination.

[50] When interrogating genomes it is often useful to first reduce the complexity of the sample and analyze one or more subsets of the genome. Subsets can be defined by many characteristics of the fragments. In a preferred embodiment of the current invention the subsets are defined by the restriction enzymes used to generate the fragments and the resulting structure of the ends of the fragments. In some embodiments, the subset is further defined by the identity of one or more nucleotides flanking one or both of the restriction enzyme sites used to generate the fragment.

[51] The genomic DNA sample of the current invention may be isolated according to methods known in the art, such as PCR, reverse transcription, and the like. It may be obtained from any biological or environmental source, including plant, animal (including human), bacteria, fungi or algae. Any suitable biological sample can be used for assay of genomic DNA. Convenient suitable samples include whole blood, tissue, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair.

[52] In a preferred embodiment of the invention, adaptors are ligated to the ends of the fragments and the fragments are amplified by PCR using primers that are designed to hybridize to sequences in the adaptors. In a particularly preferred embodiment, a single primer pair can be used for amplification.

[53] Figure 1 depicts a schematic illustration of the general steps of a preferred embodiment of the currently claimed invention. First, a sample of genomic DNA is

IDT Exhibit 1020.8910

fragmented with a first restriction enzyme, RE1, which in a preferred embodiment is a rare cutter, and a second restriction enzyme, RE2, which in a preferred embodiment is a frequent cutter, to produce three types of fragments. The most common fragments are those which have been cut on both ends with the frequent base cutter. Those fragments, which have been cut on one end with the frequent base cutter and on one end with the rare base cutter will be less frequent. The least common fragments are those which have been cut on both ends with the rare base cutter. One of the adaptors, A1, is designed to ligate to the overhang left by the rare cutter and the other adapter, A2, is designed to ligate to the overhang left by the frequent cutter. Fragments that are cut twice by RE1 will ligate to adaptor A1 on both ends. Fragments that are cut twice by RE2 will ligate to adaptor A2 on both ends. Those fragments that have been cut once with RE1 and once with RE2 will ligate to adaptor A1 on one end and to adaptor A2 on the other end. The adaptors further comprise a primer binding site.

[54] In the embodiment of the invention illustrated in Figure 1, ligation of one strand of each adapter is blocked. Blocking ligation of one strand of the adaptor to the fragments results in a gap or nick between that strand and one strand of the fragment. The gap or nick will block primer extension. In the current invention, one adaptor is designed so that ligation is blocked between the 3' end of one strand of the adaptor and the 5' end of the fragment and the other adaptor is designed to block ligation between the 5' end of one strand of the adaptor and the 3' end of the fragment.

[55] After adaptor ligation the sample is amplified by PCR. Primer P1, specific for adaptor A1, and primer P2, specific for adapter A2, are added, and the mixture is used as template for PCR. Fragments with two identical adaptors are not amplified because primer extension is blocked in both strands at the introduced gaps or nicks (*see*, Figure 2) so the first strand of DNA synthesized terminates without the primer binding site needed for second strand synthesis. However, fragments with both adaptors A1 and A2 will be amplified because both introduced gaps or nicks are in the same strand leaving a complete strand to serve as a template for PCR.

[56] In this embodiment, the gaps or nicks may be introduced, for example, by blocking ligation of one strand of the adaptor to the fragment. Both the adaptors and the fragments are double stranded so ligation of the adaptor to the fragment typically results in the formation of two phosphodiester bonds, one between one adaptor strand

IDT Exhibit 1020.8911

and one strand of the fragment and a second between the other adaptor strand and the other strand of the fragment. In one embodiment of the current invention, formation of one of these phosphodiester bonds is blocked. The adaptor is ligated to the fragment through one phosphodiester bond but one strand of the resulting ligated fragment will have a gap between the adaptor and the restriction fragment. When this strand is used as a template for primer extension the primer extension reaction will stop when the nick or gap is reached as illustrated in Figure 2.

[57] Ligation of one strand of the adaptor can be selectively blocked by any available method. In a preferred embodiment of the current invention, the adaptors carry a terminal modification that blocks ligation. For example, the presence of a ddNTP at the 3' end of an adaptor strand prevents ligation because there is no 3' hydroxyl available for the ligation reaction. Other 3' modifications that can be used to block ligation include a 3' phosphate group, phosphorothioate, biotin, fluorescent dye, 2',3' cyclic phosphate, or any other modification that results in the absence of a free 3' hydroxyl. Similarly, the absence of a free phosphate at the 5' end of an adaptor strand will block ligation. Typically oligonucleotides are synthesized without a 5' phosphate, but a phosphate can be added by, for example, a kinasing reaction or be removed by, for example, a phosphatase reaction..

[58] In another preferred embodiment, the adaptors are designed to introduce a gap that blocks ligation. (*See*, Figure 3). A gap of one or more bases can be introduced between one strand of the adaptor and one strand of the fragment. One method of introducing such a gap is by designing the adaptor overhang with either one extra base or with one fewer base. (*See*, Figure 3). In the example in figure 3, the adapter designed to hybridize to the overhang left by *Sau*3A is designed to introduce a gap between the 5' phosphate of the adaptor and the 3'OH of the fragment (5'-ATC-3' instead of 5'-GATC-3'). Alternatively, the adaptor designed to hybridize to the overhang left by *Eco*RI in figure 3 will introduce a gap between the 5' phosphate of the fragment and the 3' OH of the adaptor by adding an extra nucleotide to the overhang, (5'-AATTN-3' instead of 5'-AATT-3' where N is any nucleotide).

[59] A combination of methods may also be used, for example, ligation of one adaptor may be blocked by the absence of the necessary phosphate while ligation of the other adaptor is blocked by introduction of a gap of one or more nucleotides.

[60] In this embodiment of the current invention, ligation is preferably blocked at the 5' end of one of the two adaptors and at the 3' end of the other adaptor. More

IDT Exhibit 1020.8912

specifically, if ligation of the first adaptor is blocked between the 3' end of the adaptor and the 5' end of the fragment, then ligation of the second adaptor should be blocked between the 5' end of the adaptor and the 3' end of the fragment. Likewise, if ligation of the first adaptor is blocked between the 5' end of the adaptor and the 3' end of the fragment, then ligation of the second adaptor should be blocked between the 3' end of the adaptor and the 5' end of the fragment. As illustrated in Figure 2, this results in blocked amplification of the fragments with two similar ends but allows for amplification of the fragments with different ends because only those fragments will have a complete template strand.

[61] In another embodiment of the current invention, illustrated in Figure 4, the adaptors are designed with single stranded regions that allow for selective amplification. In the example illustrated in Figure 4, adaptor A1, which is designed to ligate to the overhang left by RE1, has an additional single stranded 3' overhang that serves as a primer binding site for primer P1. Adaptor A2, which is designed to be ligated to the overhang left by RE2, has an additional single stranded 5' overhang that serves as a template for a primer binding site for primer P2. Only fragments that have A1 ligated to one end and A2 ligated to the other end will be amplified. (*See* Figure 4).

[62] Using the current invention different combinations of enzymes can be used for digestion based upon the predicted frequency of cutting and the desired level of complexity reduction. For example, using a simplistic theoretical calculation, if a 6 base cutter is combined with a four base cutter the resulting complexity reduction is 8 fold. The 6 base cutter is predicted to cut the genome into fragments that are on average 4^6 base pairs and the 4 base cutter is predicted to cut the genome into fragments that are on average 4^4 base pairs. In the double digest, the 4^6 base pair fragments should be digested into 16 4^4 base pair fragments. Two of these fragments, or $1/8^{th}$ of the genome, should have been cut by the 6 base cutter on one end and the 4 base cutter on the other end. In a further example, if an eight base cutter is combined with a four base cutter the complexity reduction is 128 fold. (*See*, Table 1 below). Similar calculations can be done for enzymes with different length recognition sequences.

[63] In yet another embodiment, further complexity reduction is achieved by designing at least one of the primers to hybridize at its 3' end to at least one nucleotide of the variable region of the restriction fragments to be amplified. The

IDT Exhibit 1020.8913

additional selective nucleotide(s) further reduces the complexity of the resulting DNA pool because only those fragments which contain the complement of the specific nucleotide(s) incorporated in the PCR primer will be amplified. The level of complexity in the isolated pool can be varied depending upon the identity and number of nucleotides incorporated into the PCR primers and the restriction enzymes used to fragment the DNA. For example, if an enzyme with a 6 base recognition sequence is combined with an enzyme with a four base pair recognition sequence and only fragments that were digested with both enzymes are amplified on average 1 of every 8 restriction fragments will be amplified. This results in an 8 fold reduction in the complexity of the sample, reducing the 3,300 Mb human genome to 375 Mb. If an additional selective nucleotide is added to one of the primers only $1/4^{th}$ of those fragments will be amplified or 1 out of every 32 restriction fragments, resulting in a 32 fold reduction in complexity. *See*, Table 1 and Table 2. Different combinations of selective nucleotides can be used to amplify different populations of SNPs. Table 2 illustrates the use of different selective nucleotides in combination with digestion with enzymes with a 6 and 4 base pair recognition sequence. A number of embodiments of the present invention involve amplification by PCR. Any of these embodiments may be further modified to reduce complexity using the above disclosed technique.

Table 1.

# of Selective Nucleotides	Number of bases in RE1 recognition site	Number of bases in RE2 recognition site	Fraction of RE1-RE2 Fragments	Complexity Reduction from Total Genome (Mb)
0	6	4	1/8	375
1	6	4	1/32	96
2	6	4	1/128	24
3	6	4	1/512	6
0	8	4	1/128	24
1	8	4	1/512	6

Table 2.

Selective nucleotides	Estimated Complexity	# of TSC SNPs Selected
0|A	96Mb	24,000
A|A	24Mb	6,000

IDT Exhibit 1020.8914

A|T	24Mb	6,000
T|A	24Mb	6,000
C|T	24Mb	6,000
A|AT	6Mb	1,000
A|AG	6Mb	1,000

[64] There are many known methods of amplifying nucleic acid sequences including *e.g.*, PCR. *See, e.g., PCR Technology: Principles and Applications for DNA Amplification* (ed. H.A. Erlich, Freeman Press, NY, NY, 1992); *PCR Protocols: A Guide to Methods and Applications* (eds. Innis, *et al.*, Academic Press, San Diego, CA, 1990); Mattila *et al.*, *Nucleic Acids Res.* 19, 4967 (1991); Eckert *et al.*, *PCR Methods and Applications* 1, 17 (1991); *PCR* (eds. McPherson *et al.*, IRL Press, Oxford); and U.S. Patent 4,683,202, 4,683,195, 4,800,159 4,965,188 and 5,333,675 each of which is incorporated herein by reference in their entireties for all purposes.

[65] PCR is an extremely powerful technique for amplifying specific polynucleotide sequences, including genomic DNA, single-stranded cDNA, and mRNA among others. Various methods of conducting PCR amplification and primer design and construction for PCR amplification will be known to those of skill in the art. Generally, in PCR a double stranded DNA to be amplified is denatured by heating the sample. New DNA synthesis is then primed by hybridizing primers to the target sequence in the presence of DNA polymerase and excess dNTPs. In subsequent cycles, the primers hybridize to the newly synthesized DNA to produce discreet products with the primer sequences at either end. The products accumulate exponentially with each successive round of amplification. Standard protocols may be found in, for example Sambrook *et al.* which is hereby incorporated by reference for all purposes.

[66] The DNA polymerase used in PCR is often a thermostable polymerase. This allows the enzyme to continue functioning after repeated cycles of heating necessary to denature the double stranded DNA. Polymerases that are useful for PCR include, for example, *Taq* DNA polymerase, *Tth* DNA polymerase, *Tfl* DNA polymerase, *Tma* DNA polymerase, *Tli* DNA polymerase, and *Pfu* DNA polymerase. There are many commercially available modified forms of these enzymes including: AmpliTaq® and AmpliTaq Gold® both available from Applied Biosystems. Many are available with

IDT Exhibit 1020.8915

or without a 3- to 5' proofreading exonuclease activity. *See*, for example, Vent® and Vent® (exo-) available from New England Biolabs.

[67] Other suitable amplification methods include the ligase chain reaction (LCR) *(e.g.,* Wu and Wallace, *Genomics* 4, 560 (1989) and Landegren *et al.*, *Science* 241, 1077 (1988)), transcription amplification (Kwoh *et al.*, *Proc. Natl. Acad. Sci. USA* 86, 1173 (1989)), and self-sustained sequence replication (Guatelli *et al.*, *Proc. Nat. Acad. Sci. USA,* 87, 1874 (1990)) and nucleic acid based sequence amplification (NABSA). (*See*, US patents nos. 5,409,818, 5,554517, and 6,063,603). The latter two amplification methods include isothermal reactions based on isothermal transcription, which produce both single-stranded RNA (ssRNA) and double-stranded DNA (dsDNA) as the amplification products in a ratio of about 30 or 100 to 1, respectively.

[68] As those of skill in the art will appreciate, after isolation and amplification, the resulting sequences may be further analyzed using any known method including sequencing, HPLC, hybridization analysis, etc.

[69] Any of the above embodiments can be combined with *in silico* digestion techniques to predict the number and identities of SNPs predicted to be represented in a PCR product amplified using the above methods. For example, if restriction fragments that have been cleaved by *Eco*RI at one end and *Sau*3A at the other end are the subset to be amplified, a computer can be used to model an *Eco*RI and *Sau*3A digest of the genome to identify that subset of fragments and to identify the known or suspected SNPs that are present on that subset of fragments.

[70] The invention also includes kits for performing one or more of the different methods for analyzing a polynucleotide population described herein. Kits generally contain reagents necessary to perform the subject methods. The reagents may be supplied in pre-measured amounts for individual assays so as to increase reproducibility. Reagents may also be provided in a concentrated form.

[71] In one embodiment, the subject kits comprise adaptors and primers to amplify representative restriction fragments. The kits of the invention may also include one or more additional reagents required for various embodiments of the subject methods. Such additional reagents include, but are not limited to: restriction enzymes, DNA polymerases, DNA ligases, buffers, nucleotides and the like.

IDT Exhibit 1020.8916

METHODS OF USE

[72] The methods of the presently claimed invention can be used for a wide variety of applications. Any analysis of genomic DNA may be benefited by a reproducible method of complexity management. Furthermore, the methods and enriched fragments of the presently claimed invention are particularly well suited for study and characterization of extremely large regions of genomic DNA.

[73] In a preferred embodiment, the methods of the presently claimed invention are used for SNP discovery and to genotype individuals. For example, any of the procedures described above, alone or in combination, could be used to isolate the SNPs present in one or more specific regions of genomic DNA. Selection probes could be designed and manufactured to be used in combination with the methods of the invention to amplify only those fragments containing regions of interest, for example a region known to contain a SNP. Arrays could be designed and manufactured on a large scale basis to interrogate only those fragments containing the regions of interest. Thereafter, a sample from one or more individuals would be obtained and prepared using the same techniques which were used to prepare the selection probes or to design the array. Each sample can then be hybridized to an array and the hybridization pattern can be analyzed to determine the genotype of each individual or a population of individuals. Methods of use for polymorphisms and SNP discovery can be found in, for example, co-pending US application Nos. 08/813,159 and 09/428,350 which are herein incorporated by reference in their entirety for all purposes).

Correlation of Polymorphisms with Phenotypic Traits

[74] Most human sequence variation is attributable to or correlated with SNPs, with the rest attributable to insertions or deletions of one or more bases, repeat length polymorphisms and rearrangements. On average, SNPs occur every 1,000-2,000 bases when two human chromosomes are compared. (*See*, The International SNP Map Working Group, *Science* 409: 928-933 (2001) incorporated herein by reference in its entirety for all purposes.) Human diversity is limited not only by the number of SNPs occurring in the genome but further by the observation that specific combinations of alleles are found at closely linked sites.

IDT Exhibit 1020.8917

[75] Correlation of individual polymorphisms or groups of polymorphisms with phenotypic characteristics is a valuable tool in the effort to identify DNA variation that contributes to population variation in phenotypic traits. Phenotypic traits include physical characteristics, risk for disease, and response to the environment. Polymorphisms that correlate with disease are particularly interesting because they represent mechanisms to accurately diagnose disease and targets for drug treatment. Hundreds of human diseases have already been correlated with individual polymorphisms but there are many diseases that are known to have an, as yet unidentified, genetic component and many diseases for which a component is or may be genetic.

[76] Many diseases may correlate with multiple genetic changes making identification of the polymorphisms associated with a given disease more difficult. One approach to overcome this difficulty is to systematically explore the limited set of common gene variants for association with disease.

[77] To identify correlation between one or more alleles and one or more phenotypic traits, individuals are tested for the presence or absence of polymorphic markers or marker sets and for the phenotypic trait or traits of interest. The presence or absence of a set of polymorphisms is compared for individuals who exhibit a particular trait and individuals who exhibit lack of the particular trait to determine if the presence or absence of a particular allele is associated with the trait of interest. For example, it might be found that the presence of allele A1 at polymorphism A correlates with heart disease. As an example of a correlation between a phenotypic trait and more than one polymorphism, it might be found that allele A1 at polymorphism A and allele B1 at polymorphism B correlate with a phenotypic trait of interest.

Diagnosis of Disease and Predisposition to Disease

[78] Markers or groups of markers that correlate with the symptoms or occurrence of disease can be used to diagnose disease or predisposition to disease without regard to phenotypic manifestation. To diagnose disease or predisposition to disease, individuals are tested for the presence or absence of polymorphic markers or marker sets that correlate with one or more diseases. If, for example, the presence of allele A1 at polymorphism A correlates with coronary artery disease then individuals with allele A1 at polymorphism A may be at an increased risk for the condition.

IDT Exhibit 1020.8918

[79] Individuals can be tested before symptoms of the disease develop. Infants, for example, can be tested for genetic diseases such as phenylketonuria at birth. Individuals of any age could be tested to determine risk profiles for the occurrence of future disease. Often early diagnosis can lead to more effective treatment and prevention of disease through dietary, behavior or pharmaceutical interventions. Individuals can also be tested to determine carrier status for genetic disorders. Potential parents can use this information to make family planning decisions.

[80] Individuals who develop symptoms of disease that are consistent with more than one diagnosis can be tested to make a more accurate diagnosis. If, for example, symptom S is consistent with diseases X, Y or Z but allele A1 at polymorphism A correlates with disease X but not with diseases Y or Z an individual with symptom S is tested for the presence or absence of allele A1 at polymorphism A. Presence of allele A1 at polymorphism A is consistent with a diagnosis of disease X. Genetic expression information discovered through the use of arrays has been used to determine the specific type of cancer a particular patient has. (*See*, Golub *et al.* Science 286: 531-537 (2001) hereby incorporated by reference in its entirety for all purposes.)

Pharmacogenomics

[81] Pharmacogenomics refers to the study of how your genes affect your response to drugs. There is great heterogeneity in the way individuals respond to medications, in terms of both host toxicity and treatment efficacy. There are many causes of this variability, including: severity of the disease being treated; drug interactions; and the individuals age and nutritional status. Despite the importance of these clinical variables, inherited differences in the form of genetic polymorphisms can have an even greater influence on the efficacy and toxicity of medications. Genetic polymorphisms in drug-metabolizing enzymes, transporters, receptors, and other drug targets have been linked to interindividual differences in the efficacy and toxicity of many medications. (*See*, Evans and Relling, *Science* 286: 487-491 (2001) which is herein incorporated by reference for all purposes).

[82] An individual patient has an inherited ability to metabolize, eliminate and respond to specific drugs. Correlation of polymorphisms with pharmacogenomic traits identifies those polymorphisms that impact drug toxicity and treatment efficacy. This information can be used by doctors to determine what course of medicine is best

IDT Exhibit 1020.8919

for a particular patient and by pharmaceutical companies to develop new drugs that target a particular disease or particular individuals within the population, while decreasing the likelihood of adverse affects. Drugs can be targeted to groups of individuals who carry a specific allele or group of alleles. For example, individuals who carry allele A1 at polymorphism A may respond best to medication X while individuals who carry allele A2 respond best to medication Y. A trait may be the result of a single polymorphism but will often be determined by the interplay of several genes.

[83] In addition some drugs that are highly effective for a large percentage of the population, prove dangerous or even lethal for a very small percentage of the population. These drugs typically are not available to anyone. Pharmacogenomics can be used to correlate a specific genotype with an adverse drug response. If pharmaceutical companies and physicians can accurately identify those patients who would suffer adverse responses to a particular drug, the drug can be made available on a limited basis to those who would benefit from the drug.

[84] Similarly, some medications may be highly effective for only a very small percentage of the population while proving only slightly effective or even ineffective to a large percentage of patients. Pharmacogenomics allows pharamaceutical companies to predict which patients would be the ideal candidate for a particular drug, thereby dramatically reducing failure rates and providing greater incentive to companies to continue to conduct research into those drugs.

Determination of Relatedness

[85] There are many circumstances where relatedness between individuals is the subject of genotype analysis and the present invention can be applied to these procedures.

Paternity testing is commonly used to establish a biological relationship between a child and the putative father of that child. Genetic material from the child can be analyzed for occurrence of polymorphisms and compared to a similar analysis of the putative father's genetic material. Determination of relatedness is not limited to the relationship between father and child but can also be done to determine the relatedness between mother and child, (see *e.g.* Staub *et al.*, U.S. Pat. No.6,187,540) or more broadly, to determine how related one individual is to another, for example,

IDT Exhibit 1020.8920

between races or species or between individuals from geographically separated populations, (see for example H. Kaessmann, *et al. Nature Genet.* 22, 78 (1999)).

Forensics

[86] The capacity to identify a distinguishing or unique set of forensic markers in an individual is useful for forensic analysis. For example, one can determine whether a blood sample from a suspect matches a blood or other tissue sample from a crime scene by determining whether the set of polymorphic forms occupying selected polymorphic sites is the same in the suspect and the sample. If the set of polymorphic markers does not match between a suspect and a sample, it can be concluded (barring experimental error) that the suspect was not the source of the sample. If the set of markers does match, one can conclude that the DNA from the suspect is consistent with that found at the crime scene. If frequencies of the polymorphic forms at the loci tested have been determined (*e.g.*, by analysis of a suitable population of individuals), one can perform a statistical analysis to determine the probability that a match of suspect and crime scene sample would occur by chance. A similar comparison of markers can be used to identify an individual's remains. For example the U.S. armed forces collect and archive a tissue sample for each service member. If unidentified human remains are suspected to be those of an individual a sample from the remains can be analyzed for markers and compared to the markers present in the tissue sample initially collected from that individual.

Marker Assisted Breeding

[87] Genetic markers can assist breeders in the understanding, selecting and managing of the genetic complexity of animals and plants. Agriculture industry, for example, has a great deal of incentive to try to produce crops with desirable traits (high yield, disease resistance, taste, smell, color, texture, etc.) as consumer demand increases and expectations change. However, many traits, even when the molecular mechanisms are known, are too difficult or costly to monitor during production. Readily detectable polymorphisms which are in close physical proximity to the desired genes can be used as a proxy to determine whether the desired trait is present or not in a particular organism. This provides for an efficient screening tool which can accelerate the selective breeding process.

IDT Exhibit 1020.8921

EXAMPLE

Step 1. Restriction enzyme digestion of genomic DNA

[88] 2 μg human genomic DNA was digested with 20 units of *Eco*R I (New England Biolabs) and 20 units of *Sau*3A I in 80 μl 1x NEB buffer 4 with 1x BSA (NEB) by 37°C overnight incubation. The enzymes were inactivated at 65°C for 20 minute.

Step 2. Adaptor ligation

[89] 500 ng genomic DNA digested in step 1 was ligated to 30 pmol *Eco*R I adaptors (5'-AGTATAAGGGGTTCG-3' (SEQ ID NO 1) and 5'-phosphate-AATTACGAACCCCTTATACT-3' (SEQ ID NO 2)), which leaves a 1 nucleotide gap between the 3' adaptor and the restriction fragment, and 300 pmol *Sau*3A I adaptors (5'-ATCGCCTTATAGTGAGTCG-3' (SEQ ID NO 3) and 5'-CGACTCACTATAAGGC-3' (SEQ ID NO 4)), which leaves a 1 nucleotide gap between the 5' end of the adaptor and the restriction fragment, with 4000 units of T4 Ligase (NEB) in 1x T4 Ligase buffer (NEB) at 16°C for overnight. The ligase was inactivated by incubation at 65°C for 10 minutes.

[90] Because the adaptors were designed to leave a gap in one strand (see Figure 3), only fragments that were cleaved by both enzymes will ligate both the *Eco*R I and *Sau*3AI adaptor sequences to the same strand and be amplified in the following step.

Step 3. Amplification with PCR

[91] 4 ng genomic DNA ligated in step 2 was amplified with 2 units AmpliTaq Gold polymerase (PE) and 3 mM *Eco*R I primer (5'-AGTATAAGGGGTTCGTAATTCA-3'(SEQ ID NO 5)) and 3 mM *Sau3*A I primer (5'-ACTCACTATAAGGCGATCA-3'(SEQ ID NO 6)) in 50 μl 1x PCR buffer II (PE) with 2.5 mM $MgCl_2$. PCR was performed in a MJ Research thermal cycler (Watertown, MA) with 35 cycles of 94°C for 30 seconds, 48°C for 30 seconds and 72 °C for 2 minutes. 400 μl PCR reactions were pooled and purified and quantified by measuring their UV absorbance at 260 nm.

Step 4. Fragmentation, biotin labeling and hybridization to arrays

[92] 5 μg purified PCR product was digested in 40 μl 1x One-Phor-All buffer (Pharmacia) with 0.6 unit DNase I (Promega) at 37°C for 15 minutes. The DNase I was inactivated by boiling for 15 minutes. 2 μl biotin-N6-ddATP (1.0 mM, NEN)

IDT Exhibit 1020.8922

and 3 ul rTDT (15U/μl, GIBCO) were added to the DNA solution. The DNA fragments were labeled with biotin by incubation at 37°C for 1.5 hrs. Boiling for 15 minutes inactivated the enzyme. 20 μg DNA was pooled, concentrated, mixed with hybridization cocktail and hybridized to arrays on a rotisserie at 49°C overnight.

Step 5. Array washing, staining and scanning

[93] Arrays were washed with 6X SSPE buffer on a fluidics station (Affymetrix) and 0.1x MES buffer (Affymetrix) at 45°C for 30 minutes, stained and scanned under standard procedures.

CONCLUSION

[94] From the foregoing it can be seen that the present invention provides a flexible and scalable method for analyzing complex samples of DNA, such as genomic DNA. These methods are not limited to any particular type of nucleic acid sample: plant, bacterial, animal (including human) total genome DNA, RNA, cDNA and the like may be analyzed using some or all of the methods disclosed in this invention. This invention provides a powerful tool for analysis of complex nucleic acid samples. From experiment design to isolation of desired fragments and hybridization to an appropriate array, the above invention provides for fast, efficient and inexpensive methods of complex nucleic acid analysis.

[95] All publications and patent applications cited above are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication or patent application were specifically and individually indicated to be so incorporated by reference. Although the present invention has been described in some detail by way of illustration and example for purposes of clarity and understanding, it will be apparent that certain changes and modifications may be practiced within the scope of the appended claims.

IDT Exhibit 1020.8923

We claim:

1. A method of reducing the complexity of a nucleic acid sample comprising:

fragmenting the nucleic acid sample using a first and second restriction enzyme to produce fragments;

ligating adaptors to the fragments; and

selectively amplifying the fragments that were cut on one end by the first restriction enzyme and on the other end by the second restriction enzyme.

2. The method of claim 1 wherein the amplified fragments comprise at least 0.01% of the first nucleic acid sample.

3. The method of claim 1 wherein the amplified fragments comprise at least 0.5% of the first nucleic acid sample.

4. The method of claim 1 wherein the amplified fragments comprise at least 3% of the first nucleic acid sample.

5. The method of claim 1 wherein the amplified fragments comprise at least 12% of the first nucleic acid sample.

6. The method of claim 1 wherein the amplified fragments comprise at least 30% of the first nucleic acid sample.

7. The method of claim 1 wherein the amplified fragments comprise at least 50% of the first nucleic acid sample.

8. The method of claim 1 wherein the first restriction enzyme has a six base pair recognition sequence and the second restriction enzyme has a four base pair recognition sequence.

9. The method of claim 1 wherein the first restriction enzyme has an eight base pair recognition sequence and the second restriction enzyme has a four base pair recognition sequence.

IDT Exhibit 1020.8924

10. The method of claim 1 wherein the fragments are amplified by PCR.

11. The method of claim 1 wherein the nucleic acid sample is genomic DNA, DNA, cDNA derived from RNA, total RNA or mRNA.

12. The method of claim 1 wherein ligation of one strand of each adaptor is blocked.

13. The method of claim 12 wherein ligation is blocked by introducing a gap of at least one nucleotide between one strand of the adaptor and one strand of the fragment.

14. The method of claim 12 wherein ligation is blocked by the absence of a phosphate at the 5' end of an adaptor strand.

15. The method of claim 12 wherein ligation is blocked by the presence of a modified nucleotide at the 5' or 3' end of an adaptor strand.

16. The method of claim 12 wherein ligation is blocked by a terminal modification in one strand of an adaptor.

17. The method of claim 12 wherein ligation is blocked at the 5' end of one strand of one adaptor and at the 3' end of one strand of the other adaptor.

18. The method of claim 12 wherein ligation is blocked at the 5' end of one adaptor and at the 3' end of the other adaptor.

19. The method of claim 1 wherein one adaptor comprises a 5' overhang comprising a primer binding site and the other adaptor comprises a 3' overhang comprising a primer binding site.

20. A method for analyzing a nucleic acid sample comprising:

fragmenting the nucleic acid sample using a first and second restriction enzyme to produce fragments;

IDT Exhibit 1020.8925

ligating adaptors to the fragments;

selectively amplifying the fragments that were cut on one end by the first restriction enzyme and on the other end by the second restriction enzyme;

providing a nucleic acid array;

hybridizing the amplified fragments to the array; and

analyzing a hybridization pattern resulting from the hybridization.

21. The method of claim 20 wherein the method for analyzing the nucleic acid sample comprises determining whether the nucleic acid sample contains sequence variations.

22. The method of claim 21 wherein the sequence variations are single nucleotide polymorphisms (SNPs).

23. The method of claim 20 wherein the nucleic acid array is designed to query DNA fragments which have been produced by the procedures used to obtain the amplified fragments.

24. The method of claim 20 wherein a substantial amount of the sequences predicted to be contained in the amplified fragments are first determined by a computer system.

25. A method of screening for DNA sequence variations in an individual comprising:

providing a nucleic acid sample from the individual;

fragmenting the nucleic acid sample using a first and second restriction enzyme to produce fragments;

ligating adaptors to the fragments; and

selectively amplifying the fragments that were cut on one end by the first restriction enzyme and on the other end by the second restriction enzyme;

providing a nucleic acid array;

hybridizing the amplified fragments to the array;

generating a hybridization pattern resulting from the hybridization; and

IDT Exhibit 1020.8926

determining the presence or absence of DNA sequence variations in the individual based upon an analysis of the hybridization pattern.

26. The method of claim 25 wherein the sequence variation is at least one single nucleotide polymorphism (SNP).

27. The method of claim 26 wherein the at least one SNP is associated with a disease.

28. The method of claim 26 wherein the at least one SNP is associated with the efficacy of a drug.

29. A method of screening for DNA sequence variation in a population of individuals comprising:

providing a first nucleic acid sample from each of the individuals;

providing a second nucleic acid sample by:

fragmenting the first nucleic acid sample using a first and second restriction enzyme to produce fragments;

ligating adaptors to the fragments; and

selectively amplifying the fragments that were cut on one end by the first restriction enzyme and on the other end by the second restriction enzyme;

providing a plurality of identical nucleic acid arrays wherein the arrays comprise probes designed to interrogate for DNA sequence variations;

hybridizing each of the second nucleic acid samples to one of the plurality of identical arrays;

generating a plurality of hybridization patterns resulting from the hybridizations; and

analyzing the hybridization patterns to determine the presence or absence of sequence variation in the population of individuals.

30. The method of claim 29 wherein the sequence variation is at least one single nucleotide polymorphism (SNP).

IDT Exhibit 1020.8927

31. The method of claim 27 wherein the at least one SNP is associated with a disease.

32. The method of claim 27 wherein the at least one SNP is associated with the efficacy of a drug.

33. A method of genotyping an individual comprising:

providing a first nucleic acid sample from the individual;

providing a second nucleic acid sample by:

fragmenting the first nucleic acid sample using a first and second restriction enzyme to produce fragments;

ligating adaptors to the fragments; and

selectively amplifying the fragments that were cut on one end by the first restriction enzyme and on the other end by the second restriction enzyme;

hybridizing the second nucleic acid sample to an array designed to determine the presence or absence of one or more alleles of a collection of SNPs;

generating a hybridization pattern resulting from the hybridization; and

determining the presence or absence of the one or more alleles of the collection of SNPs.

34. A kit for reducing the complexity of a nucleic acid sample comprising:

buffers and restriction enzymes for fragmenting a nucleic acid sample,

a ligase and adaptors to be ligated to the fragments, the adaptors being designed for the selective amplification of the fragments that were cut on one end by a first restriction enzyme and on the other end by a second restriction enzyme,

primers for the selective amplification, and

instructions for the use of the kit.

IDT Exhibit 1020.8928

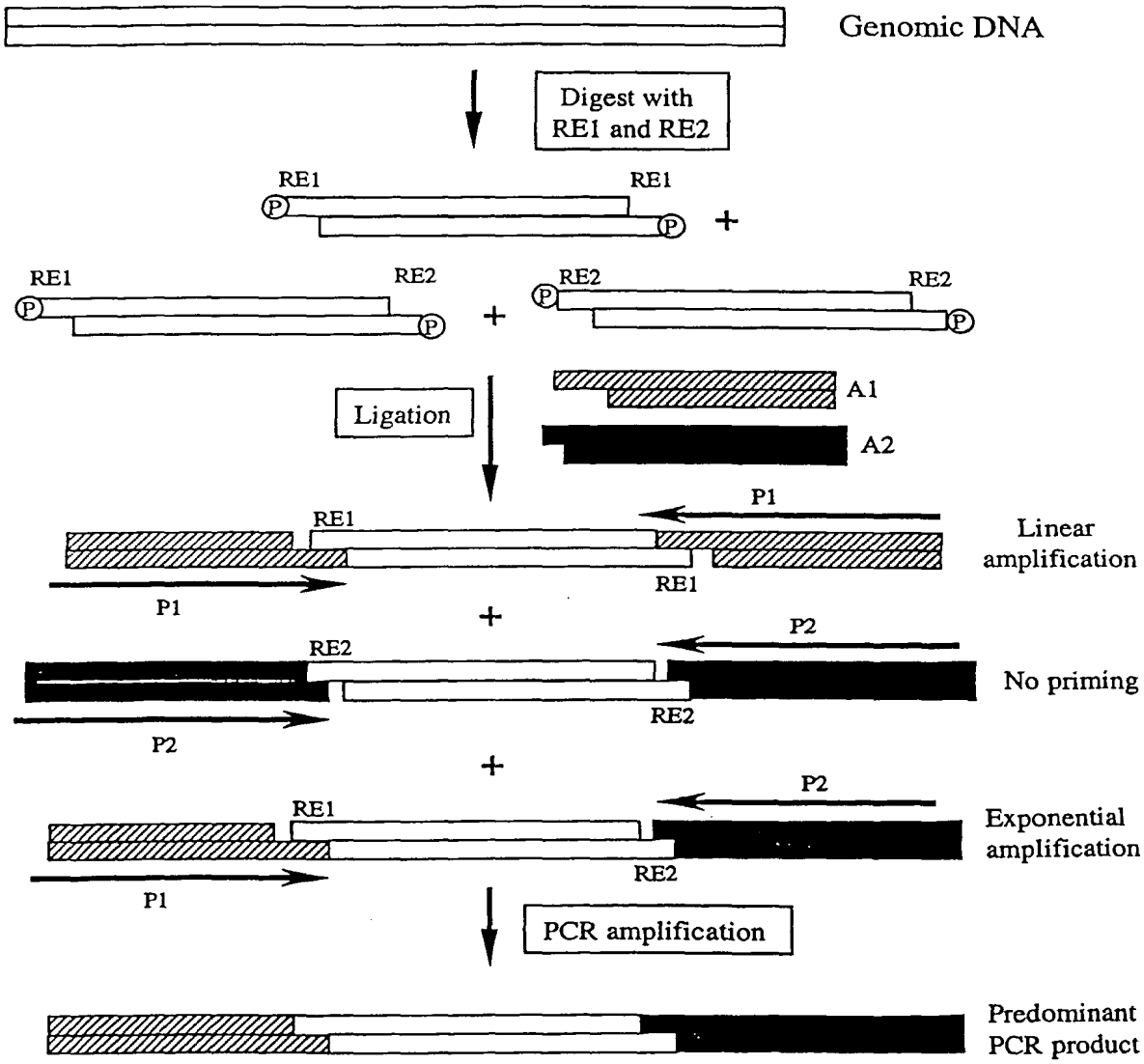

Genomic DNA
Digest with RE1 and RE2
RE1
RE1
RE1
RE2
RE2
RE2
Ligation
A1
A2
P1
RE1
RE1
P1
Linear amplification
P2
RE2
RE2
P2
No priming
P2
RE1
RE2
P1
Exponential amplification
PCR amplification
Predominant PCR product

Figure 1

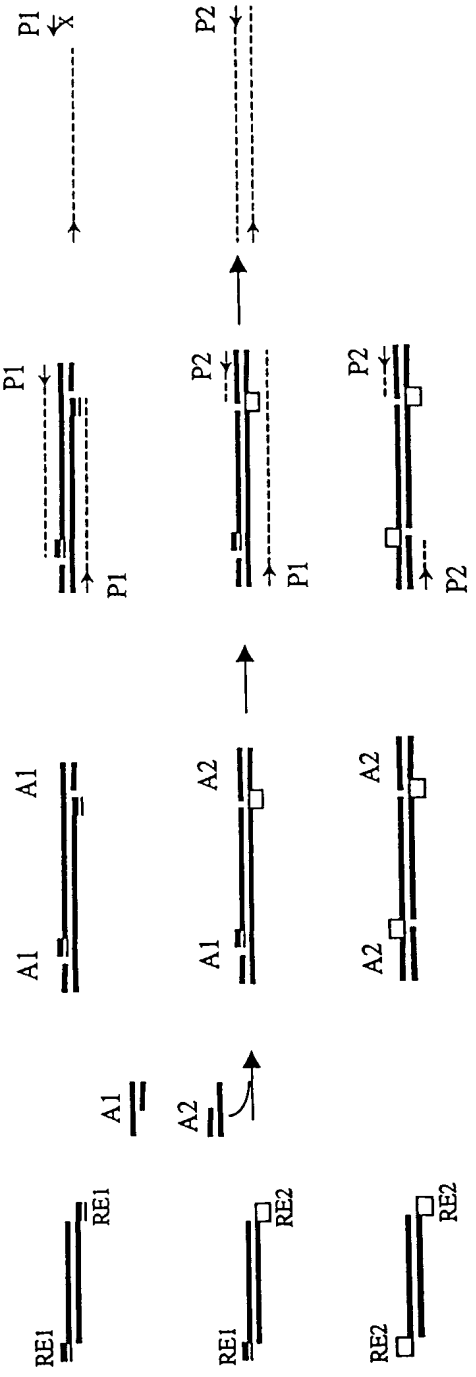
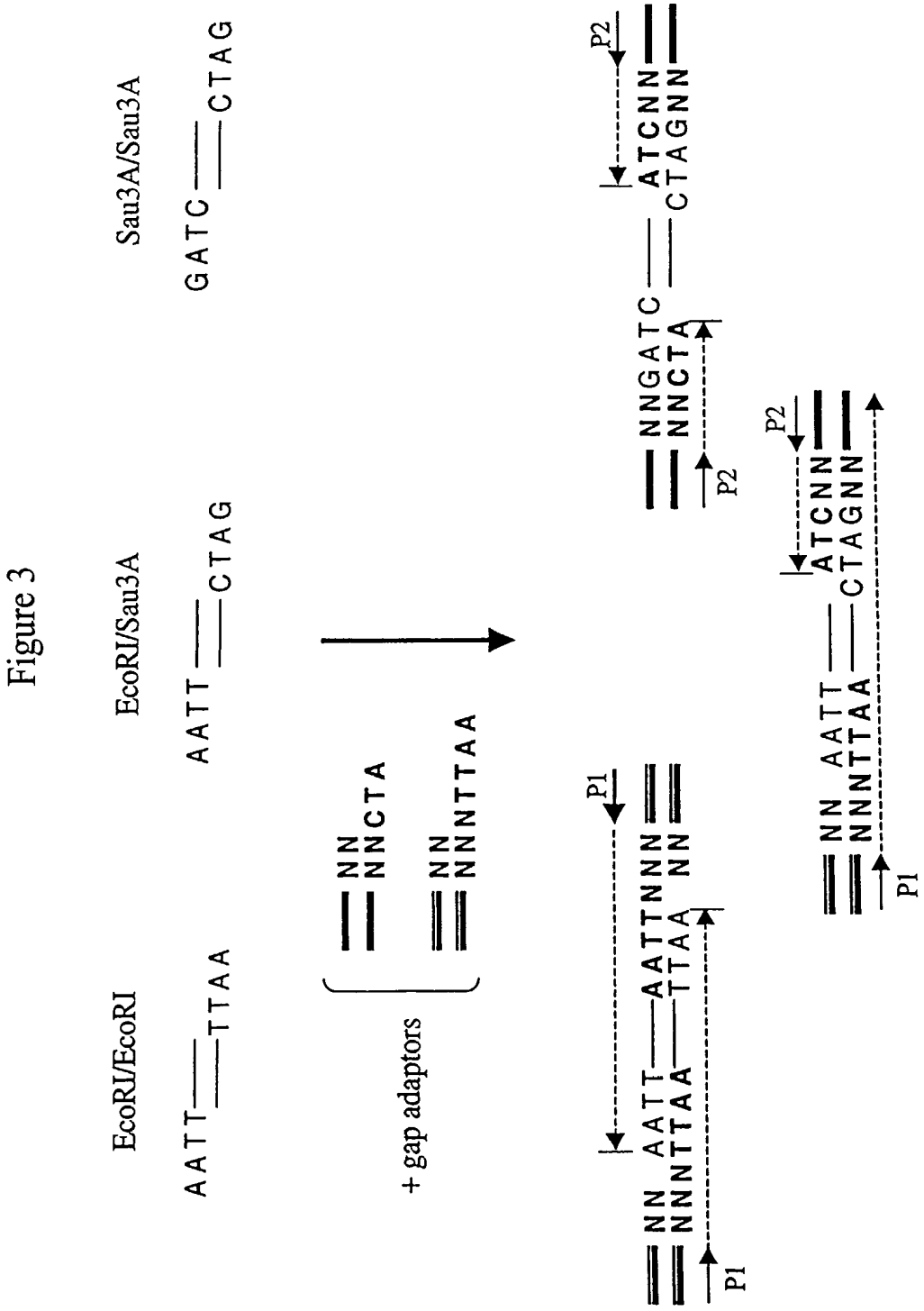

P1
P2
P1
P2
P2
P1
P1
P2
A1
A1
A2
A1
A2
A2
A1
A2
RE1
RE1
RE2
RE1
RE2
RE2

Figure 2

Figure 3


EcoRI/EcoRI
AATT
TTAA
EcoRI/Sau3A
AATT
CTAG
Sau3A/Sau3A
GATC
CTAG
+ gap adaptors
NN
NNCTA
NN
NNNTTAA
P1
P2
NNGATC
NNCTA
ATCNN
CTAGNN
AATTNNN
TTAA NN
NN AATT
NNNTTAA
NN AATT
NNNTTAA
ATCNN
CTAGNN

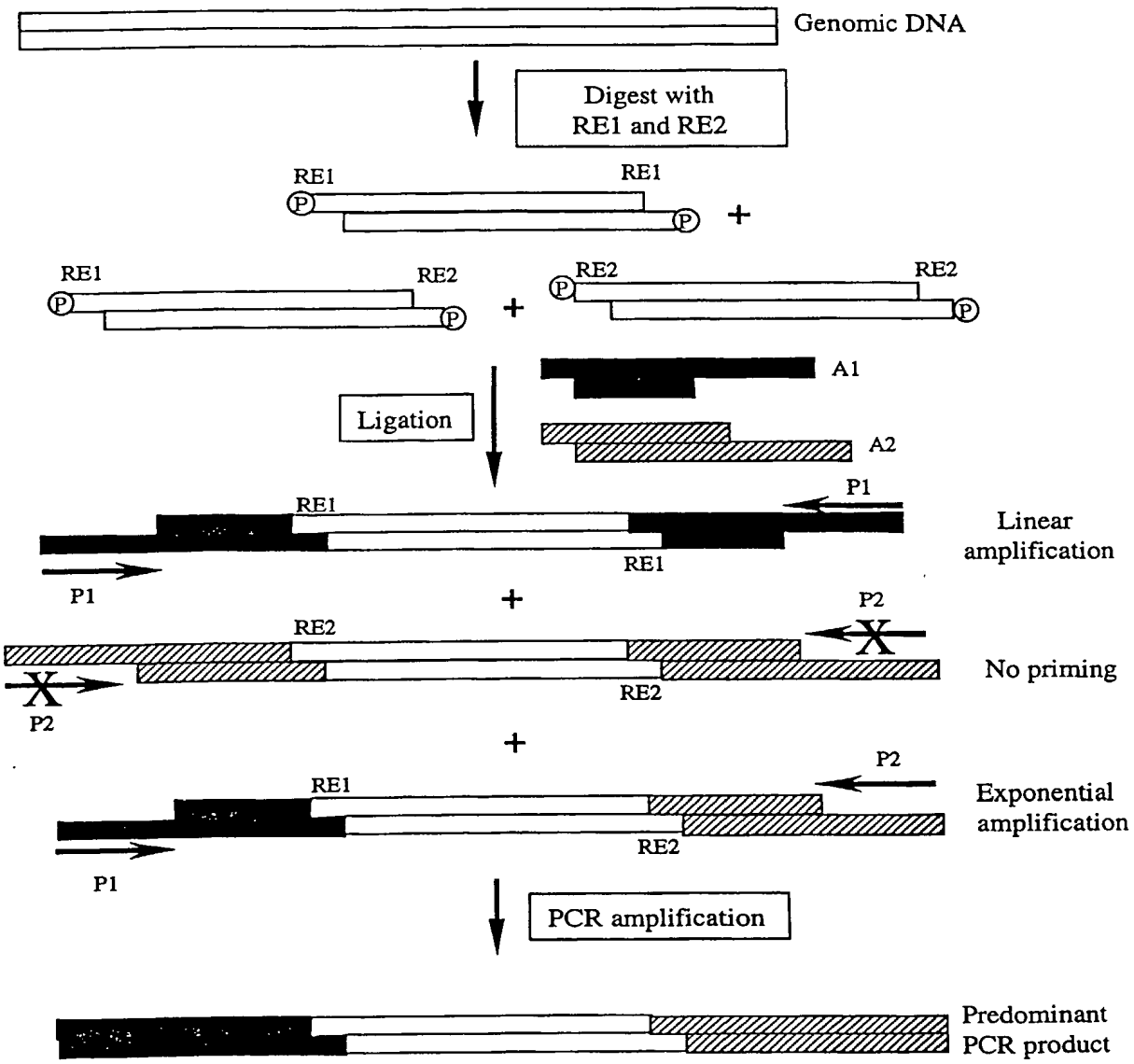

Genomic DNA
Digest with RE1 and RE2
RE1
RE1
RE1
RE2
RE2
RE2
A1
A2
Ligation
P1
RE1
RE1
Linear amplification
P1
P2
RE2
RE2
No priming
P2
P2
RE1
RE2
Exponential amplification
P1
PCR amplification
Predominant PCR product

Figure 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/23568

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12P 19/34
US CL : 435/91.1

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/91.1

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	SMITH. D.R, Ligation-mediated PCR of restriction fragments from large DNA molecules. PCR Methods and Applications. 1992. Vol 2. No 1, pages 21-27., see entire document.	1-7, 10-11 ---------- 8-9, 20-34
Y	US 6,156,502 A (BEATTIE et al) 05 December 2000 (5.12.2000), column 3, lines 13-28,see entire document.	20-34

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

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

"E" earlier application or patent 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
19 September 2002 (19.09.2002)	03 OCT 2002
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703)305-3230	Authorized officer Gary Jones Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

IDT Exhibit 1020.8933

INTERNATIONAL SEARCH REPORT

PCT/US02/23568

Continuation of B. FIELDS SEARCHED Item 3:
databases: MEDLINE, BIOSIS, CAPLUS, US PATENT DATABASE
search terms: DNA, amplification, amplify, amplified, ligate, ligated, ligation, adapter, adaptor, specific, selective, digest, digestion, restriction, restrict, endonuclease, block, blocked, extend, extension, polymerization, prevent, prevention

IDT Exhibit 1020.8934

Espacenet

Bibliographic data: WO0164952 (A2) — 2001-09-07

METHOD FOR REDUCING ARTIFACTS IN NUCLEIC ACID AMPLIFICATION

Inventor(s): DEAN FRANK B; FARUQI A FAWAD ± (DEAN, FRANK, B, ; FARUQI, A., FAWAD)

Applicant(s): MOLECULAR STAGING INC [US] ± (MOLECULAR STAGING, INC)

Classification: - international: ***C12N15/09; C12Q1/68;*** (IPC1-7): C12Q1/68
- cooperative: **C12Q1/6844**; **C12Q1/6853**

Application number: WO2001US06491 20010228 Global Dossier

Priority number (s): US20000514113 20000228

Also published as: WO0164952 (A3) AU4186401 (A) CA2401650 (A1) EP1294933 (A2) JP2003525055 (A) more

Abstract of WO0164952 (A2)

Disclosed are compositions and methods useful for reducing the formation of artifacts during nucleic acid amplification reactions. The method uses special oligonucleotides, referred to herein as template-deficient oligonucleotides, that cannot serve as a template for nucleic acid synthesis over part of their length. This prevents the oligonucleotides from serving as effective templates in the formation of artifacts. The disclosed method involves using a template-deficient oligonucleotide as at least one of the oligonucleotides (preferably a primer) in a nucleic acid amplification reaction, where the template-deficient oligonucleotide comprises one or more template-deficient nucleotides, preferably at or near the 5' end of the template-deficient oligonucleotide. The disclosed method is useful for reducing artifacts in any nucleic acid amplification reaction involving oligonucleotides. In a preferred form of the method the nucleic acid amplification reaction does not involve thermal cycling. The disclosed method is effective at reducing non-cycle oligonucleotide-based artifacts. Also disclosed are kits useful for reducing artifacts in nucleic acid amplification reactions. The disclosed kits include a template-deficient oligonucleotide, wherein the template-deficient oligonucleotide comprises one or more template-deficient nucleotides, and a nucleic acid polymerase.

IDT Exhibit 1020.8935

IDT Exhibit 1020.8936

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 September 2001 (07.09.2001)

PCT

(10) International Publication Number
WO 01/64952 A2

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/US01/06491

(22) International Filing Date: 28 February 2001 (28.02.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/514,113 28 February 2000 (28.02.2000) US

(71) Applicant: **MOLECULAR STAGING, INC.** [US/US]; 300 George Street, New Haven, CT 06511 (US).

(72) Inventors: **DEAN, Frank, B.**; 21A Pearl Street, Guilford, CT 06437 (US). **FARUQI, A., Fawad**; 985 Durham Road, Guilford, CT 06437 (US).

(74) Agents: **PERRYMAN, David, G.** et al.; Needle & Rosenberg, P.C., Suite 1200, 127 Peachtree Street, N.E., Atlanta, GA 30303-1811 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, 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 01/64952 A2

(54) Title: METHOD FOR REDUCING ARTIFACTS IN NUCLEIC ACID AMPLIFICATION

(57) Abstract: Disclosed are compositions and methods useful for reducing the formation of artifacts during nucleic acid amplification reactions. The method uses special oligonucleotides, referred to herein as template-deficient oligonucleotides, that cannot serve as a template for nucleic acid synthesis over part of their length. This prevents the oligonucleotides from serving as effective templates in the formation of artifacts. The disclosed method involves using a template-deficient oligonucleotide as at least one of the oligonucleotides (preferably a primer) in a nucleic acid amplification reaction, where the template-deficient oligonucleotide comprises one or more template-deficient nucleotides, preferably at or near the 5' end of the template-deficient oligonucleotide. The disclosed method is useful for reducing artifacts in any nucleic acid amplification reaction involving oligonucleotides. In a preferred form of the method the nucleic acid amplification reaction does not involve thermal cycling. The disclosed method is effective at reducing non-cycle oligonucleotide-based artifacts. Also disclosed are kits useful for reducing artifacts in nucleic acid amplification reactions. The disclosed kits include a template-deficient oligonucleotide, wherein the template-deficient oligonucleotide comprises one or more template-deficient nucleotides, and a nucleic acid polymerase.

IDT Exhibit 1020.8937

METHOD FOR REDUCING ARTIFACTS IN NUCLEIC ACID AMPLIFICATION

BACKGROUND OF THE INVENTION

The present invention is in the field of nucleic acid amplification, and specifically in the area of reducing amplification artifacts in nucleic acid amplification reactions.

Numerous nucleic acid amplification techniques have been devised, including strand displacement cascade amplification (SDCA)(referred to herein as exponential rolling circle amplification (ERCA)) and rolling circle amplification (RCA)(U.S. Patent No. 5,854,033; PCT Application No. WO 97/19193; Lizardi *et al.*, *Nature Genetics* 19(3):225-232 (1998)); multiple displacement amplification (MDA)(PCT Application WO 99/18241); strand displacement amplification (SDA)(Walker *et al.*, Nucleic Acids Research 20:1691-1696 (1992), Walker *et al.*, Proc. Natl. Acad. Sci. USA 89:392-396 (1992)); polymerase chain reaction (PCR) and other exponential amplification techniques involving thermal cycling, self-sustained sequence replication (3SR), nucleic acid sequence based amplification (NASBA), and amplification with Qβ replicase (Birkenmeyer and Mushahwar, *J. Virological Methods* 35:117-126 (1991); Landegren, *Trends Genetics* 9:199-202 (1993)); and various linear amplification techniques involving thermal cycling such as cycle sequencing (Craxton *et al.*, *Methods Companion Methods in Enzymology* 3:20-26 (1991)).

Artifacts--that is, unwanted, unexpected, or non-specific nucleic acid molecules--have been observed in almost all nucleic acid amplification reactions. For example, Stump *et al.*, *Nucleic Acids Research* 27:4642-4648 (1999), describes nucleic acid artifacts resulting from an illegitimate PCR process during cycle sequencing. Stump *et al.* suggests a way to avoid such artifacts by using certain primers that cannot be fully replicated. Watson, *Amplifications*, 5-6 (1989), and Ferrie *et al.*, *Am. J. Hum. Genet.* 51:251-262 (1992), describe formation of primer-dimer artifacts during PCR. Brownie *et al.*, *Nucleic Acids Research* 25:3235-3241 (1997), suggests a way to avoid primer dimer formation during PCR by adding tails at the 5' end of the PCR primers that results in formation of a non-replicable structure if primer-dimer formation is

initiated. Other forms of artifacts can occur in other types of nucleic acid amplification techniques.

Therefore, it is an object of the present invention to provide a method of reducing, preventing, or eliminating artifacts in nucleic acid amplification reactions.

It is another object of the present invention to provide oligonucleotides that, when used in a nucleic acid amplification reaction, can reduce, prevent, or eliminate artifacts in the nucleic acid amplification reaction.

It is another object of the present invention to provide kits for nucleic acid amplification that can reduce, prevent, or eliminate artifacts in the nucleic acid amplification reaction.

BRIEF SUMMARY OF THE INVENTION

Disclosed are compositions and methods useful for reducing the formation of artifacts during nucleic acid amplification reactions. The method uses special oligonucleotides, referred to herein as template-deficient oligonucleotides, that cannot serve as a template for nucleic acid synthesis over part of their length. This prevents the oligonucleotides from serving as effective templates in the formation of artifacts. The disclosed method involves using a template-deficient oligonucleotide as at least one of the oligonucleotides (preferably a primer) in a nucleic acid amplification reaction, where the template-deficient oligonucleotide comprises one or more template-deficient nucleotides, preferably at or near the 5' end of the template-deficient oligonucleotide. The disclosed method is useful for reducing artifacts in any nucleic acid amplification reaction involving oligonucleotides. In a preferred form of the method the nucleic acid amplification reaction does not involve thermal cycling. The disclosed method is effective at reducing non-cycle oligonucleotide-based artifacts. Also disclosed are kits useful for reducing artifacts in nucleic acid amplification reactions. The disclosed kits include a template-deficient oligonucleotide, wherein the template-deficient oligonucleotide comprises one or more template-deficient nucleotides, and a nucleic acid polymerase.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a diagram of the proposed "templating" process leading to formation of artifacts.

IDT Exhibit 1020.8939

Figure 2 is a diagram of an oligonucleotide with an abasic nucleotide at the 5' end showing that it cannot serve as a template for synthesis of a complete complementary strand.

Figure 3 is a diagram of an oligonucleotide with two abasic nucleotides at the 5' end showing that it cannot serve as a template for synthesis of a complete complementary strand.

Figure 4 is a diagram showing how primer dimers could be formed during PCR through non-template (that is, overhanging) addition of an adenosine at the 3' end of the synthesized strand.

Figure 5 is a diagram of various structures of examples of the disclosed template-deficient oligonucleotides (oligonucleotides 2-15). The region of each oligonucleotide that is template-deficient, template-capable, primer-deficient, and primer-capable is indicated. For comparison, oligonucleotide 1 is a normal oligonucleotide (that is, composed entirely of template- and primer-capable nucleotides).

Figure 6 is a graph of amount of nucleotide incorporated (in pmol/4 μl) versus time (in minutes) during exponential rolling circle amplification reactions using various combinations of control primers (P1 and P2) and abasic primers (P1(aba) and P2(aba)). The abasic primers had two abasic nucleotides at the 5' end. The primers used (and whether a template circle was present) are indicated by annotations to the right of the graph next to the last data point in each curve. The five annotations at the bottom (P1(aba) + P2(aba) - circle, P1 alone, P2 alone, P1(aba) alone, and P2(aba) alone) represent reactions having identical curves showing no nucleotide incorporation. The reaction number, as described in Example 1, represented by each curve is indicated in parentheses next to the annotations.

Figure 7 is a graph of amount of nucleotide incorporated (in pmol/4 μl) versus time (in hours) during exponential rolling circle amplification reactions using various combinations of control primers (P1 and P2) and template-deficient primers (P1aba, P2aba, P1C, and P4C). The P1aba and P2aba primers had two abasic nucleotides at the 5' end. The P1C and P4C primers were chimeric, having 6 deoxyribonucleotides at the 3' end and 18 ribonucleotides at the 5' end. The primers used (and whether a template

IDT Exhibit 1020.8940

circle was present) are indicated by annotations to the right of the graph next to the last data point in each curve. The five annotations at the bottom (P1aba + P2aba, P1C + P4C, P1 + P4C, P1aba + P4C, and P1C + P2aba) represent reactions having identical curves showing no nucleotide incorporation.

DETAILED DESCRIPTION OF THE INVENTION

Nucleic acid amplification techniques using primers are often compromised by the formation of nucleic acid artifacts that are not the product of legitimate amplification. For example, exponential rolling circle amplification (ERCA), an isothermal amplification reaction using two primers complementary to opposite strands of a nucleic acid molecule to be amplified (see strand displacement cascade amplification in PCT Application No. WO 97/19193), sometimes results in formation of oligonucleotide-based artifacts. These artifacts, referred to herein as non-cycle oligonucleotide-based artifacts, are different in structure from primer dimer artifacts that can occur in amplification reactions involving thermal cycling and result from a different mechanism.

For example, artifactual DNA products synthesized by Bst polymerase during ERCA consist of DNA molecules that are long, many kilobase in length, yet occur in the absence of template, or any DNA of like size that could act as a template to yield a product of such a size. The artifact DNA synthesis can occur efficiently in the presence of just two short oligonucleotide primers and Bst alone. The artifactual DNA appears as a collection of products varying greatly in size that appears as a broad smear, or ladder of closely spaced bands, when analyzed by gel electrophoresis.

It has been discovered that non-cycle oligonucleotide-based artifacts could be reduced by using oligonucleotides that could not be replicated to their full length, thereby preventing synthesis of a duplex, blunt-ended structure. Such oligonucleotides are referred to herein as template-deficient oligonucleotides. Such oligonucleotides would be poorly extended out to the ends and would only poorly be able to extend across the nick that is created by the insertion of an interloping oligonucleotide (see Figure 2).

Disclosed are compositions and methods useful for reducing the formation of primer artifacts during nucleic acid amplification reactions. The method uses special

IDT Exhibit 1020.8941

oligonucleotides, referred to herein as template-deficient oligonucleotides, that cannot serve as a template for nucleic acid synthesis over part of their length. This prevents the oligonucleotides from serving as effective templates in the formation of artifacts during a nucleic acid amplification reaction. It is preferred that the template-deficient oligonucleotide be a primer, that all of the primers used in the nucleic acid amplification reaction are template-deficient, and/or that all of the oligonucleotides used in the nucleic acid amplification reaction are template-deficient.

In one form, the disclosed method involves using a template-deficient oligonucleotide where the template-deficient oligonucleotide includes one or more template-deficient nucleotides, where the number and composition of template-capable nucleotides 3' of the template-deficient nucleotide closest to the 3' end of the template-deficient oligonucleotide is sufficient to allow the template-capable nucleotides 3' of the template-deficient nucleotide closest to the 3' end alone to effectively prime nucleic acid synthesis in the nucleic acid amplification reaction. Template-deficient oligonucleotides of this form are referred to herein as 3' open oligonucleotides. Template-deficient oligonucleotides referred to herein as 3' closed oligonucleotides are template-deficient oligonucleotides where the template-deficient oligonucleotide includes one or more template-deficient nucleotides, where the number and composition of template-capable nucleotides 3' of the template-deficient nucleotide closest to the 3' end of the template-deficient oligonucleotide is insufficient to allow the template-capable nucleotides 3' of the template-deficient nucleotide closest to the 3' end alone to effectively prime nucleic acid synthesis in the nucleic acid amplification reaction. Put another way, placing a template-deficient nucleotide at or near the 3' end of an oligonucleotide results in a 3' closed oligonucleotide.

In another form, the disclosed method involves using a template-deficient oligonucleotide where the template-deficient oligonucleotide includes one or more template-deficient nucleotides.

Numerous configurations of template-deficient nucleotides can be used in the disclosed template-deficient oligonucleotides. For example, when the template-deficient oligonucleotide contains one or more template-deficient nucleotides, the template-deficient nucleotides can be at or near the 5' end of the template-deficient

IDT Exhibit 1020.8942

oligonucleotide. When the template-deficient oligonucleotide contains two or more template-deficient nucleotides, at least two of the template-deficient nucleotides can be adjacent. In one embodiment, the adjacent template-deficient nucleotides can be within three nucleotides of the 5' end of the template-deficient oligonucleotide. In general, no particular configuration of template-deficient nucleotides is required.

Template-deficient nucleotides are selected from the group consisting of modified nucleotides, derivatized nucleotides, ribonucleotides, and nucleotide analogs. Preferred template-deficient nucleotides are modified nucleotides. Preferred modified nucleotides are abasic nucleotides. Template-deficient nucleotides include abasic nucleotides, nucleotides with an inverted base, fluoro substituted nucleotides, alkyl substituted nucleotides, nucleotides with phenyl substituted ethers, nucleotides with substituted thioethers, nucleotides with phosphate esters, α-nucleotides, 2',3'-dideoxy nucleotides, ribonucleotides, nucleotides derivatized with biotin, nucleotides derivatized with amine, nucleotides derivatized with Hex, nucleotides derivatized with Tet, nucleotides derivatized with Fam, nucleotides derivatized with fluorescein, nucleotides derivatized with rhodamine, nucleotides derivatized with alkaline phosphatase, nucleotides derivatized with horseradish peroxidase, nucleotides derivatized with spacers, nucleotides derivatized with cholesteryl, nucleotides derivatized with DNP-TEG, nucleotides derivatized with psoralen cross-linkers, nucleotides derivatized with intercalating agents, and nucleotides derivatized with PNA conjugates.

Examples of such nucleotides include abasic nucleosides (Biegelman *et al.*, *Bioorganic & Medicinal Chemistry Letters* 4(14):1715-1720 (1994); Moran *et al.*, *Nucleic Acids Res.* 24(11):2044-2052 (1996); Matray and Kool, *Nature* 399:704-708 (1999)), 5'-fluoro substituted nucleosides (Robins and Wnuk, *Tetrahedron Lett.* 29:5729 (1988)), 5'-alkyl substituted nucleosides (Ray and Jaxa-Chamiec, *Heterocycles* 31(10):1777-1780 (1990); Jun-Dong and Li-He, *Synthesis* 909-911 (1990); Tanaka *et al.*, *Tetrahedron Lett.* 30:2567-2570 (1989)), nucleosides with 5'-alkyl or phenyl substituted ethers (Jones *et al.*, *Carbohydrates, Nucleosides, Nucleotides* 4:301 (1977)), 5'-substituted thioethers (Connolly and Rider, *Nucleic Acids Res.* 13:4485 (1985); Connolly, *Nucleic Acids Res.* 15:3131-3139 (1987); Sinha and Cook, *Nucleic Acids*

IDT Exhibit 1020.8943

Res. 16:2659 (1988); Kumar *et al.*, *Nucleic Acids Res.* 19:4561 (1991); Zuckermann *et al.*, *Nucleic Acids Res.* 15:5305 (1987); Gupta *et al.*, Tetrahedron Lett. 31:2471-2474 (1990); Asslie *et al.*, *Tetrahedron* 48:1233-1254 (1992)), 5'-amines and substituted amines (Connolly and Rider, *Nucleic Acids Res.* 13:4485 (1985); Sproat *et al.*, *Nucleic Acids Res.* 15:4857 (1987); Zuckerman *et al.*, *Nucleic Acids Res.* 15:5305 (1987), Li *et al.*, *Nucleic Acids Res.* 15:5275 (1987); Dreyer and Dervan, *Proc. Natl. Acad. Sci. USA* 82:968 (1985)), phosphate esters as 5'-terminators (Tanaka *et al.*, *Tetrahedron Lett.* 30:2567-2570 (1989)), inverted bases or α-nucleosides as 5'-terminators (Bloch *et al.*, *Gene* 72:349 (1988); Sequin, *Helv. Chim. Acta.* 57:68 (1974)), 2',3'-dideoxy nucleosides as 5'-terminators (Huryn and Okabe, *Chem. Rev.* 92:1745-1768 (1992)).

The nucleotides or oligonucleotides can also be derivatized with, for example, biotin, dyes such as fluorescein or rhodamine, or proteins such as alkaline phosphatase or horseradish peroxidase. 5'-modifications useful in the disclosed oligonucleotides include 5'-spacers (Durard *et al.*, *Nucleic Acids Res.* 18:6353 (1990); Salunkhe *et al.*, *J. Amer. Chem. Soc.* 114:8768-8772 (1992); Dolinnaya *et al.*, *Nucleic Acids. Res.* 21:5403-5407 (1993); Takeshita *et al.*, *J. Biol. Chem.* 262:10171-10179 (1987); Kalin *et al.*, *Biochemistry* 27:924-931 (1998)), 5'-biotinylated primers (Cocuzza, *Tetrahedron Lett.* 30:6287-6290 (1989); Nelson *et al.*, *Nucleic Acids Res.* 20:6253-6259 (1992)), 5'-cholesteryl (Mackellar *et al.*, *Nucleic Acids. Res.* 20:3411-3417 (1992); Stein *et al.*, Biochemistry 30:2439-2444 (1991)), 5'-DNP-TEG (Will *et al.*, *Carbohydrate Research* 216:315-322 (1991); Grzybowski *et al.*, *Nucleic Acids Res.* 21:1705-1712 (1993)), 5'-psoralen cross-linkers (Pieles and Englisch, *Nucleic Acids Res.* 17:285 (1989); Taksugi *et al.*, *Proc. Natl. Acad. Sci. USA* 88:5602-5606 (1991)), 5'-intercalating agents (Thoung and Chassignol, *Tetrahedron Lett.* 29:5905 (1988)), 5'-PNA conjugates (Nielsen *et al.*, *Science* 254:1497-1500 (1991); Egholm *et al.*, *J. Am. Chem. Soc.* 114:1895-1897 (1992)), 5'-enzyme conjugates (Jablonski *et al.*, *Nucleic Acids. Res.* 14:6115-6128 (1986)), 5'-dye-label (Molecular Probes, Eugene, Oreg.; Research Organics, Cleveland, Ohio).

It is not required that all of the template-deficient nucleotides in a template-deficient oligonucleotide be the same type of nucleotide. For example, a template-deficient oligonucleotide can include both abasic nucleotides and ribonucleotides as

IDT Exhibit 1020.8944

template-deficient nucleotides. Such template-deficient nucleotides are template-deficient for different reasons (the abasic nucleotide has no base to serve as a template while the ribonucleotide is not recognized as a template by the polymerase). There are no restrictions on the combinations of different template-deficient nucleotides that can be used in the same template-deficient oligonucleotide. Similarly, it is not required that all template-deficient oligonucleotides used in a nucleic acid amplification reaction have the same types or patterns of template-deficient nucleotides. For example, a primer containing abasic nucleotides can be used with a primer containing inverted nucleotides in the same amplification reaction.

The disclosed template-deficient oligonucleotides can be used with any method of nucleic acid amplification. Preferred forms of nucleic acid amplification for use of the disclosed oligonucleotides include nucleic acid amplification reactions involving exponential amplification, either isothermal or with thermal cycling, nucleic acid amplification reactions requiring exponential amplification, either isothermal or with thermal cycling, nucleic acid amplification reactions involving isothermal linear amplification, nucleic acid amplification reactions requiring isothermal linear amplification, nucleic acid amplification reactions involving rolling circle amplification, nucleic acid amplification reactions involving the polymerase chain reaction, and nucleic acid amplification reactions not involving thermal cycling.

Examples of nucleic acid amplification reactions are exponential rolling circle amplification (ERCA)(referred to as strand displacement cascade amplification in PCT Application No. WO 97/19193 and as hyperbranched rolling circle amplification in Lizardi *et al.*, *Nature Genetics* 19(3):225-232 (1998)) and rolling circle amplification (RCA)(U.S. Patent No. 5,854,033; PCT Application No. WO 97/19193; Lizardi *et al.*, *Nature Genetics* 19(3):225-232 (1998)); multiple displacement amplification (MDA)(PCT Application WO 99/18241); strand displacement amplification (SDA)(Walker *et al.*, Nucleic Acids Research 20:1691-1696 (1992), Walker *et al.*, Proc. Natl. Acad. Sci. USA 89:392-396 (1992)); nucleic acid sequence based amplification (NASBA)(Compton, *Nature* 350:91-92 (1991)); transcription-mediated amplification (TMA)(Nelson, *Crit Rev Clin Lab Sci* 35:369-414 (1998)); polymerase chain reaction (PCR) and other exponential amplification techniques involving thermal

IDT Exhibit 1020.8945

cycling, self-sustained sequence replication (3SR), and amplification with Qβ replicase (Birkenmeyer and Mushahwar, *J. Virological Methods* 35:117-126 (1991); Landegren, *Trends Genetics* 9:199-202 (1993)); various linear amplification techniques involving thermal cycling such as cycle sequencing (Craxton *et al.*, *Methods Companion Methods in Enzymology* 3:20-26 (1991)). The disclosed oligonucleotides and method would be useful in techniques described in Lizardi *et al.*, *Nature Genetics* 3:225-232 (1998), Thomas *et al.*, *Arch Pathol Lab Med* 123:1170-1176 (1999), Baner *et al.*, *Nucleic Acids Research* 26:5073-5078 (1998), and Zhang *et al.*, *Gene* 211:277-285 (1998).

When using 3' closed oligonucleotides as primers, it is preferred that the nucleic acid amplification reaction not involve cycle sequencing, the nucleic acid amplification reaction not require linear amplification via thermal cycling, the nucleic acid amplification reaction not involve linear amplification via thermal cycling, or the nucleic acid amplification reaction involve exponential amplification via thermal cycling.

Also disclosed are kits useful for reducing artifacts in nucleic acid amplification reactions. The disclosed kits include a template-deficient oligonucleotide as described herein and a polymerase. Any form of template-deficient primer as disclosed herein can be included in the disclosed kits.

While not wishing to be limited to any particular mechanism, it is believed that non-cycle oligonucleotide-based artifacts could be generated in the following manner. First, mispriming occurs such that, for example, one oligonucleotide partially anneals to a second oligonucleotide and is extended to the end of the oligonucleotide (see Figure 1). This blunt-ended product may sit in the polymerase active site for an extended period of time. The enzyme binding site for the single-stranded template DNA strand is vacant, and then a second oligonucleotide transiently slips into the active site. Such binding may be stabilized by interaction with the amino acids in the enzyme single-stranded template binding site or the stacking energy between the 5' template-strand base and the 3' end interloping-strand base. Such binding would ordinarily not be very stable or long-lived but may be long enough for a nucleotide to be added to the 3' end of the nascent strand. This would further stabilize the interloping oligonucleotide

IDT Exhibit 1020.8946

which could then be copied quickly to its full extent. As a result, the enzyme sits with a blunt-ended duplex in its active site, and another oligonucleotide can then fit into the template-strand binding site and the cycle repeats itself. In this way, long DNA products may be generated. This process of artifactual synthesis of long products by the binding of successive oligonucleotides into the template-strand binding site is referred to herein as templating (see Figure 1).

Definitions

A template-deficient oligonucleotide is an oligonucleotide with at least one region that cannot serve as a template for nucleic acid synthesis. The functional effect is that a template-deficient oligonucleotide cannot be fully replicated. A template-deficient oligonucleotide can be made by, for example, including template-deficient nucleotides in the oligonucleotide. Whether an oligonucleotide is a template-deficient oligonucleotide can depend on the nucleic acid polymerase being used since certain types of nucleotides can be used as a template by some polymerases but not others. For example, many DNA polymerases require a DNA template. Accordingly, as used herein, whether an oligonucleotide is template-deficient is determined based on the nucleic acid polymerase being used in the nucleic acid amplification reaction. Primers that are template-deficient oligonucleotides are referred to herein as template-deficient primers.

Template-deficient nucleotides are nucleotides or nucleotide analogs that (when contained in a nucleic acid molecule) cannot serve as a template for nucleic acid synthesis. Examples of template-deficient nucleotides include abasic nucleotides and derivatized nucleotides. The functional effect is that a template-deficient nucleotide prevents synthesis of a nucleic acid strand complementary to a nucleic acid strand containing a template-deficient nucleotide at or beyond the site of the template-deficient nucleotide. Template-capable nucleotides are nucleotides that are not template-deficient. Whether a nucleotide is a template-deficient nucleotide can depend on the nucleic acid polymerase being used since certain types of nucleotides can be used as a template by some polymerases but not others. For example, many DNA polymerases require a deoxyribonucleotide template. Accordingly, as used herein,

IDT Exhibit 1020.8947

whether a nucleotide is template-deficient is determined based on the nucleic acid polymerase being used in the nucleic acid amplification reaction.

Oligonucleotide-based artifacts are nucleic acids that are formed in a nucleic acid amplification reaction in the absence of nucleic acids other than the oligonucleotides. Oligonucleotide-based artifacts are to be distinguished from artifacts that result from unintended or illegitimate priming of a template nucleic acid (most often a template that is not the intended template). It should be noted that oligonucleotide-based artifacts may occur in the presence of nucleic acids other than the oligonucleotides. The formation of oligonucleotide-based artifacts in the absence of other nucleic acids is used primarily to distinguish oligonucleotide-based artifacts from other types of nucleic acid amplification artifacts. Non-cycle oligonucleotide-based artifacts are oligonucleotide-based artifacts that are formed in nucleic acid amplification reactions not involving thermal cycling. Cycle oligonucleotide-based artifacts are oligonucleotide-based artifacts that are formed in nucleic acid amplification reactions that involve thermal cycling where formation of the artifacts depends on thermal cycling. Primer-based artifacts are oligonucleotide-based artifacts that are formed from or by primers.

Primer-deficient nucleotides are nucleotides or nucleotide analogs that (when contained in a nucleic acid molecule) cannot hybridize (i.e. cannot hydrogen bond) to a complementary nucleotide. Examples of primer-deficient nucleotides include abasic nucleotides and derivatized nucleotides. The functional effect is that a primer-deficient nucleotide prevents formation of a hybrid to a nucleic acid strand containing a primer-deficient nucleotide at the site of the primer-deficient nucleotide. Primer-capable nucleotides are nucleotides that are not primer-deficient. Examples of primer-capable nucleotides are common ribonucleotides and deoxyribonucleotides.

A 3' open oligonucleotide is a template-deficient oligonucleotide where the template-deficient oligonucleotide includes one or more template-deficient nucleotides, and where the number and composition of template-capable nucleotides 3' of the template-deficient nucleotide closest to the 3' end of the template-deficient oligonucleotide is sufficient to allow the template-capable nucleotides 3' of the template-deficient nucleotide closest to the 3' end alone to effectively prime nucleic

IDT Exhibit 1020.8948

acid synthesis in the nucleic acid amplification reaction. A 3' closed oligonucleotide is a template-deficient oligonucleotide where the template-deficient oligonucleotide includes one or more template-deficient nucleotides, where the number and composition of template-capable nucleotides 3' of the template-deficient nucleotide closest to the 3' end of the template-deficient oligonucleotide is insufficient to allow the template-capable nucleotides 3' of the template-deficient nucleotide closest to the 3' end alone to effectively prime nucleic acid synthesis in the nucleic acid amplification reaction. Put another way, placing a template-deficient nucleotide at or near the 3' end of an oligonucleotide results in a 3' closed oligonucleotide, while keeping the region of the oligonucleotide at or near the 3' end free of template-deficient nucleotides results in a 3' open oligonucleotide.

As used herein, a nucleic acid amplification reaction is said to "require" a specified component or condition when the amplification reaction would not function or would not result in the desired amplification. For example, PCR requires thermal cycling since PCR amplification depends on separation of synthesized strands and annealing of primers to both template and synthesized strands for another round of replication. As used herein, a nucleic acid amplification reaction is said to "involve" a specified component or condition when the amplification reaction uses the component or condition. Whether a given amplification reaction involves a particular component depends on the specific reaction involved. For example, a PCR reaction may be performed using labeled primers. Such a reaction can thus be said to "involve" labeled primers. However, since PCR can function without such primers, they would not be considered "required" for PCR. All amplification reactions that "require" a given component or condition also "involve" that component or condition, but not all amplification reactions that "involve" a given component or condition also "require" that component or condition.

Linear amplification refers to nucleic acid amplification that produces, or is designed to produce, an increase in the target nucleic acid directly proportional to the amount of target nucleic acid in the reaction. For example, cycle sequencing produces one strand (or, more precisely, one chain-terminated partial strand) for every target strand present. The amplification is linear since the synthesized strands are not used as

IDT Exhibit 1020.8949

templates in subsequent rounds. Exponential amplification refers to nucleic acid amplification that produces, or is designed to produce, an increase in the target nucleic acid geometrically proportional to the amount of target nucleic acid in the reaction. For example, PCR produces one strand for every original target strand and for every synthesized strand present. The amplification is exponential since the synthesized strands are used as templates in subsequent rounds. An amplification reaction need not actually produce exponentially increasing amounts of nucleic acid to be considered exponential amplification, so long as the amplification reaction is designed to produce such increases.

Cycle sequencing refers to nucleic acid sequencing reactions that require thermal cycling. Cycling, in a nucleic acid amplification reaction, refers to a periodic, repeating change in some condition of the reaction. Minor changes in the condition, such as a slight variation in temperature due to temperature-holding limits of a water bath, is not considered to be cycling, especially where such variations do not have a functional effect on the reaction. Thermal cycling, in a nucleic acid amplification reaction, refers to a periodic, repeating change in the temperature of the reaction of the reaction. PCR requires thermal cycling.

A nucleic acid amplification reaction is a reaction that results in synthesis of more than one copy of all or part of one or more nucleic acid molecules. As used herein, a nucleic acid amplification reaction includes assays, reactions, techniques, and procedures that involve nucleic acid amplification. The fact that such assays, reactions, techniques, and procedures have a goal other than or in addition to nucleic acid amplification--such as nucleic acid sequencing or nucleic acid detection--does not prevent it from being a nucleic acid amplification reaction.

As used herein, nucleoside refers to adenosine, guanosine, cytidine, uridine, 2'-deoxyadenosine, 2'-deoxyguanosine, 2'-deoxycytidine, or thymidine. A nucleoside analog is a chemically modified form of nucleoside containing a chemical modification at any position on the base or sugar portion of the nucleoside. As used herein, the term nucleoside analog encompasses, for example, both nucleoside analogs based on naturally occurring modified nucleosides, such as inosine and pseudouridine, and nucleoside analogs having other modifications, such as modifications to the 2' position

IDT Exhibit 1020.8950

of the sugar. As used herein, nucleotide refers to a phosphate derivative of nucleosides as described above, and a nucleotide analog is a phosphate derivative of nucleoside analogs as described above. The subunits of oligonucleotide analogs, such as peptide nucleic acids, are also considered to be nucleotide analogs.

As used herein, a ribonucleotide is a nucleotide having a 2' hydroxyl function. Analogously, a 2'-deoxyribonucleotide is a nucleotide having only 2' hydrogens. Thus, ribonucleotides and deoxyribonucleotides as used herein refer to naturally occurring nucleotides having nucleoside components adenosine, guanosine, cytidine, and uridine, or 2'-deoxyadenosine, 2'-deoxyguanosine, 2'-deoxycytidine, and thymidine, respectively, without any chemical modification. Ribonucleosides, deoxyribonucleosides, ribonucleoside analogs and deoxyribonucleoside analogs are similarly defined except that they lack the phosphate group, or an analog of the phosphate group, found in nucleotides and nucleotide analogs.

As used herein, oligonucleotide analogs are polymers of nucleic acid-like material with nucleic acid-like properties, such as sequence dependent hybridization, that contain at one or more positions, a modification away from a standard RNA or DNA nucleotide. A preferred example of an oligonucleotide analog is peptide nucleic acid.

The meaning of the above terms is further illustrated by usage of the terms elsewhere herein. The meaning of additional terms used herein but not defined above can generally be understood by common usage in the art and by the context of their usage. The above definitions are not intended to be an exclusive list of terms used and defined herein.

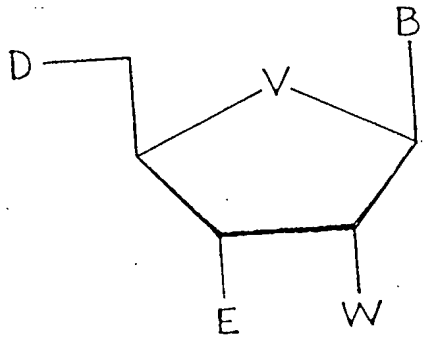
Oligonucleotides

A variety of oligonucleotide structures can be used to create template-deficient oligonucleotides. In general, template-deficient nucleotides useful for making template-deficient oligonucleotides include modified nucleotides, derivatized nucleotides, ribonucleotides, and nucleotide analogs. For example, the oligonucleotide can have one or more abasic nucleotides, one or more nucleotides with an inverted base, fluoro substituted nucleosides, one or more alkyl substituted nucleosides, one or more nucleosides with alkyl groups, one or more nucleoside with phenyl substituted

IDT Exhibit 1020.8951

ethers, one or more nucleotides with substituted thioethers, one or more nucleosides with phosphate esters, one or more α-nucleosides, one or more 2',3'-dideoxy nucleosides, or one or more nucleotides derivatized with compounds such as biotin, amine, Hex, Tet, Fam, fluorescein, rhodamine, alkaline phosphatase, horseradish peroxidase, spacers, cholesteryl, DNP-TEG, psoralen cross-linkers, intercalating agents, PNA conjugates, other enzyme conjugates, and other dye-labels.

One class of modified or derivatized nucleotides have the structure:



Each B can be -H, -OH, -COOH, $-CONH_2$, $-CONHR^1$, $-CONR^1R^2$, $-NH_2$, $-NHR^1$, $-NR^1R^2$, $-NHCOR^1$, -SH, SR^1, -F, $-ONH_2$, $-ONHR^1$, $-ONR^1R^2$, -NHOH, $-NHOR^1$, $-NR^2OH$, $-NR^2OR^1$, substituted or unsubstituted C_1-C_{10} straight chain or branched alkyl, substituted or unsubstituted C_2-C_{10} straight chain or branched alkenyl, substituted or unsubstituted C_2-C_{10} straight chain or branched alkynyl, substituted or unsubstituted C_1-C_{10} straight chain or branched alkoxy, substituted or unsubstituted C_2-C_{10} straight chain or branched alkenyloxy, and substituted or unsubstituted C_2-C_{10} straight chain or branched alkynyloxy. The substituents for W groups are independently halogen, cyano, amino, carboxy, ester, ether, carboxamide, hydroxy, or mercapto. R^1 and R^2 can be substituted or unsubstituted alkyl, alkenyl, or alkynyl groups, where the substituents are independently halogen, cyano, amino, carboxy, ester, ether, carboxamide, hydroxy, or mercapto.

Each V can be an O, S, NH, or CH_2 group.

Each W can be -H, -OH, -COOH, $-CONH_2$, $-CONHR^1$, $-CONR^1R^2$, $-NH_2$, $-NHR^1$, $-NR^1R^2$, $-NHCOR^1$, -SH, SR^1, -F, $-ONH_2$, $-ONHR^1$, $-ONR^1R^2$, -NHOH, $-NHOR^1$, $-NR^2OH$, $-NR^2OR^1$, substituted or unsubstituted C_1-C_{10} straight chain or branched alkyl, substituted or unsubstituted C_2-C_{10} straight chain or branched alkenyl,

IDT Exhibit 1020.8952

substituted or unsubstituted C_2-C_{10} straight chain or branched alkynyl, substituted or unsubstituted C_1-C_{10} straight chain or branched alkoxy, substituted or unsubstituted C_2-C_{10} straight chain or branched alkenyloxy, and substituted or unsubstituted C_2-C_{10} straight chain or branched alkynyloxy. The substituents for W groups are independently halogen, cyano, amino, carboxy, ester, ether, carboxamide, hydroxy, or mercapto. R^1 and R^2 can be substituted or unsubstituted alkyl, alkenyl, or alkynyl groups, where the substituents are independently halogen, cyano, amino, carboxy, ester, ether, carboxamide, hydroxy, or mercapto.

D and E are residues which together form a phosphodiester or phosphorothioate diester bond between adjacent nucleosides or nucleoside analogues or together form an analogue of an internucleosidic bond.

All of the above modifications may make it difficult for a polymerase to add nucleotides opposite such modified nucleotides or extend synthesis beyond such modified nucleotides. A combination of nucleotides having different modifications, derivatizations, or substitutions can be used in a single template-deficient oligonucleotide. Similarly, primers having different types of modifications, derivatizations, or substitutions can be used in the same amplification reaction (see Example 2 and Figure 7).

Template-deficient oligonucleotides can also be made without using modified nucleotides, derivatized nucleotides, or nucleotide analogs. For example, ordinary ribonucleotides can be used as template-deficient nucleotides, depending on the polymerase used. Polymerases that require a DNA template will not be able to use ribonucleotides in the primer.

It is preferred that template-deficient oligonucleotides have at least two consecutive or adjacent template-deficient nucleotides. Such an arrangement reduces the chance that synthesis could pass over a template-deficient region in the oligonucleotide. For example, some polymerases, such as Taq DNA polymerase, add an overhanging 3' nucleotide to a blunt-ended double-stranded DNA structure. This might overhang the 5' abasic base in the templating scenario and allow templating to occur anyway. Use of two consecutive abasic nucleotides should inhibit synthesis across the nick (see Figure 3). The examples shows the reduction in artifacts in an

IDT Exhibit 1020.8953

ERCA reaction when two primer oligonucleotides, P1 and P2, each with two 5' abasic nucleotides are used (see Figure 6).

A subset of the disclosed oligonucleotides are preferred for use in amplification reactions involving thermal cycling, such as PCR. Unlike the primers of Stump *et al.*--which have few enough template-capable nucleotides 3' of the most 3' template-deficient nucleotide to prevent synthesis of a strand to which the primer can effectively prime--the subset of the disclosed primers preferred for thermal cycling reactions will have enough template-capable nucleotides 3' of the most 3' template-deficient nucleotide to allow effective priming of synthesized strands.

Template-deficient nucleotides can be placed anywhere in an oligonucleotide. It may be desirable to keep the 5' end of an oligonucleotide free of template-deficient nucleotides. For example, oligonucleotides with 5' abasic residues or other modifications may not be effectively labeled by phosphorylation using T4 polynucleotide kinase and γ^{32}P-ATP.

Method

The disclosed method involves using one or more of the disclosed template-deficient oligonucleotides in a nucleic acid amplification reaction of interest. Most commonly, the disclosed oligonucleotides will be used as primers in the amplification reaction of interest. Where more than one oligonucleotide is needed or used in the reaction, it is preferred that all of the oligonucleotides in the reaction be template-deficient primers. As used herein, a nucleic acid amplification reaction includes assays, reactions, techniques, and procedures that involve nucleic acid amplification.

The advantage of a nucleic acid amplification reaction that is not susceptible artifacts is that the nucleic acid synthesis products will be specific for the input or target template. In particular, false positives in detection assays involving nucleic acid amplification--caused by artifactual nucleic acid synthesis not dependent on the presence of the target sequence--will be reduced. A preferred nucleic acid amplification reaction for use with the disclosed method if exponential rolling circle amplification (ERCA). The advantage of an ERCA reaction that is not susceptible to primer artifacts is that the DNA synthesis products will be specific for the input

IDT Exhibit 1020.8954

template and the yield of DNA synthesis products will be in direct proportion to the quantity of input template.

The disclosed method differs from the method described in Stump *et al.*, *Nucleic Acids Research* 27:4642-4648 (1999). In Stump *et al.*, the artifact resulted from the thermal cycling in a PCR reaction. This PCR artifact resulted because the product of one round of replication was a template for subsequent rounds. Such artifacts are referred to herein as cycle oligonucleotide artifacts. Stump *et al.* eliminated the cycle oligonucleotide artifact by using primers that could not be fully replicated such that the synthesized strand did not contain enough sequence complementary to the primer to allow the primer to prime synthesis on the synthesized strand (these primers are 3' closed oligonucleotides as defined herein). In contrast, 3' open oligonucleotides as defined herein do allow synthesis of a strand that contains enough sequence complementary to the primer to allow the primer to prime synthesis on the synthesized strand. Further, such primers do not yield a blunt, duplex DNA end such as is required for the templating artifact (that is, a non-cycle oligonucleotide artifact) to occur during the isothermal amplification reactions.

The disclosed method is an improvement of nucleic acid amplification reactions making them more specific. In particular, the disclosed method improves the results when ERCA is used for all applications that ERCA may be used for, such as *in vitro* diagnostics and SNP analysis in a multi-well format. The disclosed method may also yield an improvement of PCR and other amplification methods involving thermal cycling, and can be used with any PCR or other thermal cycling technique. For use in PCR and other amplification reactions involving thermal cycling, it is preferred that the primer be a 3' open oligonucleotide (that is, an oligonucleotide that includes a sufficient number of template-capable nucleotides at the 3' end of the primer to allow synthesis of a complement to a sufficient portion of the primer to allow efficient priming by the primer under the amplification conditions used).

IDT Exhibit 1020.8955

Examples

Example 1: Elimination of Primer Artifacts in Exponential Rolling Circle Amplification Using Primers With Abasic Nucleotides

This example demonstrates that significant primer artifacts are generated in ERCA when no template is present, and that template-deficient primers having two abasic nucleotides at the 5' end eliminate these primer artifacts. As with most nucleic acid amplification techniques used to detect specific nucleic acids, artifactual production of spurious nucleic acids in the absence of the nucleic acid to be detected can result in false positive assays. In assays where the nucleic acid to be detected is present, artifactual production of spurious nucleic acids can affect the accuracy of the assay both qualitatively and, most significantly, quantitatively. This example demonstrates the elimination of such artifacts.

Ten reactions were carried out under the conditions used for ERCA in order to illustrate the reduction of primer-based artifacts by using primers containing two template-deficient nucleotides at the 5' ends. Reactions (30 μl) contained 20 mM Tris-HCl, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 2 mM MgSO4, 0.1% Triton X-100 (pH 8.8 at 25°C). In addition, reactions contained 400 μM deoxyribonucleoside triphosphates, α-[^{32}P] dCTP, specific activity 169 cpm/pmol total dNTP, and 8 units Bst DNA polymerase. ERCA primers were added as indicated, where 'aba' indicates the presence of an abasic nucleotide residue.

P1 (5' end template-capable)

5' CTA AAG CTG AGA CAT GAC GAG TC 3' (SEQ ID NO:1)

P2 (5' end template-capable)

5' GTG ATT CCA CCT TCT CC 3' (SEQ ID NO:2)

P1aba (5' end template-deficient)

5' (aba)(aba) CTA AAG CTG AGA CAT GAC GAG TC 3' (SEQ ID NO:3)

P2aba (5' end template-deficient)

5' (aba)(aba) GTG ATT CCA CCT TCT CC 3' (SEQ ID NO:4)

Reactions contained single-stranded circular ERCA template DNA designated 'G542X gap padlock circle' as indicated, prepared as described by Lizardi *et al.*, *Nature Genetics* 3:225-232 (1998).

IDT Exhibit 1020.8956

Reaction	Additions
1	P1 + P2 + G542X gap padlock circle
2	P1 + P2
3	P1
4	P2
5	P1aba + P2aba + G542X gap padlock circle
6	P1aba + P2aba
7	P1aba
8	P2aba
9	P1 + P2aba
10	P1aba + P2

Reactions were assembled on ice and incubated for 30 minutes on ice. Reactions were initiated by placing them at 65°C and incubated further for 3 hours at 65°C. Aliquots (4 μl) were taken at 0.5, 1, 1.5, 2, and 3 hours and spotted onto DE81 filter paper in order to quantitate DNA synthesis by the incorporation of radioactive deoxyribonucleotide. The results are shown in Figure 6.

As can be seen, significant artifactual DNA production occurs in the absence of template when using template-capable primers (see P1 + P2 - circle curve). Such artifactual DNA is eliminated when template-deficient primers are used (see P1(aba) + P2(aba) - circle curve). The use of template-deficient primers does not prevent legitimate amplification (see P1(aba) + P2(aba) + circle curve). Use of one template-deficient primer and one template-capable primer still results in artifacts, but at a reduced level (see P1 + P2(aba) - circle and P1(aba) + P2 - circle curves).

Example 2: Elimination of Primer Artifacts in Exponential Rolling Circle Amplification Using Primers With Abasic Nucleotides, Ribonucleotides

This example demonstrates that that template-deficient primers having different types of template-deficient nucleotides, used either alone or in combination, can eliminate primer artifacts of ERCA.

Nine reactions were carried out under the conditions described in Example 1. ERCA primers were added as indicated. P1C and P4C are chimeric RNA/DNA oligonucleotides, having 2'-O-methyl ribonucleotides at the 5' end, as indicated by the

IDT Exhibit 1020.8957

use of lower case letters in the sequence. There are 20 2'-O-methyl ribonucleotides at the 5' end of P1C and 18 2'-O-methyl ribonucleotides at the 5' end of P4C. The six nucleotides at the 3' end of P1C and P4C are deoxyribonucleotides, as indicated by the use of capital letters in the sequence.

P1C (5' end template-deficient)

5' aaactaaagctgagacatga CGAGTC 3' (SEQ ID NO:5)

P4C (5' end template-deficient)

5' agtttaatacgactcact ATAGGG 3' (SEQ ID NO:6)

Reaction	Additions
1	P1 + P2 + G542X gap padlock circle
2	P1 + P2
3	P1aba + P2aba + G542X gap padlock circle
4	P1aba + P2aba
5	P1C + P4C
6	P1 + P4C
7	P1C + P2
8	P1aba + P4C
9	P1C + P2aba

Reactions were assembled on ice and incubated for 30 minutes on ice. Reactions were initiated by placing them at 65°C and incubated further for 3 hours at 65°C. Aliquots (4 µl) were taken at 0.5, 1, 2, and 3 hours and spotted onto DE81 filter paper in order to quantitate DNA synthesis by the incorporation of radioactive deoxyribonucleotide. These primers were used both in matched pairs (normal with normal, abasic with abasic, chimeric with chimeric) and in mismatched pairs (abasic with chimeric, abasic with normal, chimeric with normal). The results are shown in Figure 7.

As can be seen, all combinations of template-deficient primers prevent artifact formation (see P1aba + P2aba, P1C + P4C, P1aba + P4C and P1C + P2aba curves). This includes combinations of template-deficient primers having different types of template-deficient nucleotides (see P1aba + P4C and P1C + P2aba curves). Use of one

IDT Exhibit 1020.8958

template-deficient primer and one template-capable primer can still results in artifacts (see P1C + P2 curve), but not always (see P1 + P4C curve).

The results are summarized in the following table.

Primer Pair	Non-specific DNA Synthesis
P1 + P2	yes
P1aba + P2aba	no
P1C + P4C	no
P1 + P4C	no
P1C + P2	yes (reduced)
P1aba + P4C	no
P1C + P2aba	no

IDT Exhibit 1020.8959

CLAIMS

I claim:

1. A method of reducing formation of artifacts in a nucleic acid amplification reaction, the method comprising

using a template-deficient oligonucleotide as at least one of the oligonucleotides in the nucleic acid amplification reaction,

wherein the template-deficient oligonucleotide comprises one or more template-deficient nucleotides,

wherein the number and composition of template-capable nucleotides 3' of the template-deficient nucleotide closest to the 3' end of the template-deficient oligonucleotide is sufficient to allow the template-capable nucleotides 3' of the template-deficient nucleotide closest to the 3' end alone to effectively prime nucleic acid synthesis in the nucleic acid amplification reaction.

2. The method of claim 1 wherein the one or more template-deficient nucleotides are at or near the 5' end of the template-deficient oligonucleotide.

3. The method of claim 1 wherein the template-deficient oligonucleotide comprises two or more template-deficient nucleotides, wherein at least two of the two or more template-deficient nucleotides are adjacent.

4. The method of claim 3 wherein the two or more adjacent template-deficient nucleotides are within three nucleotides of the 5' end of the template-deficient oligonucleotide.

5. The method oligonucleotide of claim 1 wherein the template-deficient nucleotides are selected from the group consisting of modified nucleotides, derivatized nucleotides, ribonucleotides, and nucleotide analogs.

6. The method oligonucleotide of claim 1 wherein the template-deficient oligonucleotide comprises two or more template-deficient nucleotides, wherein at least two of the two or more template-deficient nucleotides are different.

7. The method of claim 1 wherein the template-deficient oligonucleotide comprises two or more template-deficient nucleotides, wherein at least two of the two or more template-deficient nucleotides are template-deficient for different reasons.

IDT Exhibit 1020.8960

8. The method of claim 5 wherein the template-deficient nucleotides are modified nucleotides.

9. The method of claim 5 wherein the modified nucleotides are abasic nucleotides.

10. The method of claim 5 wherein the template-deficient nucleotides are selected from the group consisting of abasic nucleotides, nucleotides with an inverted base, fluoro substituted nucleotides, alkyl substituted nucleotides, nucleotides with phenyl substituted ethers, nucleotides with substituted thioethers, nucleotides with phosphate esters, α-nucleotides, 2',3'-dideoxy nucleotides, ribonucleotides, nucleotides derivatized with biotin, nucleotides derivatized with amine, nucleotides derivatized with Hex, nucleotides derivatized with Tet, nucleotides derivatized with Fam, nucleotides derivatized with fluorescein, nucleotides derivatized with rhodamine, nucleotides derivatized with alkaline phosphatase, nucleotides derivatized with horseradish peroxidase, nucleotides derivatized with spacers, nucleotides derivatized with cholesteryl, nucleotides derivatized with DNP-TEG, nucleotides derivatized with psoralen cross-linkers, nucleotides derivatized with intercalating agents, and nucleotides derivatized with PNA conjugates.

11. The method of claim 1 wherein the nucleic acid amplification reaction does not involve cycle sequencing.

12. The method of claim 11 wherein the nucleic acid amplification reaction does not require linear amplification via thermal cycling.

13. The method of claim 12 wherein the nucleic acid amplification reaction does not involve linear amplification via thermal cycling.

14. The method of claim 1 wherein the nucleic acid amplification reaction involves exponential amplification via thermal cycling.

15. The method of claim 14 wherein the nucleic acid amplification reaction requires exponential amplification via thermal cycling.

16. The method of 14 wherein the nucleic acid amplification reaction involves the polymerase chain reaction.

17. The method of claim 1 wherein the nucleic acid amplification does not involve thermal cycling.

IDT Exhibit 1020.8961

18. The method of 17 wherein the nucleic acid amplification is rolling circle amplification.

19. The method of claim 1 wherein the nucleic acid amplification reaction is selected from the group consisting of exponential rolling circle amplification (ERCA), and rolling circle amplification (RCA), multiple displacement amplification (MDA), strand displacement amplification (SDA), nucleic acid sequence based amplification (NASBA), transcription-mediated amplification (TMA), polymerase chain reaction (PCR), self-sustained sequence replication (3SR), amplification with Qβ replicase, and cycle sequencing.

20. The method of claim 1 wherein the template-deficient oligonucleotide is a primer.

21. The method of claim 20 wherein all of the primers used in the nucleic acid amplification reaction are template-deficient.

22. The method of claim 1 wherein all of the oligonucleotides used in the nucleic acid amplification reaction are template-deficient.

23. A method of reducing formation of artifacts in a nucleic acid amplification reaction, the method comprising

using a template-deficient oligonucleotide as at least one of the oligonucleotides in the nucleic acid amplification reaction,

wherein the nucleic acid amplification reaction does not involve cycle sequencing.

24. The method of claim 23 wherein the nucleic acid amplification reaction does not require linear amplification via thermal cycling.

25. The method of claim 24 wherein the nucleic acid amplification reaction does not involve linear amplification via thermal cycling.

26. The method of claim 23 wherein the nucleic acid amplification does not involve thermal cycling.

27. The method of 26 wherein the nucleic acid amplification is rolling circle amplification.

28. The method of claim 23 wherein the nucleic acid amplification reaction involves exponential amplification via thermal cycling.

IDT Exhibit 1020.8962

29. The method of claim 28 wherein the nucleic acid amplification reaction requires exponential amplification via thermal cycling.

30. The method of 28 wherein the nucleic acid amplification reaction involves the polymerase chain reaction.

31. The method of claim 23 wherein the nucleic acid amplification reaction is selected from the group consisting of exponential rolling circle amplification (ERCA), and rolling circle amplification (RCA), multiple displacement amplification (MDA), strand displacement amplification (SDA), nucleic acid sequence based amplification (NASBA), transcription-mediated amplification (TMA), polymerase chain reaction (PCR), self-sustained sequence replication (3SR), and amplification with Qβ replicase.

32. The method of claim 23 wherein the template-deficient oligonucleotide comprises one or more template-deficient nucleotides.

33. The method of claim 32 wherein the one or more template-deficient nucleotides are at or near the 5' end of the template-deficient oligonucleotide.

34. The method of claim 32 wherein the template-deficient oligonucleotide comprises two or more template-deficient nucleotides, wherein at least two of the two or more template-deficient nucleotides are adjacent.

35. The method of claim 34 wherein the two or more adjacent template-deficient nucleotides are within three nucleotides of the 5' end of the template-deficient oligonucleotide.

36. The method oligonucleotide of claim 32 wherein the template-deficient nucleotides are selected from the group consisting of modified nucleotides, derivatized nucleotides, ribonucleotides, and nucleotide analogs.

37. The method oligonucleotide of claim 32 wherein the template-deficient oligonucleotide comprises two or more template-deficient nucleotides, wherein at least two of the two or more template-deficient nucleotides are different.

38. The method of claim 32 wherein the template-deficient oligonucleotide comprises two or more template-deficient nucleotides, wherein at least two of the two or more template-deficient nucleotides are template-deficient for different reasons.

39. The method of claim 36 wherein the template-deficient nucleotides are modified nucleotides.

IDT Exhibit 1020.8963

40. The method of claim 36 wherein the modified nucleotides are abasic nucleotides.

41. The method of claim 36 wherein the template-deficient nucleotides are selected from the group consisting of abasic nucleotides, nucleotides with an inverted base, fluoro substituted nucleotides, alkyl substituted nucleotides, nucleotides with phenyl substituted ethers, nucleotides with substituted thioethers, nucleotides with phosphate esters, α-nucleotides, 2',3'-dideoxy nucleotides, ribonucleotides, nucleotides derivatized with biotin, nucleotides derivatized with amine, nucleotides derivatized with Hex, nucleotides derivatized with Tet, nucleotides derivatized with Fam, nucleotides derivatized with fluorescein, nucleotides derivatized with rhodamine, nucleotides derivatized with alkaline phosphatase, nucleotides derivatized with horseradish peroxidase, nucleotides derivatized with spacers, nucleotides derivatized with cholesteryl, nucleotides derivatized with DNP-TEG, nucleotides derivatized with psoralen cross-linkers, nucleotides derivatized with intercalating agents, and nucleotides derivatized with PNA conjugates.

42. The method of claim 32 wherein the number and composition of template-capable nucleotides 3' of the template-deficient nucleotide closest to the 3' end is sufficient to allow the template-capable nucleotides 3' of the template-deficient nucleotide closest to the 3' end alone to effectively prime nucleic acid synthesis in the nucleic acid amplification reaction.

43. The method of claim 23 wherein the template-deficient oligonucleotide is a primer.

44. The method of claim 43 wherein all of the primers used in the nucleic acid amplification reaction are template-deficient.

45. The method of claim 23 wherein all of the oligonucleotides used in the nucleic acid amplification reaction are template-deficient.

46. A method of reducing formation of artifacts in a nucleic acid amplification reaction, the method comprising

using a template-deficient oligonucleotide as at least one of the oligonucleotides in the nucleic acid amplification reaction,

IDT Exhibit 1020.8964

wherein the nucleic acid amplification reaction involves exponential amplification via thermal cycling.

47. The method of claim 46 wherein the nucleic acid amplification reaction requires exponential amplification via thermal cycling.

48. The method of 46 wherein the nucleic acid amplification reaction involves the polymerase chain reaction.

49. The method of claim 46 wherein the template-deficient oligonucleotide comprises one or more template-deficient nucleotides.

50. A template-deficient oligonucleotide comprising one or more template-deficient nucleotides,

wherein the number and composition of template-capable nucleotides 3' of the template-deficient nucleotide closest to the 3' end is sufficient to allow the template-capable nucleotides 3' of the template-deficient nucleotide closest to the 3' end alone to effectively prime nucleic acid synthesis in the nucleic acid amplification reaction.

51. The template-deficient oligonucleotide of claim 50 wherein the one or more template-deficient nucleotides are at or near the 5' end of the template-deficient oligonucleotide.

52. The template-deficient oligonucleotide of claim 50 wherein the template-deficient oligonucleotide comprises two or more template-deficient nucleotides, wherein at least two of the two or more template-deficient nucleotides are adjacent.

53. The template-deficient oligonucleotide of claim 52 wherein the two or more adjacent template-deficient nucleotides are within three nucleotides of the 5' end of the template-deficient oligonucleotide.

54. The template-deficient oligonucleotide of claim 50 wherein the template-deficient nucleotides are selected from the group consisting of modified nucleotides, derivatized nucleotides, ribonucleotides, and nucleotide analogs.

55. The template-deficient oligonucleotide of claim 50 wherein the template-deficient oligonucleotide comprises two or more template-deficient nucleotides, wherein at least two of the two or more template-deficient nucleotides are different.

56. The template-deficient oligonucleotide of claim 50 wherein the template-deficient oligonucleotide comprises two or more template-deficient nucleotides,

IDT Exhibit 1020.8965

wherein at least two of the two or more template-deficient nucleotides are template-deficient for different reasons.

57. The template-deficient oligonucleotide of claim 54 wherein the template-deficient nucleotides are modified nucleotides.

58. The template-deficient oligonucleotide of claim 54 wherein the modified nucleotides are abasic nucleotides.

59. The template-deficient oligonucleotide of claim 54 wherein the template-deficient nucleotides are selected from the group consisting of abasic nucleotides, nucleotides with an inverted base, fluoro substituted nucleotides, alkyl substituted nucleotides, nucleotides with phenyl substituted ethers, nucleotides with substituted thioethers, nucleotides with phosphate esters, α-nucleotides, 2',3'-dideoxy nucleotides, ribonucleotides, nucleotides derivatized with biotin, nucleotides derivatized with amine, nucleotides derivatized with Hex, nucleotides derivatized with Tet, nucleotides derivatized with Fam, nucleotides derivatized with fluorescein, nucleotides derivatized with rhodamine, nucleotides derivatized with alkaline phosphatase, nucleotides derivatized with horseradish peroxidase, nucleotides derivatized with spacers, nucleotides derivatized with cholesteryl, nucleotides derivatized with DNP-TEG, nucleotides derivatized with psoralen cross-linkers, nucleotides derivatized with intercalating agents, and nucleotides derivatized with PNA conjugates.

60. A kit for nucleic acid amplification, the kit comprising

a template-deficient primer, wherein the template-deficient primer comprises one or more template-deficient nucleotides,

wherein the number and composition of template-capable nucleotides 3' of the template-deficient nucleotide closest to the 3' end is sufficient to allow the template-capable nucleotides 3' of the template-deficient nucleotide closest to the 3' end alone to effectively prime nucleic acid synthesis in the nucleic acid amplification reaction.

61. The kit of claim 60 wherein the one or more template-deficient nucleotides are at or near the 5' end of the template-deficient primer.

62. The kit of claim 60 wherein the template-deficient primer comprises two or more template-deficient nucleotides, wherein at least two of the two or more template-deficient nucleotides are adjacent.

IDT Exhibit 1020.8966

63. The kit of claim 62 wherein the two or more adjacent template-deficient nucleotides are within three nucleotides of the 5' end of the template-deficient primer.

64. The kit of claim 60 wherein the template-deficient nucleotides are selected from the group consisting of modified nucleotides, derivatized nucleotides, ribonucleotides, and nucleotide analogs.

65. The kit of claim 60 wherein the template-deficient primer comprises two or more template-deficient nucleotides, wherein at least two of the two or more template-deficient nucleotides are different.

66. The kit of claim 60 wherein the template-deficient primer comprises two or more template-deficient nucleotides, wherein at least two of the two or more template-deficient nucleotides are template-deficient for different reasons.

67. The kit of claim 64 wherein the template-deficient nucleotides are modified nucleotides.

68. The kit of claim 64 wherein the modified nucleotides are abasic nucleotides.

69. The kit of claim 64 wherein the template-deficient nucleotides are selected from the group consisting of abasic nucleotides, nucleotides with an inverted base, fluoro substituted nucleotides, alkyl substituted nucleotides, nucleotides with phenyl substituted ethers, nucleotides with substituted thioethers, nucleotides with phosphate esters, α-nucleotides, 2',3'-dideoxy nucleotides, ribonucleotides, nucleotides derivatized with biotin, nucleotides derivatized with amine, nucleotides derivatized with Hex, nucleotides derivatized with Tet, nucleotides derivatized with Fam, nucleotides derivatized with fluorescein, nucleotides derivatized with rhodamine, nucleotides derivatized with alkaline phosphatase, nucleotides derivatized with horseradish peroxidase, nucleotides derivatized with spacers, nucleotides derivatized with cholesteryl, nucleotides derivatized with DNP-TEG, nucleotides derivatized with psoralen cross-linkers, nucleotides derivatized with intercalating agents, and nucleotides derivatized with PNA conjugates.

70. The kit of claim 60 wherein the nucleic acid amplification reaction does not involve cycle sequencing.

IDT Exhibit 1020.8967

71. The kit of claim 70 wherein the nucleic acid amplification reaction does not require linear amplification via thermal cycling.

72. The kit of claim 74 wherein the nucleic acid amplification reaction does not involve linear amplification via thermal cycling.

73. The kit of claim 63 wherein the nucleic acid amplification reaction involves exponential amplification via thermal cycling.

74. The kit of claim 76 wherein the nucleic acid amplification reaction requires exponential amplification via thermal cycling.

75. The kit of claim 76 wherein the nucleic acid amplification reaction involves the polymerase chain reaction.

76. The kit of claim 63 wherein the nucleic acid amplification reaction is selected from the group consisting of exponential rolling circle amplification (ERCA), and rolling circle amplification (RCA), multiple displacement amplification (MDA), strand displacement amplification (SDA), nucleic acid sequence based amplification (NASBA), transcription-mediated amplification (TMA), polymerase chain reaction (PCR), self-sustained sequence replication (3SR), and amplification with Qβ replicase.

IDT Exhibit 1020.8968

FIG. 1

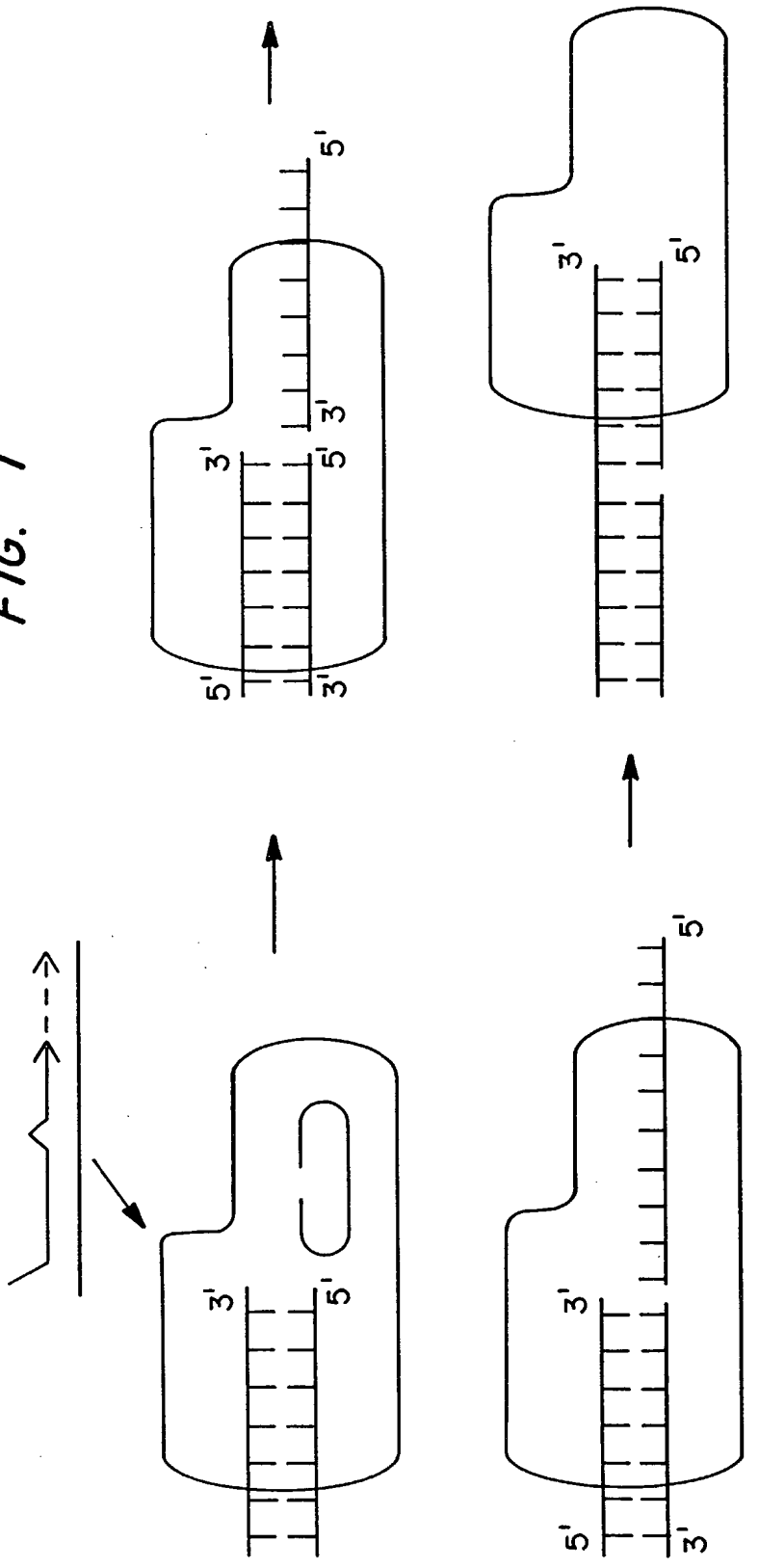
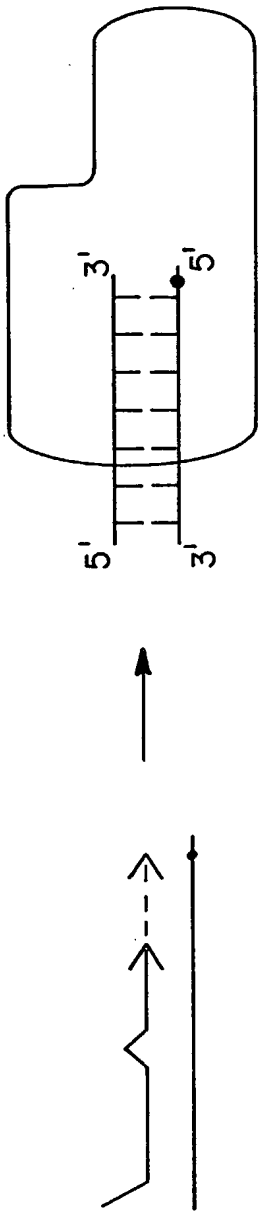


FIG. 2



IDT Exhibit 1020.8969

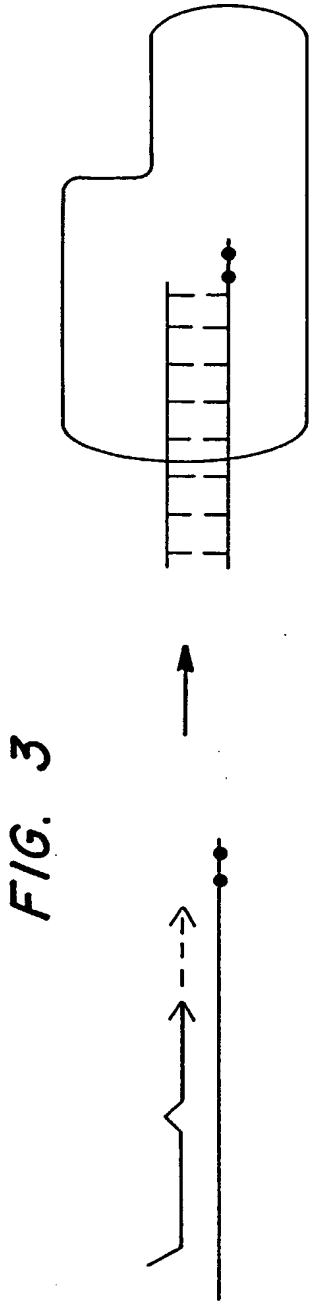


FIG. 3

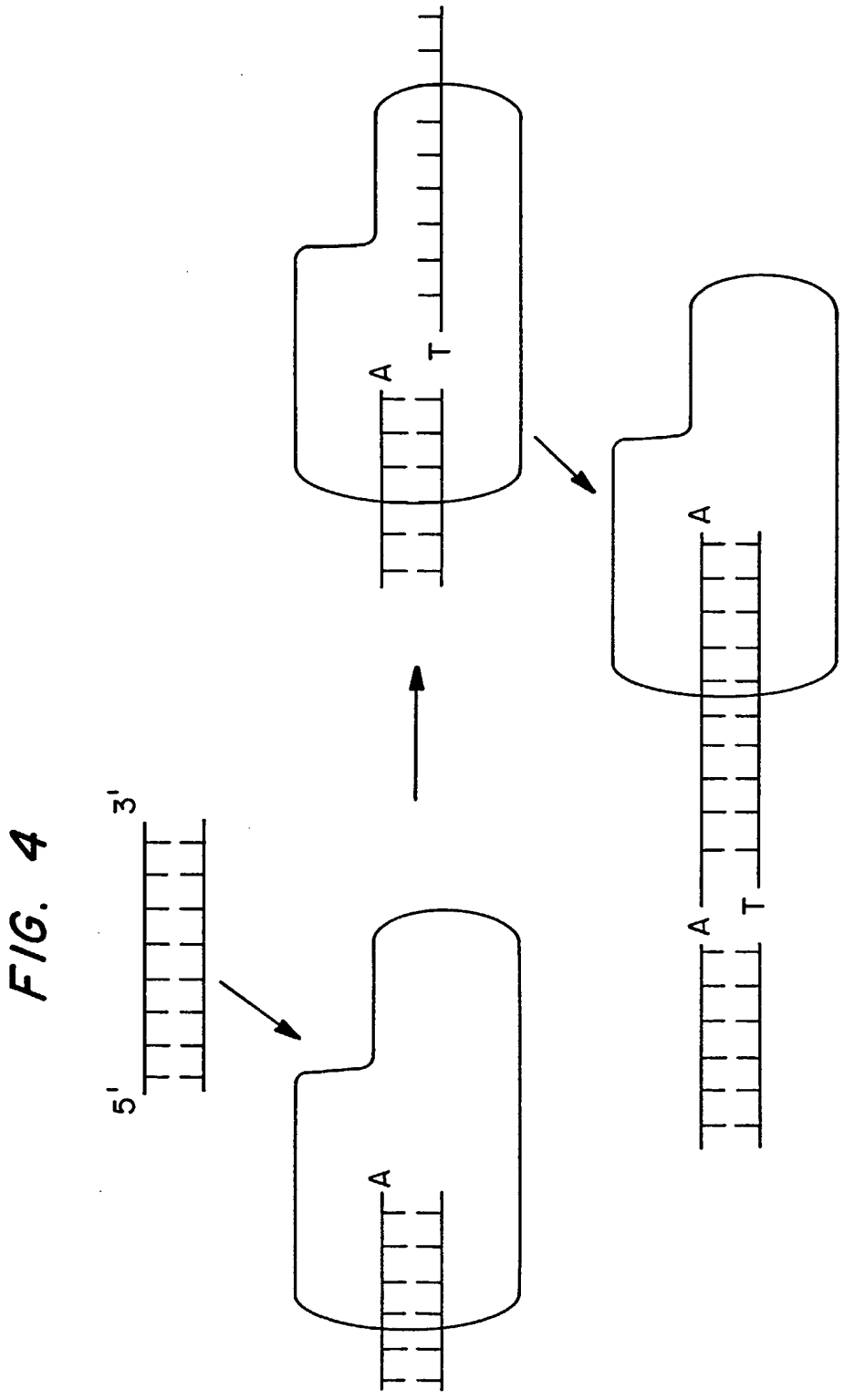
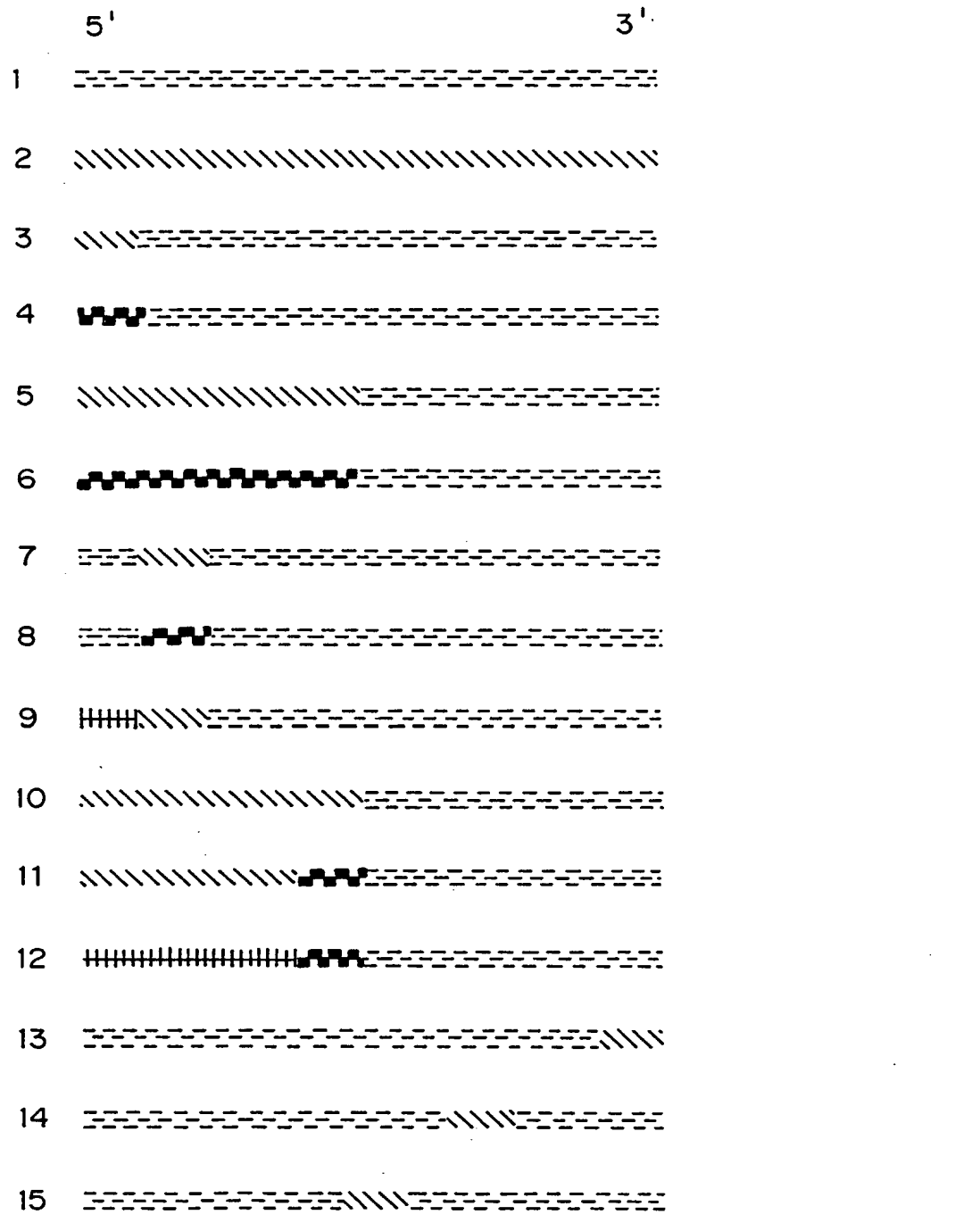


FIG. 4

IDT Exhibit 1020.8970

FIG. 5



IDT Exhibit 1020.8971

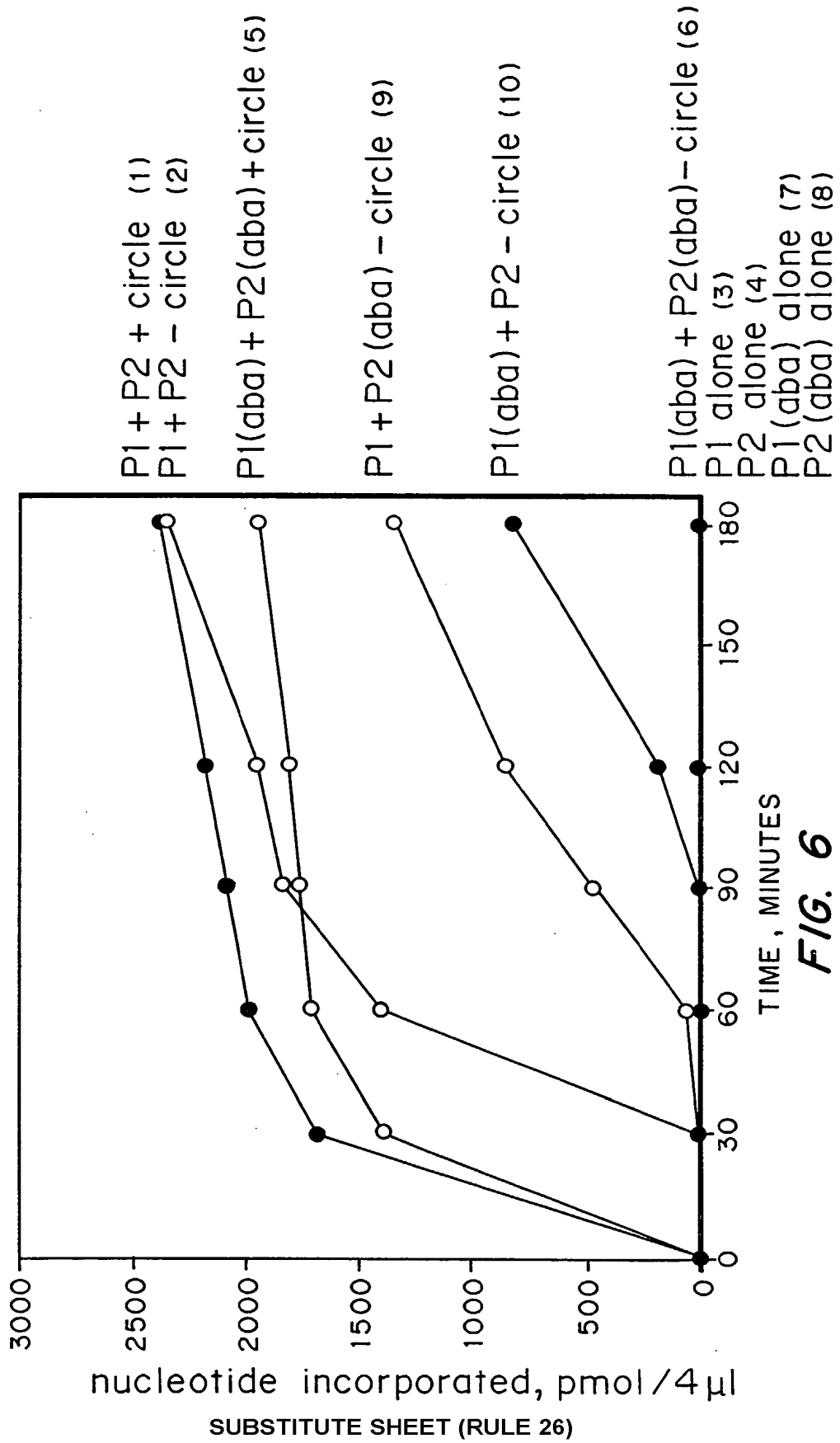


FIG. 6

IDT Exhibit 1020.8972

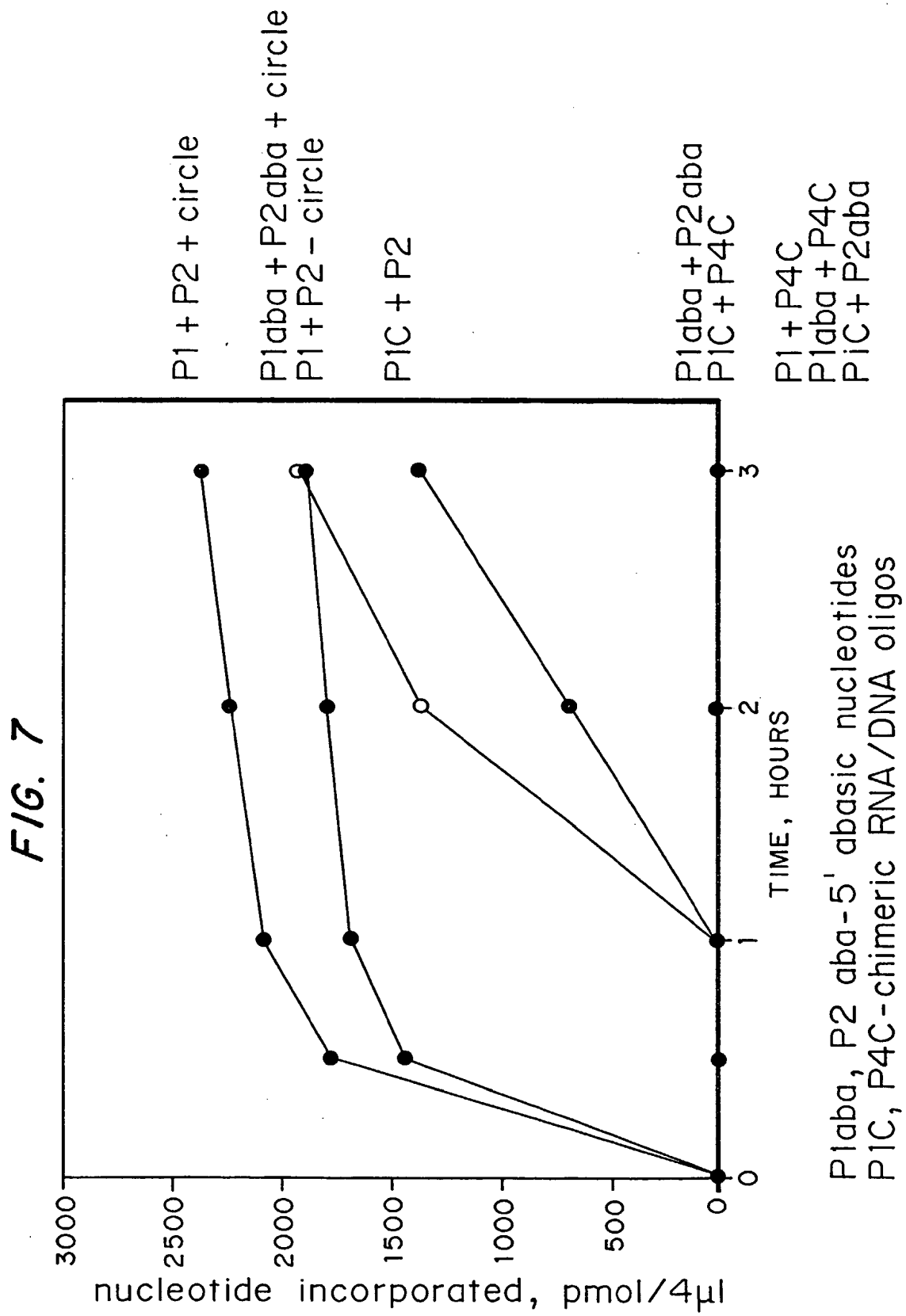


FIG. 7

IDT Exhibit 1020.8973

SEQUENCE LISTING

<110> Molecular Staging, Inc.
Dean, Frank B.
Faruqi, A. Fawad

<120> METHOD FOR REDUCING ARTIFACTS IN NUCLEIC ACID AMPLIFICATION

<130> 13172.0001P1

<150> 09/514,113
<151> 2000-02-28

<160> 6

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence:\Note = synthetic construct

<400> 1
ctaaagctga gacatgacga gtc 23

<210> 2
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence:\Note = synthetic construct

<400> 2
gtgattccac cttctcc 17

<210> 3
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence:\Note = synthetic construct

<221> misc_feature
<222> 1,2
<223> Nucleotides at positions 1 and 2 are abasic

nucleotides

<400> 3
nnctaaagct gagacatgac gagtc 25

<210> 4
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence:\Note =
synthetic construct

<221> misc_feature
<222> 1,2
<223> Nucleotides at positions 1 and 2 are abasic
nucleotides

<400> 4
nngtgattcc accttctcc 19

<210> 5
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence:\Note =
synthetic construct

<400> 5
aaactaaagc tgagacatga cgagtc 26

<210> 6
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence:\Note =
synthetic construct

<400> 6
agtttaatac gactcactat aggg 24

Espacenet

Bibliographic data: WO0008208 (A2) — 2000-02-17

REVERSE TRANSCRIPTION AND AMPLIFICATION PROCESSES AND PRIMERS THEREFOR

Inventor(s): FREEMAN THOMAS CHARLES [GB]; RICHARDSON PETER JOHN [GB]; DIXON ALISTAIR K [GB] ± (FREEMAN, THOMAS, CHARLES, ; RICHARDSON, PETER, JOHN, ; DIXON, ALISTAIR, K)

Applicant(s): MEDICAL RES COUNCIL [GB]; FREEMAN THOMAS CHARLES [GB]; RICHARDSON PETER JOHN [GB]; DIXON ALISTAIR K [GB] ± (MEDICAL RESEARCH COUNCIL, ; FREEMAN, THOMAS, CHARLES, ; RICHARDSON, PETER, JOHN, ; DIXON, ALISTAIR, K)

Classification: **- international:** ***C12Q1/68;*** (IPC1-7): C12Q1/68
- cooperative: **C12Q1/6853;** C12Q2600/158 more

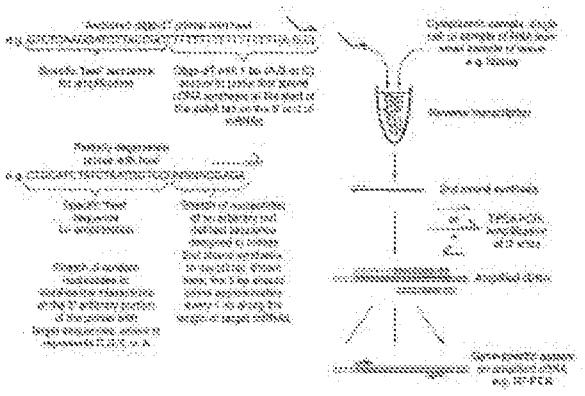
Application number: WO1999GB02579 19990805 Global Dossier

Priority number(s): GB19980017055 19980805

Also published as: WO0008208 (A3) AU5294399 (A)

Abstract of WO0008208 (A2)

A method is provided for the expression profiling of single cells. The method employs a first heeled primer for reverse transcription of mRNA in a sample to provide first strand cDNA species, and then a second heeled primer population to generate second strand cDNAs. The non-heeled portion of the second heeled primers are capable of hybridizing to the reverse transcribed first strands of cDNA species, at least one along the lengths thereof. Due to the presence of random and preselected sequences in the second primers a qualitatively more uniform and therefore representative cDNA profile is produced from cellular mRNAs.



IDT Exhibit 1020.8976

IDT Exhibit 1020.8977

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12Q 1/68

A2

(11) International Publication Number: **WO 00/08208**

(43) International Publication Date: 17 February 2000 (17.02.00)

(21) International Application Number: PCT/GB99/02579

(22) International Filing Date: 5 August 1999 (05.08.99)

(30) Priority Data:
9817055.8 5 August 1998 (05.08.98) GB

(71) Applicant *(for all designated States except US)*: MEDICAL RESEARCH COUNCIL [GB/GB]; 20 Park Crescent, London W1 4AL (GB).

(72) Inventors; and

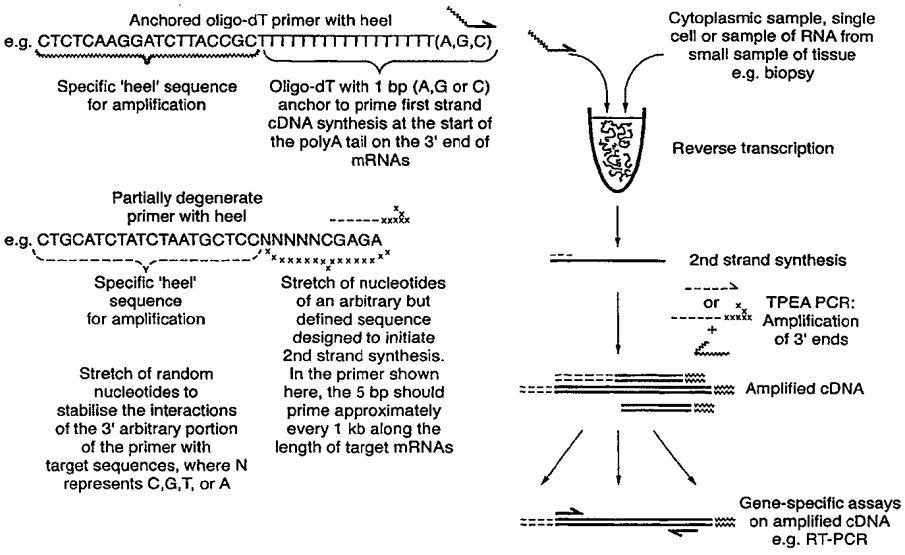
(75) Inventors/Applicants *(for US only)*: FREEMAN, Thomas, Charles [GB/GB]; 80 Whitehill Road, Cambridge CB5 8LX (GB). RICHARDSON, Peter, John [GB/GB]; Hillcrest Cottage, Bartlow Road, Hadstock CB1 6PF (GB). DIXON, Alistair, K. [GB/GB]; 108 Gwydir Street, Cambridge CB1 2LL (GB).

(74) Agent: HEPWORTH LAWRENCE BRYER & BIZLEY; Merlin House, Falconry Court, Bakers Lane, Epping, Essex CM16 5DQ (GB).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: REVERSE TRANSCRIPTION AND AMPLIFICATION PROCESSES AND PRIMERS THEREFOR



(57) Abstract

A method is provided for the expression profiling of single cells. The method employs a first heeled primer for reverse transcription of mRNA in a sample to provide first strand cDNA species, and then a second heeled primer population to generate second strand cDNAs. The non–heeled portion of the second heeled primers are capable of hybridizing to the reverse transcribed first strands of cDNA species, at least one along the lengths thereof. Due to the presence of random and preselected sequences in the second primers a qualitatively more uniform and therefore representative cDNA profile is produced from cellular mRNAs.

IDT Exhibit 1020.8978

FOR THE PURPOSES OF INFORMATION ONLY

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

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

IDT Exhibit 1020.8979

REVERSE TRANSCRIPTION AND AMPLIFICATION PROCESSES AND PRIMERS THEREFOR

The present invention relates to processes for the reverse transcription of mRNA to provide cDNA. Additionally, the invention relates to processes involving reverse transcription of mRNA and concomitant or subsequent amplification of cDNA. The mRNA samples are typically those obtainable from cell or tissue samples of organisms, such as biopsy or blood samples, for example. In particular, the invention relates to processes for reverse transcription or reverse transcription and amplification of a population of mRNA obtainable from a single cell. Thus, the invention concerns the field of analysis of gene expression (i.e. "expression profiling") of cells and tissues, even down to level of a single cell. The invention also relates to polynucleotide primers adapted for the performance of reverse transcription on population of mRNA species with concomitant or subsequent amplification procedures.

DNA sequence information resulting from genome and expressed sequence tag (EST) sequencing projects is expected to provide the basis for furthering understanding of the control and mode of action of individual gene products. In this respect, expression profiling is regarded as playing a role in the functional characterisation of newly identified genes (Lander, E S (1996), Science, 274:536-539 and Strachan, T (1997), Nature Genetics, 16:126-132).

IDT Exhibit 1020.8980

Many tissues, such as the immune and nervous systems, are composed of highly heterogeneous cell populations. A key factor in understanding their physiology, and the role of specific gene products expressed within them, is to examine gene usage in the context of this cellular diversity. In the past, methods such as Northern blotting and nuclease protection assays were employed to study gene expression. More recently, techniques have been developed for assessing simultaneously the expression of large numbers of genes, e.g. (deRisi, J *et al* (1996), Nature Genetics, 14:457-460; Chee, M *et al* (1996), Science, 274:610-614; Lockhart, D J *et al* (1996), Nature Biotech, 14:1675-1680; Madden S L *et al* (1997), Oncogene, 15:1079-1085; and Marshall A & Hodgson J (1998), Nature Biotechnology, 16:27-31). All these techniques, however, require relatively large amounts of RNA and currently lack the sensitivity to analyse specimens derived from small populations of cells or indeed from an individual cell.

At present, methods for the analysis of gene expression within single cells or small tissue samples are limiting. Whilst *in situ* hybridization techniques provide detailed information about the cellular expression pattern of a gene in intact tissue, be it whole-mounts or tissue sections, the technique is relatively laborious and unable to analyse multiple transcripts in a single preparation.

In recent years, the polymerase chain reaction (PCR) has been used successfully to investigate gene

expression in cytoplasmic samples derived from single cells. The nested-primer approach has a good level of sensitivity, but the analysis is restricted to just a small number of closely related genes from specific gene families (Lambolez, B *et al* (1992) Neuron, 9:247-258; Sucher, N J *et al* (1993), J Biol Chem, 268:22299-22304; and Yan, Z & Surmeier, D.J (1997), Neuron, 19:1115-1126).

Other techniques known to be capable of detecting the expression of unrelated genes in a single cell include T7 RNA polymerase amplification of mRNA (Eberwine, J *et al* (1992) Proc Natl Acad Sci USA, 89:3010-3014; Van Gelder *et al* (1990), Proc Natl Acad Sci USA, 87:1663-1667). Also known are methods employing PCR after prior homopolymeric tailing of the first strand cDNA (Brady G & Iscove, N.N. (1993), Methods in Enzymol, 225:611-623; Frohman *et al* (1988), Proc Natl Acad Sci USA, 85:8998-9002; and Jena P K *et al*, (1996), J Immunol Methods, 190:199-213). However, neither of the aforementioned approaches have been demonstrated to be capable of analysing more than a small number of genes of the total number of genes expressed in any given cell. Moreover, the techniques are not widely used. The former procedures suffer from technical difficulties, whilst the latter procedures are biased against long transcripts and often requires subsequent cloning of the amplified products. In view of the known technologies, there is still a great need for improved methods which can quickly, easily, sensitively and simultaneously reveal for analysis the spectrum of gene expression at the cellular level, even right down to the level of a single cell.

Significantly, there is also a need for understanding gene expression at the level of single cells and small biopsy samples. A method capable of assessing gene expression in samples as small as a single cell would be of considerable benefit to the understanding of gene expression and so the molecular basis of cell function, the identification of specific types of diseased (e.g. cancerous) cells and in the future routine analysis of gene expression in the scientific and wider communities. (With the increasing influence of genomics on Biomedical Science, the ability of scientists and medical doctors to assess gene expression in small tissue samples will be of increasing utility.)

Currently, there is also considerable interest in the use of multiple gene arrays to assess the expression in tissue samples of many hundreds or thousands of genes. However, although this approach allows many genes to be examined simultaneously in a single sample, relatively large amounts of starting material are required. Thus, any data obtained cannot be unequivocally assigned to individual cell types in an original sample and because of this the samples have to be large (requiring more than 10^7 cells). This causes considerable problems in the collection of samples from humans and in the interpretation of the data. This problem is particularly acute in the CNS and immune systems where the heterogeneity of the cells means that such data is extremely difficult to interpret.

Previous attempts at assessing gene expression at the level of single cells has been of only limited success. The following are presented as examples of known techniques.

Nested PCR

Nested PCR has been used successfully in many scientific laboratories and it relies on two sequential amplification steps, both targeted to the genes of interest. This means, therefore that in any one cell sample the expression of only a few genes (of up to 3 gene families) can be examined. The technique does not ensure that all members of cDNA populations in a complex mixture are amplified, nor that all the amplicons are of similar sizes. The technique suffers from the drawback that only a few gene families can be examined at a time. This is wholly unsatisfactory from the point of view of expression profiling.

RNA Amplification

This is a technically complex procedure and its manifest difficulties are seen in the few laboratories in the world that are able to use this technique regularly. The technique is just not suited to a simultaneous analysis of a large number of cell samples.

In Situ Hybridization

This technique is suited to the analysis of the expression of no more than about 3 genes in any cell.

The technique is useful for simultaneous expression screening of a large number of dead cells in fixed tissue slices *In situ* hybridization usually involves the use of radiolabels which are inconvenient and the technique as a whole is quite unsuitable for expression profiling of living cells.

cDNA Tailing

This technique used by a number of research groups employs terminal deoxynucleotide transferase to attach a unique priming site on to the 3' end of a first strand cDNA following reverse transcription. Following PCR of the cDNA with the unique priming site, the expression profile of a number genes from a single cell can be analysed. The technique has a number of drawbacks. First, there is the need to use homopolymeric PCR primers capable of annealing to sites in a DNA sequence. Also, there is an unequal amplification of cDNA because unequal lengths of the cDNA transcripts are amplified. The amplification efficiency is low. The initial PCR reactions for the different transcripts operate at vastly different efficiencies, and so bias the procedure in favour of shorter gene transcripts.

Ligation

In this technique, a primer sequence is ligated to the 3' end of cDNA to provide a second amplification primer site. However, this technique suffers from the same problems as the cDNA tailing technique referred to above. PCR of the different cDNA species in the reverse transcribed sample takes place at

different levels of efficiency, depending on the length of the cDNA molecule being amplified. Additionally, the ligation reactions can be difficult to control with multiple priming sites being ligated.

Each of the aforementioned techniques suffers from a variety of limitations. Indeed, an analysis of the scientific literature suggests that none of these techniques (other than *in situ* hybridization) are widely used in practice. Existing methods for analysis of gene expression in small samples of RNA, particularly from single cells, are severely limited in terms of the number and diversity of genes that it is possible to analyse, and the difficulty of experimental procedures involved.

The present inventors have sought to develop a more straightforward, reproducible and reliable cDNA amplification procedure for small mRNA samples wherein expression profiling can be conducted, thereby avoiding the various problems outlined above.

Accordingly, in first aspect the present invention provides a process of reverse transcribing mRNA species present in a sample from an organism comprising:

- reverse transcribing the mRNA species using a first heeled primer, thereby to provide first strand cDNA species;

- synthesising second cDNA strands using a second heeled primer population, the non-heel portion of the second primers being capable of hybridizing to

the reverse transcribed first strand cDNA species at least once along the lengths thereof.

In second aspect the invention provides a process of reverse transcribing and amplifying mRNA species present in a sample from an organism comprising:

- reverse transcribing the mRNA species using a first heeled primer, thereby to provide first strand cDNA species;

- synthesising second cDNA strands using a second heeled primer population, the non-heel portion of the second heeled primers being capable of hybridizing to the reverse transcribed first strand cDNA species at least once along the lengths thereof;

- amplifying the resulting cDNA species.

In other words, in both of the first and second aspects of the invention described above, the particular sequence of the non-heel portion of each second primer in the second heeled primer population is such that at least one second heeled primer from the population is capable of hybridizing to each reverse transcribed first strand cDNA species. For each first strand cDNA species, the number and identity of individual second heeled primers which hybridize thereto is expected to be different but this is not an expectation which excludes the possibility of similarities in hybridizations with cDNAs arising. Furthermore, due to the sequence characters of the non-heel sequences of the second

primers, their hybridization with the first cDNA strands is possible at a multiplicity of sufficiently complementary sites along the lengths of the cDNA strands.

The particular temperatures, enzymes and reagents (other than the first primer) used in the process of reverse transcription may be those already known in the art.

A "heeled" primer will be readily understood in the art to be a primer comprising a hybridizing portion and a non-hybridizing portion, wherein the non-hybridizing portion represents the "heel" of the primer.

The second primer is actually a population of individual primer species. When the first strand cDNA population is contacted with the second primer population under appropriate hybridizing conditions then because of the selection of nucleotide sequences amongst the second primers, each cDNA species will hybridize with at least one second primer, second cDNA strand synthesis then proceeds in a 5' to 3' direction from the hybridized second primer. Although the inventors do not wish to be bound by any particular theory, what appears to happen is that when more than one second primer species hybridizes to any given cDNA species then second strand cDNA synthesis proceeds in its 5' to 3' direction from the second primer whose 3' end is not obstructed along the first strand cDNA template by any other second primer hybridized thereto. As a result, and particularly where a multiplicity of second primers

hybridize to the first strand cDNA template, there is more of a tendency to generate second strand cDNA molecules starting at points further upstream on the first strand cDNA template, i.e. further downstream on the original mRNA molecules. Coincidentally, the 3' end portions of mRNA molecules are generally more diverse in their sequences.

A net result of the reverse transcription process of the invention is that there appears to be a bias towards a more uniform length of cDNA molecule. This in turn impacts on any subsequent amplification procedure. When the amplification is PCR for example then more uniform length cDNA molecules lend themselves more to amplification than a population of cDNA molecules less uniform in length.

Advantageously, the processes of the invention generate cDNA molecules highly representative of the spectrum of mRNA molecules in a sample. The element of bias towards more uniform length cDNA molecules ensures that even relatively low abundance mRNA species are transcribed, and optionally amplified, to the same level of efficiency as more abundant mRNA species. Thus, a much better qualitative profile of expressed gene sequences in samples can be achieved than was hitherto possible.

The degree of sensitivity of the processes of this invention can of course be varied by modifying the numbers and sequences of the second primer population species thereby modifying the frequency with which the second primers hybridize to the first strand cDNA templates.

Where further amplification of the cDNA products is required, the amplification of the cDNA species resulting from the reverse transcription preferably employs a third primer comprising at least a part of the heel portion of the first heeled primer and a fourth primer comprising at least part of the heel portion of the second heeled primer. Further amplification may be advantageous where subsequent analysis of cDNA species involves less sensitive detection means or where a larger sample is required for analysis by methods which require larger quantities of cDNA material.

The second heeled primer population may comprise primers differing by up to five nucleotide bases the population preferably comprising a number of primers in the range 1000 to 100,000 primers, more preferably in the range 1024 to 65536 primers. In order to achieve this, the primers of the second heeled primer population preferably each comprise a random sequence of nucleotides in the range of 5 to 8 nucleotides 3' to the heel and a further sequence of at least 5 nucleotides contiguous 3' therewith. As will be appreciated, where there are 5 random nucleotides (which is preferred) there will be 4^5 (i.e. 1024) possible pentamer sequences.

The further sequence of nucleotides may be selected by sequence analysis of known sequences so as to promote the ability of the second heeled primer as a whole to hybridize to the transcribed cDNA species. The sequence analysis can be carried out through databases of DNA or RNA sequences. In particular,

the known sequences of the organism of interest are preferably consulted. The further sequence of nucleotides preferably comprises a number of nucleotides in the range 2 to 10 nucleotides. In a particularly preferred embodiment the further sequence of nucleotides may comprise a number of nucleotides equivalent to the number of nucleotides in the random sequence of nucleotides.

The further nucleotide sequence of the second heeled primers is preferably constant throughout the population of these primers and it is selected so as to stabilise the primers and to ensure optimal efficiency of hybridization to target first strand cDNA species.

In preferred embodiments, the second heeled primer from the population of second primers preferably hybridises on average once in every 1kb portion of first strand cDNA species. This has been found to provide a relatively efficient and uniform reverse transcription and optionally amplification of mRNAs in a sample.

A particularly preferred further sequence of nucleotides in the second primers is:

CGAGA

and a particularly preferred second heeled primer population is represented by:

CTGCATCTATCTAATGCTCCNNNNNCGAGA

wherein N is independently selected from C G T or A.

The heel portion of the first and second heeled primers are preferably selected so that they lack the ability to hybridise to mRNA or first strand cDNA respectively. The heel portions, like the further sequence portions of the second primers, are selected by an analysis of known nucleotide sequence information. In particularly preferred embodiments, the heel portions preferably comprise sequences absent from the mRNA species in the sample, although the heel portions may simply comprise sequences absent from the genome of the organism from which the sample is taken. The heel portions preferably comprise a number of nucleotides in the range 15 to 50, more preferably 18 to 22 nucleotides although somewhat fewer or somewhat more nucleotides may be acceptable.

In preferred embodiments, the first heeled primer is preferably an anchored primer comprising an oligo (dT) sequence. The nature of anchored primers is already well known in this art. For example, provision of a few non-T bases at the 5' end of the primer ensures that hybridization of the primer occurs at the 5' end of the mRNA poly A tail.

The fourth primer is preferably the heel of the second heeled primer, or at least a portion thereof.

The third primer is preferably the heel of the first heeled primer, or at least a portion thereof.

The third primer may be the same as the first heeled primer and this can be advantageous in reducing the numbers of reagents needed to perform the processes of the invention.

The frequency with which individual second primer population species hybridize along a given length of nucleic acid may be adjusted by employing suitable hybridizing conditions. Preferably, the hybridization conditions are of limited stringency so that the random sequences of oligonucleotides in the second primers have a significant effect on whether hybridization occurs or not. The degree of stringency of hybridizing conditions and the number of contiguous random bases in the second primers may be varied according to routine trial and error in order to achieve a desired frequency of hybridization of second primer species along a given length of nucleic acid material.

The amplification of the resulting cDNA species preferably comprises more than one round of amplification cycles. Preferably, each further round of amplification comprises addition of further second and third primers and amplification reagents, optionally a fourth primer as well.

Each round of amplification may comprise 5 to 45 cycles, more preferably 10 to 40 cycles. In particular, the first round of amplification may comprise a lesser number of cycles than any further rounds of amplification.

The amplification process is preferably PCR although modified PCR procedures or other compatible amplification procedures may be used.

Preferred PCR cycles comprise X_1°C for Y_1 min; then X_2°C for Y_2 min; then X_3°C extension for Y_3 minute; then Y_4 min extension, wherein X_1, $>X_3$ $>X_2$ and Y_4 $>Y_3$ $>Y_2$ $>Y_1$. X_1 is in the range 90 to 94°C, X_2 is in the range 45 to 70°C, X_3 is in the range 65 to 75°C and Y_1, Y_2, Y_3 are in the range 15 seconds to 4 minutes and Y_4 is in the range 2.5 to 20 minutes. Particularly preferred values are X_1 = 92°C, X_2 = 60°C, X_3 = 72°C, Y_1 = 2.5 minutes, Y_2 = 1.5 minutes, Y_3 = 1 minute and Y_4 = 10 minutes.

The sample from an organism will preferably include or be derived from tissue or cells. The sample may be comprised of whole cells, possibly comprising a single cell type, even comprising just a single cell. In particularly preferred embodiments the samples are composed of the cytoplasm of cells (same cell type or mixture) or more preferably the cytoplasm of a single cell. Biopsy samples may provide useful sources of sample material ranging from a few grams to a few micrograms of tissue/cell material.

The cytoplasm may be obtained by lysis or aspiration of a cell or cells and such cell may be obtained by fluorescence activated cell sorting (FACS). The processes of the invention are sufficiently reliable, sensitive and efficient that substantially all mRNA species in a sample are reverse transcribed and optionally amplified to approximately the same degree. A more accurate and reliable picture of gene

expression can be obtained for a cell sample. Advantageously, the processes of the invention permit single cell gene profiling.

In third aspect the invention provides a method of reverse transcribing expressed gene sequences in a sample from an organism comprising reverse transcribing the mRNA in the sample using a first primer to produce first strand cDNA species, synthesising second cDNA strands using a population of second primers, wherein at least one second primer in the population hybridises to a given first strand cDNA species. This method may further comprise the amplification of the resulting double stranded cDNA. Preferred or alternative versions of this method may comprise one or more of the further features of the other processes of the invention as hereinbefore described.

In fourth aspect the invention provides a polynucleotide primer for reverse transcription of mRNA species comprising an oligo (dT) sequence and 5' thereto a polynucleotide heel sequence, wherein the heel sequence is substantially incapable of hybridisation to the mRNA species. The primer is preferably an anchored primer and may comprise the further features as hereinbefore described.

In fifth aspect the invention provides a polynucleotide primer population for synthesis of second strand cDNA species from first strand cDNA species, wherein at least one primer in the population is capable of hybridising to a given first strand of cDNA. At least one primer in this primer

population is preferably capable of hybridising approximately at least once in any given 1kb of first strand cDNA. The primers may further comprise one or more additional features of such primers as hereinbefore described.

In sixth aspect the invention provides polynucleotide primers for amplification of cDNA comprising a reverse transcription primer (ie the first primer) as hereinbefore described and a primer comprising at least a portion of the heel portion of the second primer population as hereinbefore described.

In seventh aspect, the invention provides the use of a polynucleotide comprising an oligo (dT) sequence and a heel sequence 5' thereto for the reverse transcription of mRNA species in a sample. Such a polynucleotide may further comprise one or more of the characteristics of the first primer as hereinbefore described.

In eighth aspect the invention provides the use of a polynucleotide primer population as hereinbefore defined for the synthesis of second strand cDNA from a population of first strand cDNA species.

In ninth aspect the invention provides a cDNA library preparation obtainable by a process or method as hereinbefore defined, said library comprising substantially all cDNA species corresponding to genes expressed by a single cell, cell type or tissue.

In tenth aspect the invention provides a kit for the production of cDNA from mRNA in a sample from an

organism comprising a primer of the fourth aspect of the invention and a primer population of the fifth aspect of the invention. The kit may further comprise at least one further primer for achieving amplification of the cDNA. Particularly preferred kits comprise a primer pair for amplification of the cDNA.

The invention thus provides a rapid, robust and reproducible procedure, called Three Prime End Amplification (TPEA), optionally with PCR (TPEA-PCR), capable of amplifying 3' fragments of cDNA prior to analysis by other techniques. An important advantage of TPEA-PCR, is the relative ease of performing the method. Other known procedures are generally time consuming and complex, involving DNA purification and precipitation from one step to another. The present cDNA amplification technique however, can be carried out in a single tube with a need for only limited manual intervention. This therefore makes it possible to amplify large numbers of samples relatively easily. The ability to then analyse the expression of many genes of unrelated sequence, both at high and low abundance, in samples taken from as little as a single cell, will potentially allow it to be used in high throughput screening systems.

The invention can be used to analyse gene expression in samples as small as just a single cell (Figure 4), or in much larger samples such as 100 cells (Figure 2). Amplification from a single cell currently provides enough material for approximately 40 gene specific PCR reactions. Whilst this is already an improvement over existing protocols, it should

theoretically be possible to improve the efficiency of the TPEA reaction to provide far higher yields of 3' cDNA product. This would then not only allow the number of gene-specific PCR reactions performed on each sample to be increased, but more importantly allow the procedure to be linked to other analysis procedures. With improvements to the amplification regime and addition of fluorescent or radioactive nucleotide label to the reaction, it should be possible in the future to analyse the product using array based hybridization technologies (deRisi J *et al* (1996), Nature Genetics,14:457-460, Chee M *et al* (1996), Science, 274:610-614, Lockhart D J *et al* (1996), Nature Biotech, 14:1675-1680) which currently require relatively large amounts of RNA for a single assay. Such developments would potentially allow the expression profiling of hundreds or thousands of genes in samples derived from biopsies or single cells.

The inventors provide a rapid, robust and reproducible procedure, called Three Prime End Amplification optionally with PCR (TPEA-PCR) (TPEA), capable of amplifying 3' fragments of cDNA prior to analysis by other techniques. In the method of the invention, the 3' region of mRNA is amplified arbitrarily by PCR using a combination of primers. The amplified cDNA, which represents the most diverse region of gene sequence, can then be analysed by a second round of PCR using gene-specific primers. Using the invention it is possible to analyse the expression of, for example, up to 40 genes (20 in duplicate) in single human lymphoblastoma cells. The method is also suited to the analysis of genes

expressed at low levels in small populations of cells, eg expression of the adenosine A_{2a} receptor in cholinergic neurons of the rat striatum.

Sequence diversity between genes is at its greatest in the 3' untranslated region and this region provides the most unique target for gene-specific assays; this is especially important when wishing to differentiate between closely related members of a gene family. The procedures of the invention preferentially amplifies this portion of the mRNA sequences. cDNA synthesis by reverse transcriptase is initiated by an anchored oligo-dT priming so that the 3' region of all genes is represented in the resulting single-stranded cDNA. A 5'-specific heel may be incorporated into this primer for use in the subsequent amplification procedure. For second-strand synthesis, it is desireable that about 1 kb from the 3' end of each gene is selected and amplified by PCR Assuming a completely random length of nucleotide sequence, it would be expected that a given 5 base sequence would appear every 1024 bases (4^5), even though some nucleotide sequences are more common than others (Lopez-Nieto & Nigam (1996) Nature Biotech 14:857-861). A pentameric sequence is preferentially selected so that the primer initiates second-strand synthesis in an arbitrary manner within 1 kb of the 3' end of the mRNA. A search of 30 gene sequences reveals that at least one copy of this 5 base sequence was present in this region of each gene. 5' to this, 5 bases of random sequence (N_5) are preferably incorporated in order to stabilise the interaction of the arbitrary pentameric sequence, which in turn was flanked by a specific heel

sequence. After a single round of second-strand synthesis, each DNA strand contains a specific priming site 5' and 3' to the region of interest, thus allowing amplification of the intervening sequence. The majority of the mRNA species represented in the first-strand cDNA pool before amplification, as detected by conventional RT-PCR are also detectable after amplification.

The invention should permit the analysis of gene expression in samples obtained from small samples of tissue or single cells. In so doing, it should allow the utilisation of the wealth of new sequence data now available, to further understanding of disease processes and the cellular physiology of complex issues.

A major utility of TPEA-PCR will be in sampling single cells in tissues and in culture conditions to conduct detailed studies of temporal gene expression, changes in gene expression in response to growth conditions or environmental insults, and in identifying hitherto undetected gene activity associated with particular cellular events. Apart from the study of cells to understand their innate biology and responses, new approaches to toxicology profiling are promised as well as means to molecularly classify rare cell types.

The invention has many possible applications including

1. Making a single cell cDNA libraries for subsequent detailed analysis of gene expression, and the discovery of novel genes.
2. Real time profiles of gene expression in selected cells.
3. Preamplification of small (and single cell) samples for subsequent analysis of gene expression using hybridization based assays including those using cDNA and oligonucleotide arrays.
4. Analysis of gene expression in small tissue samples (diagnosis of cancer and other disease states).
5. Analysis of gene expression by PCR.
6. Amplification of full length RNA samples from single cells and small samples, for subsequent library making or expresion in expression systems.

Preferred embodiments of the invention will now be described in more detail by way of specific examples and drawings in which:

Figure 1 shows a schematic illustration of the process of TPEA-PCR. Details of the protocol are given in the Examples, as are the particular sequences of the anchored oligo (dT) primer and the partially degenerate second strand primer.

Figure 2 shows cDNA amplification. cDNA was prepared from varying numbers of sorted lymphoblastoma cells, and analysed by RT-PCR for the expression of 8 "housekeeping" genes, before (left) and after (right) TPEA-PCR. After cDNA amplification, expression of all 8 genes assessed could be detected after carrying out gene specific PCR on only 5% of the amplified cDNA generated from a single cell. Without cDNA amplification however, expression of none of the genes assayed could be detected at this level. Genes assayed: RPL21 (riboprotein L21), RP27a (riboprotein 27a), RPL28 (riboprotein L28), RPS5 (riboprotein S5), HSKPQZ7 (housekeeping protein), ACTB (Cytoplasmic beta-actin), G-3-PDH (Glyceraldehyde-3-phosphate dehydrogenase) and EF1 (Elongation factor 1).

Figure 3 shows multiple gene expression analysis in single lymphoblastoma cells. Four cells (1-4) were lysed, reverse transcribed, the cDNA amplified and gene specific PCR performed on the product. CD2, SI and intron primer pairs serve as negative controls to check for genomic contamination of the samples. Eleven of the other genes are expressed in all four cells in duplicate, while JUND, EF1 CDC25B and CD19 expression was not consistent between cells. Genes assayed: RPL5 (riboprotein L5), RPL21 (riboprotein L21), RP27a (riboprotein 27a), RPL28 (riboprotein L28), RPS5 (riboprotein S5), RPS9 (riboprotein S9), RPSI0 (riboprotein SI0), RPS29 (riboprotein S29), HSKPQZ7 (Housekeeping protein), ACTB (Cytoplasmic beta-Actin), G-3-PDH (Glyceraldehyde-3-phosphate dehydrogenase), EF1 (Elongation factor 1), JUND (JUND), CDC25B (cell cycle factor CDC25b) and the cell surface antigens CD19, CD79a, CD2, IGM

(immunoglobulin IgM), SI (the intestine-specific enzyme, sucrase-isomaltase).

Figure 4 shows adenosine 2a receptor expression in striatal cholinergic interneurons. Panel a shows an infrared video image of a rat striatal cholinergic interneuron during electrophysiological characterisation and panel b, after aspiration of cytoplasm. The expression of four housekeeping genes, the transmitter synthesising enzymes choline acetyltransferase (found only in cholinergic neurons) and glutamic acid decarboxylase (found in GABAergic, medium spiny neurons), three tachykinin (NK) receptors and the adenosine A_{2a} receptor was assessed in twenty six striatal cholinergic neurons. Two representative neurons are shown, neuron 1 expresses the A_{2a} receptor, neuron 2 does not, while the expression of the other genes tested are the same in both.

Example I - TPEA-PCR of Lymphoblastoma Cells

TPEA-PCR assay was performed on lymphoblastoma cells in the G0/G1 phase of the cell cycle. Groups of 100, 10 and single cells were flow sorted into wells containing lysis buffer and the mRNA reverse transcribed. A proportion of the sorted cells then underwent 3' end amplification, as described hereinafter. Figure 1 shows a schematic summary of the TPEA-PCR procedure. Gene-specific PCR assays for 8 'housekeeping' genes were carried out on lymphoblastoma cDNA, before and after cDNA amplification. Following reverse transcription only, the expression of each of the housekeeping genes

could be detected when cDNA generated from between 1 and 100 cells was used in each PCR assay as shown in Figure 2. It was not however possible to detect the expression of any of these genes when cDNA equivalent to less than one cell was assayed. Following TPEA-PCR however, the expression of all eight of the genes was detectable even when amplified cDNA generated from the equivalent of as little as 5% of one cell was present in the PCR assay (Figure 2).

Lymphoblastoma Cell Sorting

An Epstein Barr virus transformed lymphoblastoid cell line (HRC575, ECACC, Porton Down, UK) was maintained in log phase growth in RPMI 1640 medium supplemented with 16% foetal calf serum, 2mM L-glutamine and penicillin-streptomycin (100 U/ml and 100 mg/ml respectively). Cells at approximately 10^8 per ml were stained with the bixbenzimadazole dye Hoechsht 33342 (Sigma, Poole, UK) at 1 µg/ml for 30 minutes at 37°C. Cells were sorted by using the Autoclone attachment of a Coulter Elite ESP flow cytometer, 300 mW of all lines UV from a Coherent 306 laser and by using single drop and complete abort sorting settings. Time of flight, forward and right angle scatter, and Hoechst fluorescence peak and area measurements were used to ensure the sorting of single cells. The accuracy of sorting (both spatial and numerical) was tested by sorting single fluorescent beads (DNA Check, Coulter Corp) into 96 well plates and viewing the plates on a fluorescence microscope.

Reverse Transcription (RT) and cDNA Amplification

Lymphoblastoma cells were FACS sorted into 96 well plates containing 7 µl of freshly prepared lysis buffer (50 mM Tris-HC (pH 8.3), 75 mM KCl, 3mM $MgCl_2$, 5mM NP-40 (Sigma) and 1.5 units of RNase inhibitor (Pharmacia, Milton Keynes, UK). This buffer leaves the nucleus intact (Jena *et al* (1996), J Immunol Methods, 190:199-213). After 5 min on ice, nuclei were removed by centrifugation (8,000g, 5 min at 4°C), and the supernatant aspirated and the RNA reverse transcribed in a reaction volume of 10 µl containing; 1 x first-strand buffer, 200 Units M-MLV reverse transcriptase (Gibco BRL, Paisley, UK), 0.5 ng reverse transcription primer for 60 min at 37°C. The RT primer was composed of an anchored oligo(dT) primer with a specific 5' heel sequence: CTCTCAAGGATCTTACCGCTTTTTTTTTTTTTTT (A,G,C).

Second-strand cDNA synthesis was initiated by incubation of the first-strand cDNA with 1 ng of a primer consisting of (5' to 3'); a 20 base sequence selected due to its absence from the mammalian databases, a stretch of five random nucleotides and a defined pentameric sequence (CTGCATCTATCTAATGCTCCNNNNNCGAGA where N represents C, G, T or A) for 15 mins at 50°C under amplification conditions described below. Although this primer will undoubtedly prime second-strand DNA synthesis at many sites on the first strand cDNA, the subsequent PCR between the heel sequence of the oligo(dT) primer and the arbitrary primer closest to the 5' end, ensures amplification of cDNA sequences complementary to the 3' ends of the polyA tail. After allowing the second-strand primer to anneal, primer extension was

performed at 72°C for 10 min using AmpliTaq DNA polymerase (0.35 units, Applied Biosystems, Warrington, UK) in PCR buffer containing 67 mM Tris HCl (pH 8.3), 4.5 mM $MgCl_2$ and 0.5 mM dNTPs. Subsequently, 0.4 ng of 3' heel primer (CTCTCAAGGATCTTACCGC) was added and the reaction subjected to 10 cycles of 92°C for 2.5 min, 60°C for 1.5 min and 72°C extension of 1 min, followed by a final 10 min extension. A further 125 ng of second-strand primer and 50 ng of 3' heel primer were then added in 10 µl of PCR reaction mix. After 15 cycles (as before), a further 10 µl of PCR reaction mix containing 125 ng second-strand primer and 50 ng of 3' heel primer were added to the reaction and subjected to another 15 rounds of PCR. The final reaction mixture was then diluted to 200 µl with 10 mM Tris/0.1 mM EDTA (pH 8.1). 5 µl samples used for subsequent gene specific PCR assays.

Gene-specific PCR

Samples (5 µl) of amplified cDNA were subjected to hot-lid PCR carried out in 1 x PCR buffer (3.5 mM $MgCl_2$ pH 8.8) containing, 12.5% sucrose, 0.1 mM cresol red, 12 mM β-mercaptoethanol, 0.5 mM dNTPs (Pharmacia), 0.6 U AmpliTaq DNA polymerase (Applied Biosystems), and primers were used at 100 ng/reaction. Amplifications were carried out on PTC-225 thermal cyclers (Tetrad, MJ Research, US). Following an initial 2 min denaturing step (92°C), each PCR cycle consisted of 30 sec denaturing (92°C), 90 sec annealing (55°C), and 60 sec elongation (72°C). After the final cycle the reaction was held

for 10 min at 72°C. The PCR products were then separated on a 2.5% agarose gel, stained with ethidium bromide and photographed. All gene-specific primers are listed in table 1 set out below.

IDT Exhibit 1020.9007

Table 1. Primer Sequences used for gene specific assays

Human Primers

Gene	Symbol	Acc. No.	Forward	Reverse
Riboprotein 1.5	RPL5	U14966	GAACAGCGTAACTCCAGACATG	CCGCTCCTGAGCTCTGAG
Riboprotein L21	RPL21	U14967	GTGCGTATTGAGCACATAAAGC	TTCCAGCAGCTCAGGCTC
Riboprotein L27a	RPL27a	U14968	GACAAATTGTGGACTTTGGTCA	ATGTGCTTCAAGCCACCAG
Riboprotein L28	RPL28	U14969	ATCAACAAGAATGCTCGCG	TGACTTTATTGCTCTGGGGG
Riboprotein S5	RPS5	U14970	GTGAACCAGGCCATCTGG	CCAAAGGGCAGACAGGTTTA
Riboprotein S9	RPS9	U14971	CATTGTCCGCCTGGATTC	TGTTTATTTGGCAGGAAAACG
Riboprotein S10	RPS10	U14972	GGAGATTGTGCCTGCCAC	AAACTGGAATTCGGTTGCTG
Riboprotein S29	RPS29	U14973	ATTCGGCCAGGGTTCTC	TTTTTCATTGAGTAGATGCCCC
Housekeeping protein Q27	HSKPQZ7	M81806	GGAACTTCCTCTGGGAACCT	GGAGGTCAAGTCAAGCTCCA
Glyceraldehyde-3-Phosphate Dehydrogenese	G-3-PDH	M33197	CGACCACTTTGTCAAGCTCA	AGGGGTCTACATGGCAACTG
Cytoplasmic beta-actin	ACTB	AB004047	CGTGGACATCCGTAAAGACC	ACATCTGCTGGAAGGTGGAC
JUND	JUND	X56681	CTCGGGGAACAAACGTTG	ACAAACAGGAATGTGGACTCG
Elongation factor 1	EF1	M82882	CCAAGAAAGCAACTTGAGCC	TGCAGAAGACATTTTTTTCTTG
Cell division cycle 25B	CDC25B	M81934	AGTGCCTTGCATACCCAAAC	CCCCCATCTTGAGATTCAGA
CD19	CD19	X13312	CCAACCTCTGGAGCAATGTT	GGAATACAAAGGGGACTGGA
CD79A	CD79A	L32754	CTTCTGGGGGCTTCCTTAGT	GTTAGGAGGTGGGGCAGTTT3
Immunoglobulin M	IGM	X17115	GTCCCCGGAGAAGTATGTGA	GACTTGTCCACGGTCCTCTC
CD2	CD2	M16445	TCTTCGAACTCAGCCATGTG	GGCTGCTTGTAGTGAGACCC
Sucrase-isomaltase	SI	X63597	CAGGGTTCTCTGTTTTTGGGA	GGAAGCGAATTTTTATTTCCG
INTRON			GTGGATATAGTTGAGTCC	CACTGTTGACCACTCCAT

IDT Exhibit 1020.9008

Rat Primers

Gene	Symbol	Acc. No.	Forward	Reverse
Beta-actin	Aclb	V01217	CATCGATGCCCTGAGTCC	ACACCTCAAACCACACTCCCAG
Alpha-tubulin	Tuba	V01226	CAGTGGTACGTGGGTGAGG	TTTGACATGATACAGGGACTGC
Cyclophilin	Cyca	M19533	ACTGCCAAGACTGAGTGGCT	AATGGTTTGATGGGTAAAATGC
Ribosomal protein L18	Rpl18	M20156	AGGTGTACCGACACACTTTGGC	TTTTGTAGCCTCGGCTGG
Choline acetyllransferase	Cha1	*	TACTAAGCTCTGTTCCCATCCC	ACCCAGGTTGCTTCCAAAC
Glulamic acid decarboxylase	Gad67	X57573	ATCTTGCTTCAGTAGCCTTTGC	TGTCTTCAAAAACACTTGTGGG
Neurokinin receptor 1	Nk1	J05097	CTGGAAAGAGGAGCCTTGTG	CTGAGACGGAAAGGAACAGC
Neurokinin receptor 2	Nk2	M31838	TTCTGCAGTGAGGAGCTGG	TTGGCTTTCAGAGGGCAC
Neurokinin receptor 3	Nk3	J05189	TTTGTGACCGCAGAAGAGC	TTGTGTGTGGGAAGTCAGTCA
Adenosine receptor 2a	Adora2a	M91214	TCTGACCAACAAAGCTGGC	TGGAAGGAAAGGCAGTAGTCA

* Brice A., et al (1989) J Neurosci. Res. 23:266-273

IDT Exhibit 1020.9009

Example II

Reproducibility and Specificity of 3' Amplification of cDNA from Single Cells

The reproducibility of this technique was assessed when the expression of 20 genes was examined in duplicate on single lymphoblastoma cells. Duplicate PCR assays revealed almost identical expression profiles both between assays on a single cell and between different cells (Figure 3). The expression of 11 housekeeping genes was reproducible in all of these single cell expression profiles, demonstrating consistent amplification of cDNA from the transcripts of these genes. However four genes (JUND, EF1 CDC25B and CD19) were not found to be expressed in every cell tested. Two gene-specific assays, sucrase-isomaltase and CD2, were included as negative controls as they were not expected to be expressed in these cells since their mRNAs have only previously been observed in the gastrointestinal tract (Chadrasena G *et al* (1992), Cell Mol Biol, 38:243-254) and in populations of T-cells (Sewell W *et al* (1986) Proc Natl Acad Sci USA, 83:8718-8722), respectively. Another set of primers, designed to amplify an intronic sequence from a gene found in the Xq 2.5 region were used to detect genomic contamination. Since these three sets of primers were negative in all the cells examined (n=20), and expression of any gene was not detected without prior reverse transcription, false positives arising from amplification of genes not expressed in these cells

(SI and CD2) or genomic contamination, do not appear to complicate interpretation of results.

Figure 3 clearly demonstrates the reproducibility of the gene-specific assays following amplification from single cells, with each of 11 housekeeping gene assays being detectable in duplicate reactions on each of four lymphoblastoma cells. There was some apparent variability in the expression of four genes (JUND, EF1 CDC25B and CD19) studied in these cells. It is unclear at this stage however, whether this represents experimental variability, fluctuations in transcriptional activity within these cells or 'real' consistent differences in the expression of these genes in cells that we can only assume to be homogeneous. Such variability in gene expression has been encountered in other cell groups thought to be homogeneous (O'Dowd, D K & Smith M A (1996), Mol Neurobiol, 13:199-211). This analysis also demonstrates that the procedure does not lead to false positives due to either over amplification (as no signal was detected for genes (SI and CD2) known not to be expressed in these cells) or genomic contamination (as demonstrated by the lack of signal from intronic primers or from the gene-specific primers when the cell contents were not reverse transcribed prior to amplification).

The power of this technique lies in its potential to facilitate expression profiling of cells derived from complex cell populations, even when they form only a small proportion of the population as a whole, and in its ability to detect low abundance transcripts. Expression of the neurokinin (NK) receptors and the

adenosine A_{2a} receptor were investigate in single striatal cholinergic interneurons which constitute a small fraction of the total cellular mass of the striatum. Of the neurokinin receptors, the NK1 receptor is widely accepted as being expressed in these cells (Kawaguchi Y *et al* (1995), Trends in Neurosci, 18:527-535). The expression of this gene was examined as an example of an mRNA species expressed at far lower levels than housekeeping genes. In contrast, there is considerable controversy as to whether the A_{2a} receptor is expressed in cholinergic interneurons (Schiffman S N *et al* (1991), J Neurochem, 57:1062-1067; Fredholm B B & Svenningsson P (1998), Trends in Pharmacol Sci, 19:46-47, Richardson P J *et al* (1997), Trends in Pharmacol Sci, 18:338-344, Richardson P J *et al* (1998), Trends in Pharmacol Sci, 19:47-48), suggesting that the corresponding mRNA species may be present at low levels as suggested by one *in situ* hybridization study (Dixon A K *et al* (1996), Br J Pharmacol, 188:1461-1468) or not present at all, as suggested by other studies (Svenningsson P *et al* (1997) Neurosci, 80:1171-1185). Expression of the A_{2a} receptor in these cells has important implications for the mechanism of action of adenosine and because of the potential of the A_{2a} receptor to act as a target for novel drugs for the amelioration of Parkinson's Disease (Richardson P J *et al* (1997), Trends in Pharmacol Sci, 18:338-344). Using a patch pipette it was possible to harvest upwards of an estimated 40% of the cellular contents of these neurons (figure 4a). This was then subjected to the cDNA amplification procedure, followed by gene-specific PCR assays. Testing for the expression of

four housekeeping genes and the marker enzymes choline acetyltransferase and *Gad67*, was included in order to demonstrate the quality of collection and amplification, and to corroborate the cell lineage, respectively. NK1 receptor mRNA was detected in all the cholinergic neurons tested, although the NK2 and NK3 receptors were not, confirming that Substance P exerts its effects on these cells via the NK1 receptor (Bell M I *et al* (1998), Neurosci in press). Expression of the adenosine A_{2a} receptor mRNA was detected in 27% of the cholinergic neurons assayed (Figure 4b), a percentage close to that observed previously (Dixon A K *et al* (1996), Br J Pharmacol, 118:1461-1468). It is not clear whether the apparent heterogeneity in expression of this receptor is due to differences in the temporal expression of this gene (i.e. that all of these cells possess A_{2a} receptor protein, but only 27% actually express the gene at any one time), or to an absolute difference in gene expression within the striatal cholinergic interneuron population.

Example III

Extraction of Neuronal Contents

300 μm coronal slices from 14-28 day-old male Sprague Dawley rats containing the striatum were viewed with a Zeiss Axioskop microscope (Carl Zeiss Ltd, Welwyn Garden City, UK) fitted with a x64 water-immersion objective lens together with gradient contrast optics (Luigs and Neumann, Ratingen, Germany). Light in the infrared range (>740 nm) was used in conjunction with a contrast-enhancing Newvicon camera (Hamamatsu,

Hamamatsu City, Japan) to resolve individual neurones within slices (Stuart G J *et al* (1993), Pflügers Archiv, 423:511-518). The cytoplasm from large cells (>30 μm in one dimension) was gently aspirated under visual control into a patch-clamp recording electrode until at least 40% of the somatic cytoplasm had been collected. The electrode was then withdrawn from the cell to form an outside-out patch which prevented contamination when the electrode were forced into a microtube and reverse transcribed, subjected to 3' cDNA amplification, and 2.5% of the product used in each gene specific PCR reaction.

Example IV

Analysis of Complex Cell Systems: Expression Profiling of Neurons

Having shown that TPEA-PCR permits expression profiling of single cells, the analysis of a complex cell system can be made *in vivo*. Striatal cholinergic interneurons are readily identifiable in rat brain slices due to their large size (>30 μm diameter) when compared to surrounding cell types, which are predominately medium spiny neurons (<15 μm diameter). After electrophysical characterisation of the cells (Lee K *et al* (1997) J Neurochem, 69:1774-1776), cytoplasmic samples were taken using a patch pipette, reverse transcribed and the cDNA amplified. The expression of a variety of genes of was then investigated (two representative expression profiles of cholinergic interneurons are shown in Figure 4). The expression of four known housekeeping genes was demonstrated, confirming the integrity of sample

collection and the RNA. In addition, the expression of choline acetyltransferase (the acetylcholine synthesising enzyme) was observed in each sample, thus unequivocally confirming the cholinergic phenotype of the sampled neurons. In order to control against the possibility of contamination by the surrounding population of medium spiny neurons, which are known to express the adenosine A_{2a} receptor (Schiffman S N *et al* ((1991), J Neurochem, 57:1062-1067), the samples were assayed for the presence of the mRNA of glutamic acid decarboxylase (*Gad67*), which is highly expressed in medium spiny, but not cholinergic neurons. The *Gad67* primers amplify relevant mRNA from medium spiney neurons. Given the absence of any *Gad67* expression in cholinergic interneurons,the failure to see any amplification of *Gad67* mRNA in the sample demonstrates that samples were not contaminated with medium spiney neurons. All neurons tested for the expression of the neurokinin (NK) receptors were negative for NK2 and NK3 receptor mRNA, but positive for that of the NK1 receptor. 27% (7/26) of the cholinergic interneurones tested expressed the adenosine A_{2a} receptor.

IDT Exhibit 1020.9015

Claims:

1. A process of reverse transcribing mRNA species present in a sample from an organism comprising:

- reverse transcribing the mRNA species using a first heeled primer, thereby to provide first strand cDNA species;

- synthesising second cDNA strands using a second heeled primer population, the nucleotide sequences of the non-heel portions of the second heeled primers being such that the reverse transcribed first strand cDNA species are capable of hybridising to at least one second primer.

2. A process of reverse transcribing and amplifying mRNA species present in a sample from an organism comprising:

- reverse transcribing the mRNA species using a first heeled primer, thereby to provide first strand cDNA species;

- synthesising second cDNA strands using a second heeled primer population, the nucleotide sequences of the non-heel portions of the second heeled primers being such that the reverse transcribed first strand cDNA species are capable of hybridising to at least one second primer.

- amplifying the resulting cDNA species.

IDT Exhibit 1020.9016

3. A process as claimed in claim 2, wherein the amplification of the resulting cDNA species employs a third primer comprising at least a part of the heel portion of the first heeled primer and a fourth primer comprising at least part of the heel portion of the second heeled primer.

4. A process as claimed in any one of claims 1 to 3, wherein the second heeled primer population comprises primers differing by up to five nucleotide bases, the population comprising a number of primers in the range 1000 to 100,000 primers, preferably in the range 1024 to 65536 primers.

5. A process as claimed in claim 4, wherein the primers of the second heeled primer population each comprise a random sequence of nucleotides in the range of 5 to 8 nucleotides 3' to the heel and a further sequence of at least 5 nucleotides contiguous 3' therewith.

6. A process as claimed in claim 5, wherein the said further sequence of nucleotides is selected by sequence analysis of known sequences so as to promote the ability of the second heeled primer as a whole to hybridise to the transcribed cDNA species.

7. A process as claimed in claim 6, wherein the said further sequence of nucleotides comprises a number of nucleotides in the range 2 to 10 nucleotides.

8. A process as claimed in claim 7, wherein the said further sequence of nucleotides comprises a

IDT Exhibit 1020.9017

number of nucleotides equivalent to the number of nucleotides in the random sequence of nucleotides.

9. A process as claimed in any preceding claim, wherein a second heeled primer hybridises on average once in every 1kb portion of a first strand cDNA species.

10. A process as claimed in any one of claims 5 to 9, wherein the said further sequence of nucleotides is:

CGAGA

11. A process as claimed in any one of claims 5 to 10, wherein the second heeled primer is:

CTGCATCTATCTAATGCTCCNNNNNCGAGA

wherein N is independently selected from C G T or A.

12. A process as claimed in any preceding claim, wherein the heel portion of the first and second heeled primers are selected so that they lack the ability to hybridise to mRNA or first strand cDNA respectively

13. A process as claimed in claim 12, wherein the heel portions are selected by an analysis of known nucleotide sequence information.

14. A process as claimed in claim 12 or claim 13, wherein the heel portions comprise sequences absent from the mRNA species in the sample.

IDT Exhibit 1020.9018

15. A process as claimed in any one of claims 13 to 14, wherein the heel portions comprise sequences absent from the genome of the organism from which the sample is taken.

16. A process as claimed in any preceding claim, wherein the heel portions comprise a number of nucleotides in the range 15 to 30, preferably 18 to 22 nucleotides.

17. A process as claimed in any preceding claim, wherein the first heeled primer is an anchored primer comprising an oligo d(T) sequence.

18. A process as claimed in any preceding claim, wherein the fourth primer is the heel of the second heeled primer.

19. A process as claimed in any preceding claim, wherein the third primer is the heel of the first heeled primer.

20. A process as claimed in any one of claims 1 to 16, wherein the third primer is the same as the first heeled primer.

21. A process as claimed in any preceding claim, wherein the amplification of the resulting cDNA species comprises more than one round of amplification cycles.

22. A process as claimed in claim 21, wherein each further round of amplification comprises addition of

IDT Exhibit 1020.9019

further second and third primers and amplification reagents.

23. A process as claimed in any preceding claim, wherein each round of amplification comprises 5 to 45 cycles, preferably 10 to 40 cycles.

24. A process as claimed in any one of claims 21 to 23, wherein the first round of amplification comprises a lesser number of cycles than any further rounds of amplification.

25. A process as claimed in any one of claims 21 to 24, wherein the amplification process is PCR.

26. A process as claimed in any one of claims 21 to 25, wherein each PCR cycle comprises X_1°C for Y_1 min; then X_2°C for Y_2 min; then X_3°C extension for Y_3 minute; then Y_4 min extension, wherein $X_1 > X_3 > X_2$ and $Y_4 > Y_3 > Y_2 > Y_1$ and wherein X_1 is in the range 90 to 94°C, X_2 is in the range 45 to 70°C, X_3 is in the range 65 to 75°C and Y_1, Y_2, Y_3 are in the range 15 seconds to 4 minutes and Y_4 is in the range 2.5 to 20 minutes.

27. A process as claimed in claim 26, wherein X_1 = 92 °C, X_2 = 60°C, X_3 = 72°C, Y_1 = 2.5 minutes, Y_2 = 1.5 minutes, Y_3 = 1 minute and Y_4 = 10 minutes.

28. A process as claimed in any preceding claim, wherein during the amplification of the cDNA strands, further primers are included for amplification of

IDT Exhibit 1020.9020

contaminating nucleotide sequences suspected as being present in the sample.

29. A process as claimed in any preceding claim, wherein the sample from an organism is a cytoplasmic sample obtained from a single cell type, preferably from a single cell.

30. A process as claimed in claim 29, wherein the cytoplasm is obtained by lysis or aspiration of a cell or cells.

31. A process as claimed in claim 29 or claim 30, wherein cells are obtained by fluorescence activated cell sorting (FACS).

32. A method of reverse transcribing expressed gene sequences in a sample from an organism comprising reverse transcribing the mRNA in the sample using a first primer to produce first strand cDNA species, synthesising second cDNA strands using a population of second primers, wherein at least one second primer in the population hybridises to a given first strand cDNA species.

33. A method as claimed in claim 28, further comprising amplification of the resulting double stranded cDNA.

34. A method as claimed in claim 32 or claim 33 further comprising the features of any of claims 1 to 31.

IDT Exhibit 1020.9021

35. A polynucleotide primer for reverse transcription of mRNA species comprising an oligo (dT) sequence and 5' thereto a polynucleotide heel sequence, wherein the heel sequence is substantially incapable of hybridisation to the mRNA species.

36. A primer as claimed in claim 35 being an anchored primer.

37. A primer as claimed in claim 35 or claim 36 further comprising the features of any of claims 12 to 16.

38. A polynucleotide primer population for synthesis of second strand cDNA species from first strand cDNA species, wherein at least one primer in the population is capable of hybridising to a given first strand of cDNA.

39. A primer population as claimed in claim 38, wherein at least one primer in the population is capable of hybridising at least once in any given 1kb of first strand cDNA.

40. A primer population as claimed in claim 38 or claim 39, wherein the primers further comprise the features of any of claims 4 to 16.

41. Polynucleotide primers for amplification of cDNA comprising a primer as claimed in any one of claims 35 to 37 and a primer comprising at least a portion of the heel portion of the primers claimed in claim 40 or claim 40 when dependent on any of claims 4 to 16.

IDT Exhibit 1020.9022

42. The use of a polynucleotide comprising an oligo d(T) sequence and a heel sequence 5' thereto for the reverse transcription of mRNA species in a sample.

43. The use as claimed in claim 42, wherein the polynucleotide further comprises the features of any of claims 35 to 37.

44. The use of a polynucleotide primer population of any of claims 38 to 40 for the synthesis of second strand cDNA from a population of first strand cDNA species.

45. A cDNA library preparation obtainable by a process of any of claims 1 to 31 or a method of any of claims 32 to 34, said library comprising substantially all cDNA species corresponding to genes expressed by a single cell, cell type or tissue.

46. A kit for the production of cDNA from mRNA in a sample from an organism comprising a primer of any of claims 35 to 37 and a primer population of any of claims 38 to 40.

47. A kit as claimed in claim 46, further comprising at least one further primer for achieving amplification of the cDNA.

48. A kit as claimed in claim 46, further comprising a primer pair for amplification of the cDNA.

IDT Exhibit 1020.9023

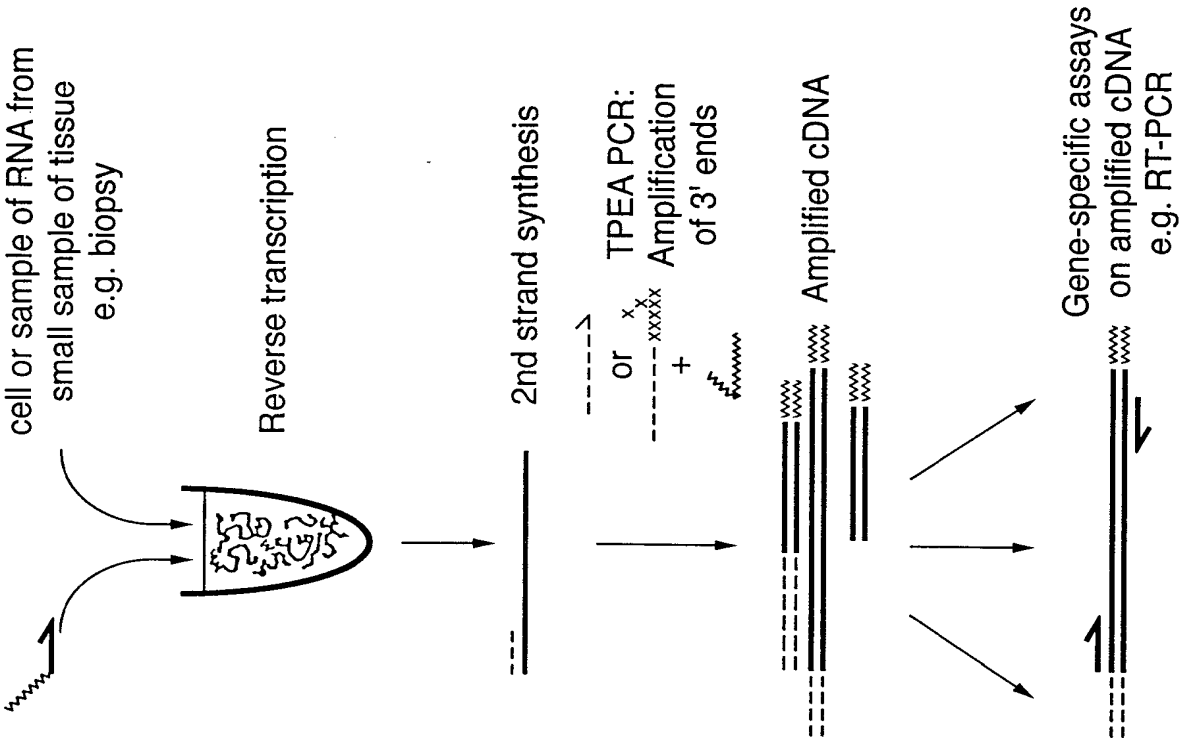


Fig.1.

Anchored oligo-dT primer with heel

e.g. CTCTCAAGGATCTTACCGCTTTTTTTTTTTTTTT(A,G,C)

Specific 'heel' sequence for amplification

Oligo-dT with 1 bp (A,G or C) anchor to prime first strand cDNA synthesis at the start of the polyA tail on the 3' end of mRNAs

Partially degenerate primer with heel

e.g. CTGCATCTATCTAATGCTCCNNNNCGAGA

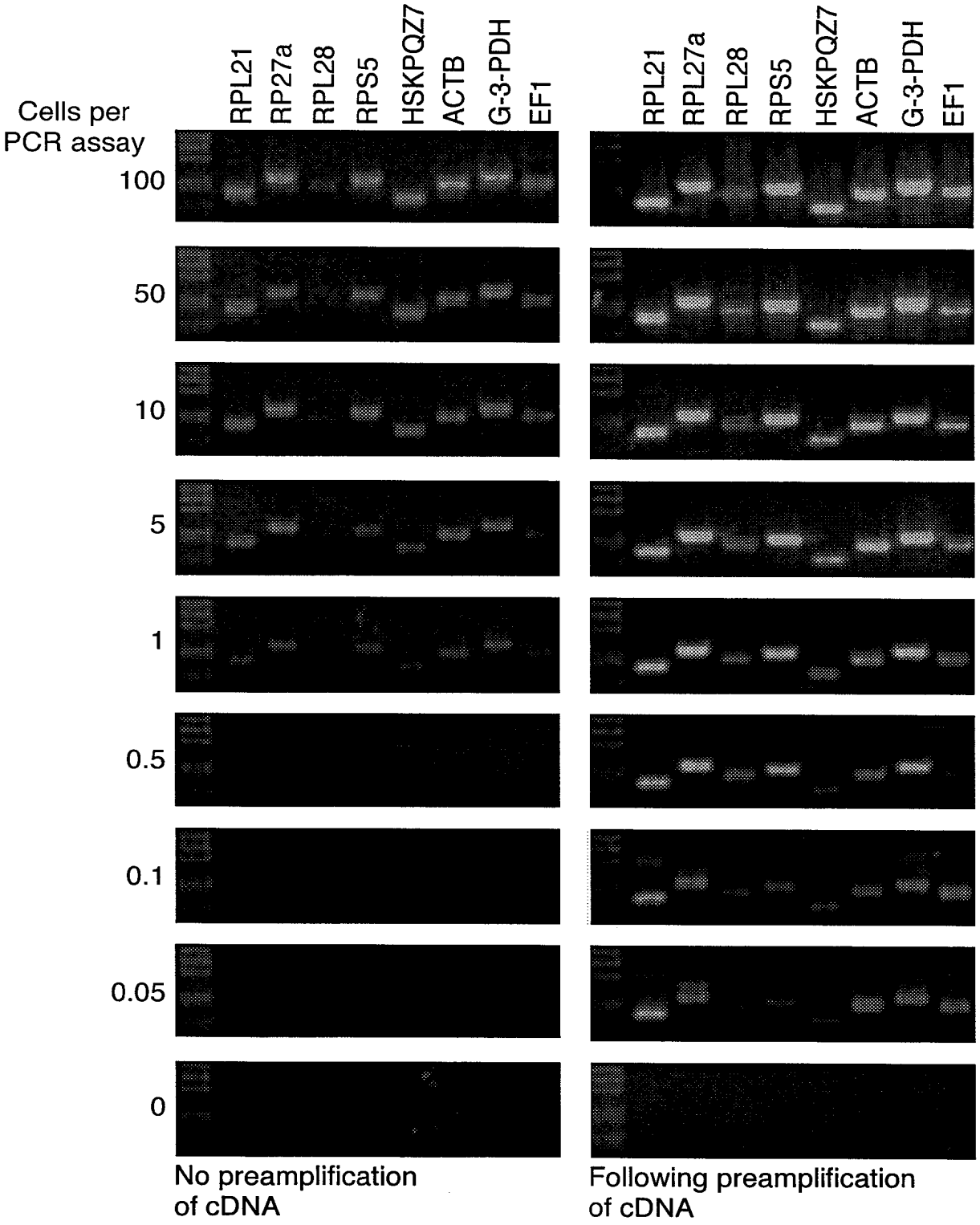
Specific 'heel' sequence for amplification

Stretch of nucleotides of an arbitrary but defined sequence designed to initiate 2nd strand synthesis. In the primer shown here, the 5 bp should prime approximately every 1 kb along the length of target mRNAs

Stretch of random nucleotides to stabilise the interactions of the 3' arbitrary portion of the primer with target sequences, where N represents C,G,T, or A

IDT Exhibit 1020.9024

Fig.2.



IDT Exhibit 1020.9025