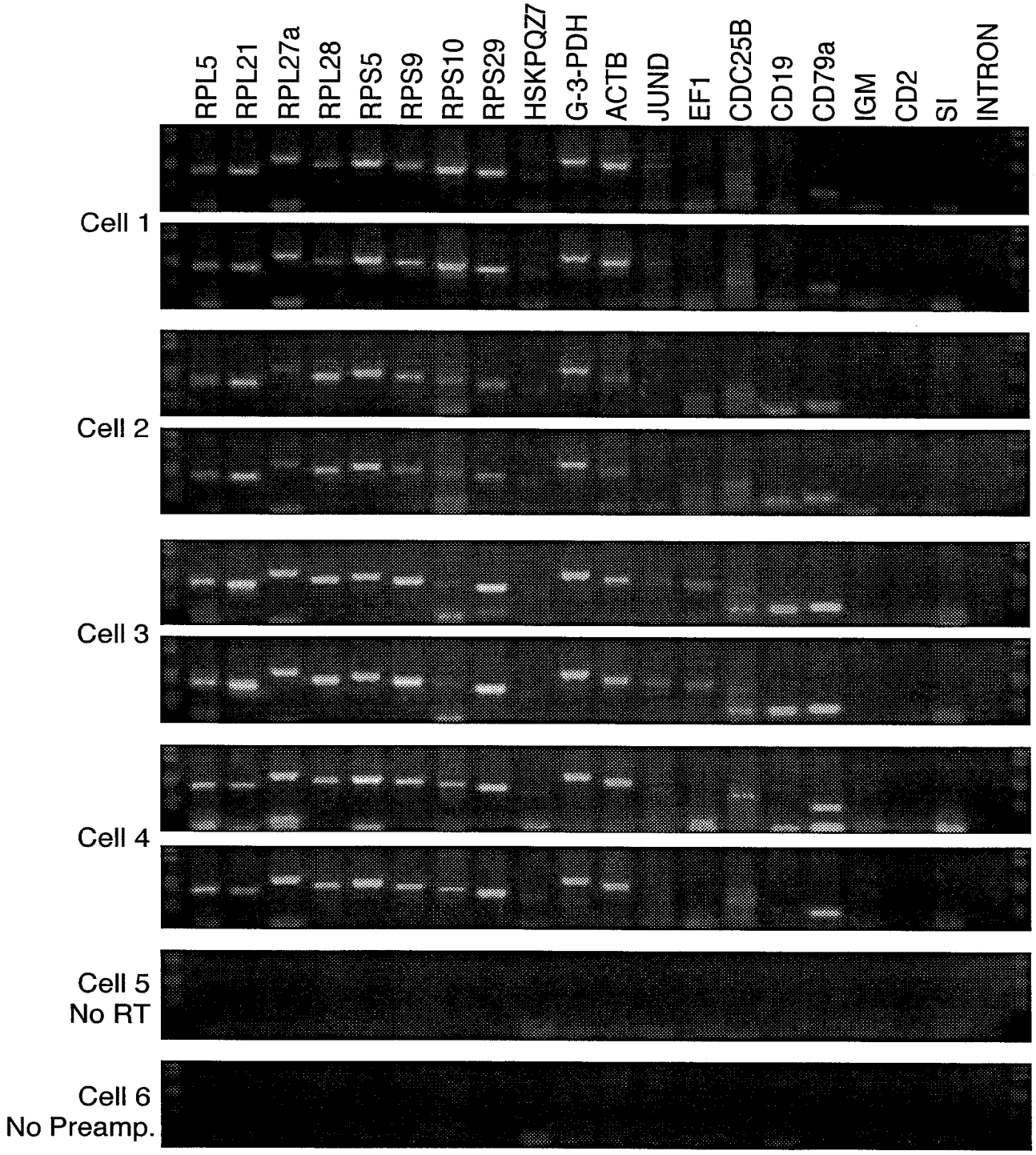
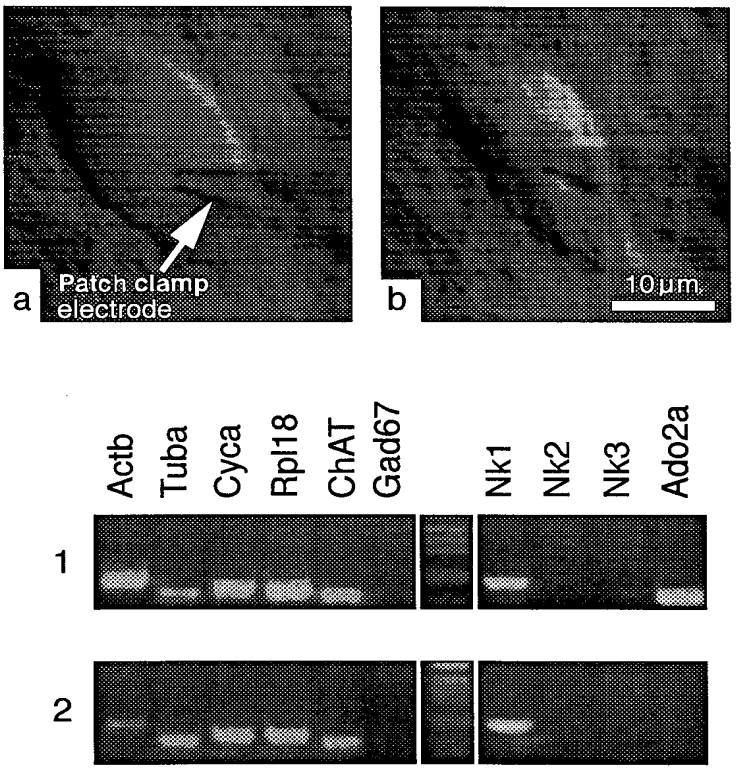
Fig.3.



IDT Exhibit 1020.9026

Fig.4.



IDT Exhibit 1020.9027

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 667 393 A2**

(12) **EUROPEAN PATENT APPLICATION**

(21) Application number: **95100438.1**

(51) Int. Cl.[6]: **C12N 15/10**, C12Q 1/68

(22) Date of filing: **13.01.95**

(30) Priority: **13.01.94 US 182621**

(43) Date of publication of application:
16.08.95 Bulletin 95/33

(84) Designated Contracting States:
AT BE CH DE DK ES FR GB GR IE IT LI LU MC NL PT SE

(71) Applicant: **ENZO DIAGNOSTICS, INC.**
60 Executive Boulevard
Farmingdale,
New York 11735-4716 (US)

(72) Inventor: **Engelhardt, Dean L.**
173 Riverside Drive (No. 19A)
New York,
New York 10025 (US)

Inventor: **Stavrianopoulos, Jannis G.**
99 South Clinton (No.11D)
Bay Shore,
New York 11706 (US)

Inventor: **Rabbani, Elazar**
69 Fifth Avenue (No. 19A)
New York,
New York 10003 (US)

Inventor: **Donegan, James J.**
210 East Broadway (No. 3G)
Long Beach,
New York 11561 (US)

(74) Representative: **VOSSIUS & PARTNER**
Siebertstrasse 4
D-81675 München (DE)

(54) **Novel process, construct and conjugate for producing multiple nucleic acid copies.**

(57) This invention provides *inter alia* an *in vitro* process for producing multiple specific nucleic acid copies in which the copies are produced under isostatic conditions, e.g., temperature, buffer and ionic strength, and independently of any requirement for introducing an intermediate structure for producing the copies. In other aspects, the invention provides *in vitro* processes for producing multiple specific nucleic acid copies in which the products are substantially free of any primer-coded sequences, such sequences having been substantially or all removed from the product to regenerate a primer binding site, thereby allowing new priming events to occur and multiple nucleic acid copies to be produced. This invention further provides a promoter-independent non-naturally occurring nucleic acid construct that produces a nucleic acid copy or copies without using or relying on any gene product that may be coded by the nucleic acid construct. Another aspect of this invention concerns a protein-nucleic acid construct in the form of a conjugate linked variously, e.g., covalent linkage, complementary nucleic acid base-pairing, nucleic acid binding proteins, or ligand receptor binding. Further disclosed in this invention is an *in vivo* process for producing a specific nucleic acid in which such a protein-nucleic acid construct conjugate is introduced into a cell. A still further aspect of the invention relates to a construct comprising a host promoter, second promoter and DNA sequence uniquely located on the construct. The host transcribes a sequence in the construct coding for a different RNA polymerase which after translation is capable of recognizing its cognate promoter and transcribing from a DNA sequence of interest in the construct with the cognate promoter oriented such that it does not promote transcription from the construct of the different RNA polymerase.

EP 0 667 393 A2

IDT Exhibit 1020.9028

This invention relates to the field of *in vitro* and *in vivo* production of nucleic acid production and to nucleic constructs and protein-nucleic acid conjugates for use in such production.

All patents, patent publications, scientific articles, and videocassettes cited or identified in this application are hereby incorporated by reference in their entirety in order to describe more fully the state of the art to which the present invention pertains.

Current methodology cited heretofore in the literature relating to amplification of a specific target nucleic acid sequence *in vitro* essentially involve 2 distinct elements:

1. repeated strand separation or displacement or a specific "intermediate" structure such as a promoter sequence linked to the primer or introduction an assymetric restrictrion site not originally present in the nucleic acid target; followed by
2. production of nucleic acid on the separated strand or from an "intermediate" structure.

Separation can be accomplished thermally or by enzymatic means. Following this separation, production is accomplished enzymatically using the separated strands as templates.

Of the established amplification procedures, Polymerase Chain Reaction (PCR) is the most widely used. This procedure relies on thermal strand separation, or reverse transcription of RNA strands followed by thermal dissociation. At least one primer per strand is used and in each cycle only one copy per separated strand is produced. This procedure is complicated by the requirement for cycling equipment, high reaction temperatures and specific thermostable enzymes. (Saiki, et al., Science 230:1350-1354 (1985); Mullis and Faloona, Methods in Enzymology 155: 335-351 (1987); U.S. Patent Nos. 4,683,195 and 4,883,202).

Other processes, such as the Ligase Chain Reaction (LCR) (Backman, K., European Patent Application Publication No. 0 320 308; Landegren, U., et al. Science 241 1077 (1988); Wu, D. and Wallace, R.B. Genomics 4 560 (1989); Barany, F. Proc. Nat. Acad. Sci USA 88:189 (1991)), and Repair Chain Ligase Reaction (RLCR) or Gap Ligase Chain Reaction (GLCR) (Backman, K. et al. (1991) European Patent Application Publication No. 0 439 182 A; Segev, D. (1991) European Patent Application Publication No. 0 450 594) also use repeated thermal separation of the strands and each cycle produces only one ligated product. These procedures are more complicated than PCR because they require the use of an additional thermostable enzyme such as a ligase.

More complicated procedures are the Nucleic Acid Sequence Based Amplification (NASBA) and Self Sustained Sequence Reaction (3SR) amplification procedures. (Kwoh, D.Y. et al., Proc Nat Acad Sci., USA., 86:1173-1177 (1989); Guatelli,J.C. et al., 1990 Proc Nat Acad Sci., USA 87:1874-1878 (1990) and the Nucleic Acids Sequence Based Amplification (NASBA) (Kievits,T., et al J. Virol. Methods 35:273-286 (1991); and Malek, L.T., U.S. Patent No. 5,130,238). These procedures rely on the formation of a new "intermediate" structure and an array of different enzymes, such as reverse transcriptase, ribonuclease H, T7 RNA polymerase or other promotor dependant RNA polymerases and they are further disadvantaged by the simultaneous presence of ribo- and deoxyribonucleotide tripohsphates precursors.

For the intermediate construct formation, the primer must contain the promotor for the DNA dependent RNA polymerase. The process is further complicated because the primer is, by itself, a template for the RNA polymerase, due to its single-stranded nature.

The last of the major amplification procedures is Strand Displacement Amplification (SDA) (Walker, G.T. and Schram, J.L., European Patent Application Publication No. 0 500 224 A2; Walker, G.T. et al. European Patent Application No. 0 543 612 A2; Walker, G.T., European Patent Application Publication No. 0 497 272 A1; Walker, G.T. et al., Proc Natl Acad Sci USA 89:392-396 (1992); and Walker, G.T. et al., Nuc Acids Res. 20:1691-1696 (1992)). The intermediate structure of this procedure is formed by the introduction of an artificial sequence not present in the specific target nucleic acid and which is required for the assymetric recognition site of the restriction enzyme. Again this procedure involves more than one enzyme and the use of thio nucleotide triphosphate precursors in order to produce this assymetric site necessary for the production step of this amplification scheme.

The random priming amplification procedure (Hartley, J.L., U.S. Patent No. 5,043,272) does not relate to specific target nucleic acid amplification.

Probe amplification systems have been disclosed which rely on either the amplification of the probe nucleic acid or the probe signal following hybridization between probe and target. As an example of probe amplification is the Q-Beta Replicase System (Qβ) developed by Lizardi and Kramer and their colleagues. Qβ amplification is based upon the RNA-dependent RNA polymerase derived from the bacteriophage Qβ. This enzyme can synthesize large quantities of product strand from a small amount of template strand, roughly on the order of 10^6 to 10^9 (million to billion) increases. The Qβ replicase system and its replicatable RNA probes are described by Lizardi et al., "Exponential amplification of recombinant RNA hybridization probes," Biotechnology 6:1197-1202 (1988); Chu et al., U.S. Patent No. 4,957,858; and well as by Keller and Manak (DNA Probes, MacMillan Publishers Ltd, Great Britain, and Stockton Press (U.S. and Canada,

IDT Exhibit 1020.9029

1989, pages 225-228). As discussed in the latter, the Qβ replicase system is disadvantaged by non-specific amplification, that is, the amplification of non-hybridized probe material, which contributes to high backgrounds and low signal-to-noise ratios. Such attendent background significantly reduces probe amplification from its potential of a billion-fold amplification to something on the order of 10^4 (10,000 fold). In addition, the Q beta amplification procedure is a signal amplification - and not a target amplification.

In vivo

Literature covering the introduction of genes or antisense nucleic acids into a cell or organism is very extensive (Larrick, J.W. and Burck, K. Gene Therapy Elsevier Science Publishing Co., Inc, New York (1991); Murray, J.A.H. ed Antisense RNA and DNA, Wiley-Liss, Inc., New York (1992)). The biological function of these vectors generally requires inclusion of at least one host polymerase promoter.

The present invention as it relates to *in vitro* and *in vivo* production of nucleic acids is based on novel processes, constructs and conjugates which overcome the complexity and limitations of the above-mentioned documents.

The present invention provides an *in vitro* process for producing more than one copy of a specific nucleic acid in which the process is independent of any requirement for the introduction of an intermediate structure for the production of the specific nucleic acid. The process comprises three steps, including (a) providing a nucleic acid sample containing or suspected of containing the sequence of the specific nucleic acid; (b) contacting the sample with a three component reaction mixture; and (c) allowing the mixture to react under isostatic conditions of temperature, buffer and ionic strength, thereby producing more than one copy of the specific nucleic acid. The reaction mixture comprises: (i) nucleic acid precursors, (ii) one or more specific nucleic acid primers each of which is complementary to a distinct sequence of the specific nucleic acid, and (iii) an effective amount of a nucleic acid producing catalyst.

In another aspect, the present invention provides an *in vitro* process for producing more than one copy of a specific nucleic acid in which the products are substantially free of any primer-coded sequences. Such a process comprises the following steps, including (a) providing a nucleic acid sample containing or suspected of containing the sequence of the specific nucleic acid; (b) contacting the sample with a three component mixture (the mixture comprising (i) nucleic acid precursors, (ii) one or more specific polynucleotide primers comprising at least one ribonucleic acid segment each of which primer is substantially complementary to a distinct sequence of the specific nucleic acid, and (iii) an effective amount of a nucleic acid producing catalyst); and (c) allowing the mixture to react under isostatic conditions of temperature, buffer and ionic strength, thereby producing at least one copy of the specific nucleic acid; and (d) removing substantially or all primer-coded sequences from the product produced in step (c). By removing such sequences, a primer binding site is regenerated, thereby allowing a new priming event to occur and producing more than one copy of the specific nucleic acid.

The present invention also provides an *in vitro* process for producing more than one copy of a specific nucleic acid in which the products are substantially free of any primer-coded sequences. In the steps of this process, said process comprising a nucleic acid sample containing or suspected of containing the sequence of the specific nucleic acid is provided, and contacted with a reaction mixture. The mixture comprises (i) unmodified nucleic acid precursors, (ii) one or more specific chemically-modified primers each of which primer is substantially complementary to a distinct sequence of said specific nucleic acid, and (iii) an effective amount of a nucleic acid producing catalyst. The mixture thus contacted is allowed to react under isostatic conditions of temperature, buffer and ionic strength, thereby producing at least one copy of the specific nucleic acid. In a further step, substantially or all primer-coded sequences from the product produced in the reacting step is removed to regenerate a primer binding site. The regeneration of a primer binding site thereby allows a new priming event to occur and the production of more than one copy of said specific nucleic acid.

An additional provision of the present invention is an *in vitro* process for producing more than one copy of a specific nucleic acid in which the products are substantially free of any primer-coded sequences. In this instance, the process comprises the steps of: (a) providing a nucleic acid sample containing or suspected of containing the sequence of the specific nucleic acid; and (b) contacting the sample with a reaction mixture (the mixture comprising (i) unmodified nucleic acid precursors, (ii) one or more specific unmodified primers comprising at least segment each of which primer comprises at least one non-complementary sequence to a distinct sequence of the specific nucleic acid, such that upon hybridization to the specific nucleic acid, at least one loop structure is formed, and (iii) an effective amount of a nucleic acid producing catalyst). The mixture so formed is allowed to react in step (c) under isostatic conditions of temperature, buffer and ionic strength, thereby producing at least one copy of the specific nucleic acid;

IDT Exhibit 1020.9030

which step is followed by (d) removing substantially or all primer-coded sequences from the product produced in step (c) to regenerate a primer binding site. The regeneration of a primer binding site thereby allows a new priming event to occur and the production of more than one copy of said specific nucleic acid.

Another embodiment of the present invention concerns a promoter-independent non-naturally occurring nucleic acid construct which when present in a cell produces a nucleic acid without the use of any gene product coded by the construct.

In yet another embodiment, the present invention provides a conjugate comprising a protein-nucleic acid construct in which the nucleic acid construct does not code for said protein, and which conjugate produces a nucleic acid when present in a cell.

The present invention also has significant *in vivo* applications. In one such application, an *in vivo* process is provided for producing a specific nucleic acid. The *in vivo* process comprises the steps of (a) providing a conjugate comprising a protein-nucleic acid construct, the conjugate being capable of producing a nucleic acid when present in a cell; and (b) introducing such a conjugate into a cell, thereby producing the specific nucleic acid.

Another significant aspect of the present invention relates to a construct comprising a host promoter located on the construct such that the host transcribes a sequence in the construct coding for a different RNA polymerase, which after translation is capable of recognizing its cognate promoter and transcribing from a DNA sequence of interest from the construct with the cognate promoter oriented such that it does not promote transcription from the construct of the different RNA polymerase.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 (A-F) depicts various nucleic acid construct forms contemplated by the invention in which at least one single-stranded region are located therein.

Figure 2 (A-F) depicts the functional forms of the nucleic acid constructs illustrated in Figure 1 (A-F).

Figure 3 (A-C) is an illustration of three nucleic acid constructs with an RNA polymerase covalently attached to a transcribing cassette.

Figure 4 (A-C) illustrates three nucleic acid constructs with promoters for endogenous RNA polymerase.

Figure 5 is a nucleic acid sequence for M13mp18.

Figure 6 shows the sequence and the positions of the primers derived from M13mp18 which were employed in the present invention for nucleic acid production.

Figure 7 illustrates appropriate restriction sites in M13mp18.

Figure 8 is an agarose gel with a lane legend illustrating the experimental results in Example 5 in which amplification of the M13 fragment was carried out in the presence of a large excess (1500 fold) of irrelevant DNA.

Figure 9 is an agarose gel with a lane legend illustrating the results in Example 8 in which the effect of variations of reaction conditions on the product obtained in Example 3 was investigated.

Figure 10 is an agarose gel with a lane legend that illustrates the results of a qualitative analysis of the effects observed in Example 9 of various buffers on the amplification reaction in accordance with the present invention.

Figure 11 is a southern blot (with lane legend) obtained from Example 10 in which two buffers, DMAB and DMG, were separately employed in nucleic acid production.

Figure 12 is an agarose gel and lane legend obtained in Example 11 in which the nature of the ends of amplified product was investigated.

Figure 13 is an agarose gel obtained in Example 12 in which amplification from non-denatured template was examined.

Figure 14 is an agarose gel obtained in Example 13 in which amplification from an RNA template was examined.

Figure 15 is a southern blot of the gel obtained in Figure 14.

Figure 16 is a fluorescence spectrum illustrating the results obtained in Example 14 in which the phenomenon of "strand displacement" using ethidium-labeled oligonucleotides in accordance with the present invention was investigated.

Figure 17 is a fluorescence spectrum illustrating the results obtained in Example 15 in which a T7 promoter oligonucleotide 50 mer labeled with ethidium was employed to study its effects on *in vitro* transcription by T7 and T3 polymerases from an IBI 31 plasmid (pIBI 31-BH5-2) and from a BlueScript II plasmid construct (pBSII//HCV).

Figure 18 depicts the polylinker sequences of the IBI 31 plasmid (pIBI 31-BH5-2) and the BlueScript II plasmid construct (pBSII//HCV).

IDT Exhibit 1020.9031

The present invention describes novel methods and constructs for production of multiple copies of specific nucleic acid sequences *in vitro* and *in vivo*

One aspect of this invention represents an *in vitro* process for the production of more than one copy of nucleic acid from specific target nucleic acid (either DNA or RNA) sequences utilizing a biological catalyst, e.g., a DNA polymerase, primer oligonucleotides complementary to sequences (primer sites) in the target nucleic acid. The production process can proceed in the presence of a large excess of other nucleic acids and does not require thermal cycling or the introduction of specific intermediate constructs such as promoters or assymetric restriction sites, etc.

More particularly, this invention provides an *in vitro* process for producing more than one copy of a specific nucleic acid, the process being independent of a requirement for the introduction of an intermediate structure for the production of any such specific nucleic acid. The *in vitro* production process comprises the steps of: (a) providing a nucleic acid sample containing or suspected of containing the sequence of the specific nucleic acid; (b) contacting the sample with a three component mixture; and (c) allowing the thus-contacted mixture to react under isostatic conditions of temperature, buffer and ionic strength, thereby producing more than one copy of the specific nucleic acid. The three component mixture just alluded will generally comprise (i) nucleic acid precursors, (ii) one or more specific nucleic acid primers each of which is complementary to a distinct sequence of the specific nucleic acid, and (iii) an effective amount of a nucleic acid producing catalyst. In other aspects, the specific nucleic acid may be single-stranded or double-stranded, and may take the form of deoxyribonucleic acid, ribonucleic acid, a DNA.RNA hybrid or a polymer capable of acting as a template for a nucleic acid polymerizing catalyst.

In addition, the specific nucleic acid can be in solution in which case the above-described *in vitro* process may further comprise the step of treating the specific nucleic acid with a blunt-end promoting restriction enzyme. Further, isolation or purification procedures can be employed to enrich the specific nucleic acid. Such procedures are well-known in the art, and may be carried out on the specific nucleic acid prior to the contacting step (b) or the reacting step (c). One means of isolation or purification of a nucleic acid involves its immobilization, for example, by sandwich hybridization (Ranki et al., 1983), or sandwich capture. Particularly significant in the latter methodology is the disclosure of Engelhardt and Rabbani, U.S. Patent Application Serial No. 07/968,706, filed on October 30, 1992, entitled "Capture Sandwich Hybridization Method and Composition," now allowed, that was published as European Patent Application Publication No. 0 159 719 A2 on October 30, 1985. The contents of the foregoing U.S. patent application is incorporated herein by reference.

The target nucleic can be be present in a variety of sources. For purposes of disease diagnosis these would include blood, pus, feces, urine, sputum, synovial fluid, cerebral spinal fluid, cells, tissues, and other sources. The production process can be performed on target nucleic that is present in samples which are free of interfering substances, or the production process can be performed on target nucleic acid separated from the sample. The nucleic acid can be in solution or bound to a solid support. While the replication process can be carried out in the presence of nonrelevant nucleic acids, certain applications may require prior separation of the target sequences. Methods such as sandwich hybridization or sandwich capture referenced above can then be applied to immobilize target sequences. In such instances where sandwich hybridization or sandwich capture is carried out, the above-described *in vitro* process may further comprise the step of releasing the captured nucleic acid, e.g., by means of a restriction enzyme.

As described above, the target sequence need not be limited to a double-stranded DNA molecule. Target molecules could also be single stranded DNA or RNA. For example, replication of a single-stranded target DNA could proceed using primers complementary to both the single-stranded DNA target and to the produced complementary sequence. Following the initial synthesis of the complementary sequence DNA, production from this strand would begin. RNA can serve as the template using a DNA polymerase I, e.g., Klenow, which can reverse transcribe under conditions that have been described (Karkas, J.D. et al., Proc Nat Acad Sci U.S.A. 69:398-402 (1972)).

In case the target nucleic acid is double stranded, a restriction digest or sonication, partial endonuclease treatment or denaturation could be employed for the preparation of the target nucleic acid before the onset of amplification.

An aspect of this invention concerns its use in determining whether a specific target nucleic acid was derived from a living or a deceased organism. To make such a determination, one could in parallel amplify and detect the presence of a specific target DNA or a specific target RNA associated with the genomic makeup of the organism; and thereafter amplify and detect the presence of a specific RNA target associated to the biological function (living function) of the organism which does not survive if the organism is deceased.

IDT Exhibit 1020.9032

The nucleic acid precursors contemplated for use in the present invention are by and large well-known to those skilled in the art. Such precursors may take the form of nucleoside triphosphates and nucleoside triphosphate analogs, or even combinations thereof. More particularly, such nucleoside triphosphates are selected from deoxyadenosine 5'-triphosphate, deoxyguanosine 5'-triphosphate, deoxythymidine 5'-triphosphate, deoxycytidine 5'-triphosphate, adenosine 5'-triphosphate, guanosine 5'-triphosphate, uridine 5'-triphosphate and cytidine 5'-triphosphate, or a combination of any of the foregoing. Such nucleoside triphosphates are widely available commercially, or they may be synthesized by techniques or equipment using commercially available precursors.

In the case where the nucleic acid precursors comprise nucleoside triphosphate analogs, these are also widely available from a number of commercial sources, or they may be manufactured using known techniques. Such nucleoside triphosphate analogs can be in the form of naturally occurring or synthetic analogs, or both.

It should not go unrecognized or even unappreciated that the foregoing nucleoside triphosphate and nucleoside triphosphate analogs can be unmodified or modified, the latter involving modifications to the sugar, phosphate or base moieties. For examples of such modifications, see Ward et al., U.S. Patent No. 4,711,955; Engelhardt et al., U.S. Patent No. 5,241,060; Stavrianopoulos, U.S. Patent No. 4,707,440; and Wetmur, Quartin and Engelhardt, U.S. Patent Application Serial No. 07/499,938, filed on March 26, 1990, the latter having been disclosed in European Patent Application Serial No. 0 450 370 A1, published on October 9, 1991. The contents of the foregoing U.S. patents and patent application are incorporated by their entirety into the present application.

The primers, one or more, described herein bind to specific sequences on the target nucleic acids and initiate the polymerizing reaction. While oligo deoxynucleotide primers may be preferred, polydeoxynucleotide as well as oligo and polyribonucleotide or nucleotide copolymer primers can be used (Kornberg, A. and T. A. Baker, second edition, 1992, W.H. Freeman and Co. New York, Karkas, J.D., PNAS 69:2288-2291 (1972); and Karkas, J.D. et al., Proc. Natl. Acad. Sci. U.S.A. 69:398-402 (1972)). Thus, the specific nucleic acid primers may be selected from deoxyribonucleic acid, ribonucleic acid, a DNA.RNA copolymer, or a polymer capable of hybridizing or forming a base-specific pairing complex and initiating nucleic acid polymerization. Under conditions where the primer is an oligoribonucleotide or copolymer, the primer can be removed from its cognate binding site using specific enzymatic digestion (e.g., RNase H, restriction enzymes and other suitable nucleases) such that another primer can bind and initiate synthesis. This can be used as a system for the multiple initiation of the synthesis of polynucleotide or oligonucleotide product.

Modifications, including chemical modifications, in the composition of the primers would provide for several novel variations of the invention. See, for example, U.S. Patent Nos. 4,711,955; 5,241,060; 4,707,440; and U.S. Patent Application Serial No. 07/499,938, *supra*. For example, substitution of the 3' hydroxyl group of the primer by an isoteric configuration of heteroatoms, e.g., a primary amine or a thiol group, would produce chemically cleavable linkers. In the case of thiol excess of another thiol in the reaction mixture will cleave the phosphorothioate linkers which is formed after the initiation of polymerization, thus allowing the DNA polymerase to reinitiate polymerization with the same primer. Thus, in this variation repeated syntheses can begin from a modified, hybridized primer providing a significant increase in the synthesis of DNA.

In another aspect of the invention, the specific nucleic acid primers are not substantially complementary to one another, having for example, no more than five complementary base-pairs in the sequences therein.

In another variation, the primer could contain some noncomplementary sequences to the target, whereupon hybridization would form at least one loop or bubble which could be used as a substrate for a specific endonuclease such that the primer could be removed from the target by enzymatic digestion thus allowing reinitiation. Furthermore, the primer could contain additional sequences noncomplementary to the target nucleic acid. Thus, the specific nucleic acid primers may comprise at least one non-complementary nucleotide or nucleotide analog base, or at least one sequence thereof. The range of non-complementarity may range in some cases from about 1 to about 200 noncomplementary nucleotide or nucleotide analogs, and in other cases, from about 5 to about 20 nucleotides. Such noncomplementary base sequence or sequences can be linked by other than a phosphodiester bond.

As used herein, the term "nucleic acid producing catalyst" is intended to cover any agent, biological, physical or chemical in nature, that is capable of adding nucleotides (e.g., nucleoside triphosphates, nucleoside triphosphate analogs, etc.) to the hydroxyl group on the terminal end of a specific primer (DNA or RNA) using a pre-existing strand as a template. A number of biological enzymes are known in the art which are useful as polymerizing agents. These include, but are not limited to E. coli DNA polymerase I, Klenow polymerase (a large proteolytic fragment of E. coli DNA polymerase I), bacteriophage T7 RNA

IDT Exhibit 1020.9033

polymerase, and polymerases derived from thermophilic bacteria, such as Thermus aquaticus. The latter polymerase are known for their high temperature insensitivity, and include, for example, the *Taq* DNA polymerase I. A thermostable *Taq* DNA polymerase is disclosed in Gelfand et al., U.S. Patent No. 4,889,818. Preferred as a polymerizing agent in the present invention is the *Taq* DNA Polymerase I. Many if not all of the foregoing examples of polymerizing agents are available commercially from a number of sources, e.g., Boehringer-Mannheim (Indianapolis, IND). Particularly suitable as nucleic acid producing catalysts are DNA polymerase and reverse transcriptase, or both. As used herein, "the effective amount" of the nucleic acid producing catalyst is an art-recognized term reflecting that amount of the catalyst which will allow for polymerization to occur in accordance with the present invention.

Since the rate and extent of hybridization of the primers is dependent upon the standard conditions of hybridization (Wetmur, J.G. and Davidson, N. J., Mol. Biol. 31:349 (1968)), the concentration and nucleotide sequence complexity of the total primers added to the reaction mixture will directly affect the rate at which they hybridize and accordingly the extent to which they will initiate nucleic acid synthesis. In addition, if the reaction is run under conditions where the guanosine triphosphate is replaced by inosine triphosphate or other modified nucleoside triphosphates such that the presence of this modified nucleotide in the product nucleic acid would lower the melting temperature of the product:template double helix, then at any given temperature of the reaction the extent of breathing of the double helix will be increased and the extent of binding of the primers to the target strand will be enhanced.

Furthermore, primers could displace the strands at the ends of the double stranded target and hybridize with one of the two strands and, this displacement hybridization reaction (or D loop formation reaction) is favored by adding more than one primer molecule. In general, as the total amount of the sequence complexity of the primers complementary to the target nucleic acid is increased a greater nucleic acid production is obtained (see Example 3 below).

Modification of the primers could either increase or decrease the binding of primer to the target at a given pH, temperature and ionic strength, in other words, at isostatic conditions of pH, temperature and ionic strength, e.g., ionic salt. Other primer modifications can be employed which would facilitate polymerization from the primer sites, even when the initiation site is within a double helix. For example, once an oligo primer is introduced into a target double stranded nucleic acid molecule, if such an oligo primer is modified with ethidium or any moiety that increases the melting temperature of the double stranded structure formed by the oligo and a target nucleic acid, it forms a relatively more stable single stranded structure because of the nucleotide modifications. This produces a primer initiation site. In fact, the nucleic acid precursors or the specific primers (or both) can be modified by at least one intercalating agent, such as ethidium, in which case it may be useful to carry out an additional step (d) of detecting any product produced in step (c), as set forth above. In such a step where desirable, detection can be carried out by means of incorporating into the product a labeled primer, a labeled precursor, or a combination thereof.

Another additional aspect of the *in vitro* process, above-described, is the inclusion of a further step of regenerating one or more specific nucleic acid primers, as described elsewhere in this disclosure, including immediately below.

As described in the summary of this invention, an *in vitro* process for multiple nucleic acid production is provided in which the products are substantially free of any primer-coded sequences. In such process, the removing step (d) is carried out by digestion with an enzyme, e.g., ribonuclease H. In one aspect of this invention, the nucleic acid precursors are modified or unmodified in the instance where one or more specific polynucleotide primers are used, the primers comprising at least one ribonucleic acid segment and wherein each primer is substantially complementary to a distinct sequence of the specific nucleic acid. Thus, the specific polynucleotide primers may further comprise deoxyribonucleic acid. In another feature of this particular *in vitro* process, the specific polynucleotide primers contain a 3'-hydroxyl group or an isoteric configuration of heteroatoms, e.g., nitrogen, sulfur, or both. In addition, the polynucleotide primers in this instance may further comprise from about 1 to about 200 noncomplementary nucleotide or nucleotide analogs.

In yet a further *in vitro* process for producing more than one copy of a specific nucleic acid is provided (as described in the summary), the products being substantially free of any primer-coded sequences. In this instance, unmodified nucleic acid precursors are reacted in a mixture with one or more chemically-modified primers each of which is substantially complementary to a distinct sequence of the specific nucleic acid. An effective amount of a nucleic acid producing catalyst is also provided in the mixture. As in the case of the last-described *in vitro* process, the removing step (d) may be carried out by digestion with an enzyme, e.g., ribonuclease H. The specific chemically modified primers are selected, for example, from ribonucleic acid, deoxyribonucleic acid, a DNA.RNA copolymer, and a polymer capable of hybridizing or forming a base-specific pairing complex and initiating nucleic acid polymerization, or a combination of any of the foregoing.

IDT Exhibit 1020.9034

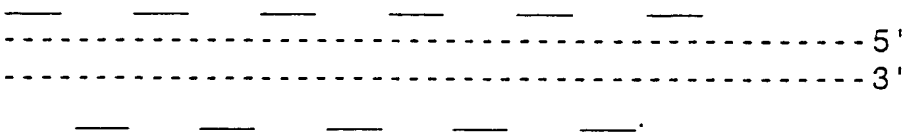
The specific chemically modified primers may contain a 3'-hydroxyl group or an isosteric configuration of heteroatoms, N, S, or both, as described above in other *in vitro* processes. Further, the specific chemically modified primers can be selected from nucleoside triphosphates and nucleoside triphosphate analogs, or a combination thereof, wherein at least one of said nucleoside triphosphates or analogs is modified on the sugar, phosphate or base. Also as in other *in vitro* processes, the specific chemically modified primers may further comprise from about 1 to about 200 noncomplementary nucleotide or nucleotide analogs.

In still yet another of the *in vitro* processes for multiple nucleic acid production, described previously in the summary of this invention, unmodified nucleic acid precursors are provided in the mixture and reacting step (c), together with one or more specific unmodified primers comprising at least one segment, each of which primer comprises at least one non-complementary sequence to a distinct sequence of the specific nucleic acid, such that upon hybridization to the specific nucleic acid at least one loop structure is formed. As in the other instances, digestion with an enzyme, e.g., ribonuclease H, may be employed in the removing step (d). In one feature of this process, specific unmodified primers are selected from ribonucleic acid, deoxyribonucleic acid, a DNA.RNA copolymer, and a polymer capable of hybridizing or forming a base-specific pairing complex and initiating nucleic acid polymerization, or a combination of any of the foregoing. Further, the specific unmodified primers may further comprise from about 1 to about 200 noncomplementary nucleotide or nucleotide analogs, in accordance with the present invention.

The rate of hybridization of the primer to target nucleic acids and, in particular, to target double stranded nucleic acids can be facilitated by binding of the primer with various proteins, e.g., rec A proteins. For example, if the primer is modified with an intercalating agent, e.g., ethidium (or any moiety that increases the melting temperature of the double stranded structure), the addition of this primer to or with a protein such as rec A, either free or bound, would facilitate the introduction of the primer into the double stranded target. (Komberg and Baker, *supra*, pages 797-800). This could produce a suitable primer initiation site.

The arrangement of primer binding sites on the template nucleic acid can be varied as desired. For example, the distance between successive primer binding sites on one strand can also be varied as desired. Also specific primers can be employed that initiate synthesis upstream of the sequence sought to be copied. Under this scenario, multiple copies of nucleic acid are made without successive denaturation or use of other enzymes or the introduction of intermediate structures for their production.

When primer sites on double stranded DNA are arranged as shown, specific DNA production is increased.



When the target sequences are substantially covered by their complementary primers, a further increase in the production of multiple copies of nucleic acid is favored due to the increase in initiation points and destabilization of the double stranded template molecule.

Finally, if an oligo is modified such that it will form a stable hybrid, even in the presence of the complementary nucleic acid strand, then the modified oligo can act as a 'helper' oligo. 'Helper' oligo in this context is defined as a oligo that does not necessarily act as a primer but will accelerate the binding and priming activity of other oligos in the vicinity to the binding site of the 'helper' oligo. Vicinity is here being defined as the location of a nucleotide sequence or the complementary nucleotide sequence close enough to the binding site of the 'helper' oligo to have the rate or extent of hybridization of the primer affected by the binding of the 'helper' oligo. The 'helper' oligo can be modified such that it does not initiate polymerization as for example through the use of a dideoxy 3' terminal nucleotide or other nucleotide with blocked 3' ends. The 'helper' oligo can also be modified in such a manner that the double helix formed by the 'helper' oligo and the target nucleic acid strand or the 'helper' and the complementary strand to the target strand is more stable or has a higher melting temperature than the equivalent double helix of unmodified 'helper' oligo and the target or the strand complementary to the target strand. Such modifications can include halogenation of certain bases, ethenyl pyrimidines (C:C triple bonds, propyne amine derivatives; the addition of ethidium or other intercalating molecules (see Stavrianopoulos and Rabbani, U.S. Patent Application Serial No. 07/956,566, filed on October 5, 1992, the contents of which are incorporated herein by reference and which were disclosed in European Patent Application Publication No. 0 231 495 A2,

IDT Exhibit 1020.9035

published on August 12, 1987); the supplementation of the oligo with certain proteins that stabilize the double helix and any other treatment or procedure or the addition of any other adduct that serves to stabilize the portion of the double helix with the 'helper' bound or to increase the melting temperature of portion of the double helix with the 'helper' bound.

In vivo Synthesis of Nucleic Acid

This invention describes a casette or nucleic acid construct into which any nucleic acid sequence can be inserted and which can be used as a template for the production of more than one copy of the specific sequence. This cassette is a nucleic acid construct containing a sequence of interest, which within or present within, the cell produces nucleic acid product which is independent or only partially dependent on the host system. The cassette or nucleic acid construct may be characterized as a promoter-independent non-naturally occurring, and in one embodiment comprises double-stranded and single-stranded nucleic acid regions. This construct contains a region in which a portion of the opposite strands are not substantially complementary, e.g., a bubble (even comprising at least one polyT sequence), or loop, or the construct comprises at least one single-stranded region. The construct is composed of naturally occurring nucleotides or chemically modified nucleotides or a synthetic polymer in part or a combination thereof. These structures are designed to provide binding of polymerizing enzymes or primers and the modifications provide for nuclease resistance or facilitate uptake by the target cell.

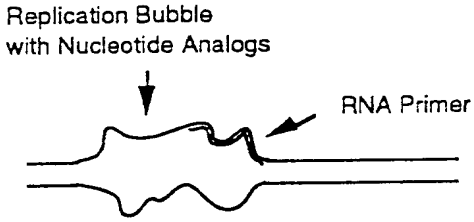
Referring to the constructs (A-F) depicted in Figure 3, the single stranded regions described in the constructs will contain coding sequences for nucleic acid primers present in the cell to facilitate initiation points of DNA polymerase in said cell. In the case of RNA polymerase, these constructs constitute promotor independent binding and initiation of RNA polymerase reaction. These constructs can be used *in vitro* and *in vivo* for production of nucleic acids. The position of the single stranded region adjacent to the double stranded specific sequence would provide a specific and consistent transcription of these specific sequences, both *in vitro* and *in vivo* independent of promotor. The replication (DNA) or transcription (RNA) products of these constructs can be single stranded nucleic acid which could have a sense or antisense function or could be double stranded nucleic acid.

In Figure 13(A), a large bubble is located in the construct. In Figure 13(B), the two strands are noncomplementary at their ends, and thus do not form a bubble. In Figure 13(C), a double bubble is formed due to noncomplementarity at both ends. In Figure 13(D), a single-stranded region is shown in the middle of the construct leading to a partially single-stranded region (and no bubble formation). Figure 13(E) depicts a bubble at one end of the construct (compare with the two bubbles in the construct shown in Figure 13(C)). In Figure 13(F), a single bubble in the middle of the construct is shown. It should be readily appreciated by those skilled in this art that the above-depicted embodiments are representative embodiments not intended to be limiting, particularly in light of the present disclosure.

In vivo these constructs, with a specific primer present in the cell can initiate nucleic acid synthesis. When these primers are RNA, after initiation of nucleic acid synthesis, they can be removed by the action of ribonuclease H, thus vacating the primer binding sequence and allowing other primer molecules to bind and reinitiate synthesis. The cellular nucleic acid synthesizing enzymes can use these constructs to produce copies of a specific nucleic acid from the construct. Shown in Figure 14 (A-F) are corresponding illustrations of the constructs in Figure 13 (A-F), except that the production arrows (points and directions) are indicated.

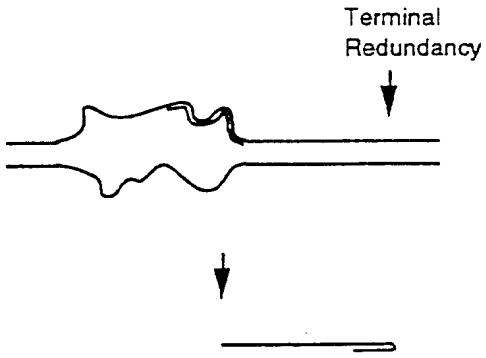
These constructs could contain more than one specific nucleic acid sequence which in turn could produce more than one copy of each specific nucleic acid sequence. If two independent nucleic acid products are complementary, then they could hybridize and form muliple copies of a new double stranded construct that could have the properties of the novel construct. Furthermore they could contain promotor sites such as the host promotor therefore serving as an independent nucleic acid production source (the progeny).

IDT Exhibit 1020.9036



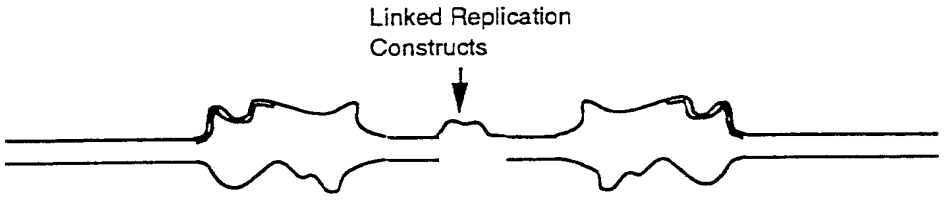
Primer-Dependent DNA Production Using Nucleic Acid Construct

The replication of this structure could result in the production of one strand of DNA product. Several alternative events may occur allowing for the formation of a second complementary strand. For example, a terminal loop could be inserted at the end of the construct such that the single stranded product will code for the synthesis of the complementary strand using the repair enzyme. Constructs can be made that produce single stranded DNA product that has a hairpin loop and therefore, can be used to form a double-stranded product. Alternatively, constructs can be formed that produce nucleic acid in both polarity.



Hairpin Product

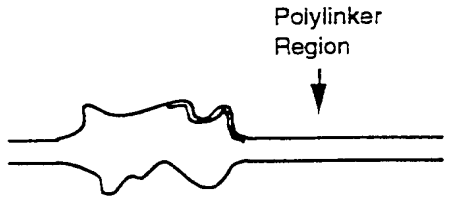
An alternative approach to the production of double stranded product is to covalently link two constructs that make complementary DNA strands.



Linked Complementary Production Constructs

The construct can be made to contain a poly linker region into which any sequence can be cloned. The result will be a transient accumulation of expressing genes within the cell to deliver sense, antisense or protein or any other gene product into the target cell.

IDT Exhibit 1020.9037



Cloning Site in Production Constructs

Other processes within the invention herein described apply to the production of more than one copy of functional genes or antisense DNA or RNA in target cells.

Production of Primers

Primers can be produced by several methods. Single-stranded oligonucleotides in the range from between from about 5 to about 100 bases long, and preferably between from about 10 to about 40, and more preferably, between from about 8 to about 20 nucleotides. These ranges may further vary with optimally between from about 13 to about 30 for bacterial nucleic acid, and optimally between from about 17 to about 35 for eukaryote nucleotides would appear to be appropriate for most applications although it may be desirable in some or numerous instances to vary the length of the primers. Oligonucleotide primers can be most conveniently produced by automated chemical methods. In this way modified bases can be introduced. Manual methods can be used and may in some cases be used in combination with automated methods. All of these methods and automation are known and available in the art.

In addition nucleic acid primers can be produced readily by the action of T7 RNA polymerase, T3 polymerase, SP6 polymerase or any appropriate DNA or RNA polymerase on DNA templates or RNA templates containing the primer sequences extended from the corresponding RNA polymerase promoter sites or other nucleic acid synthesis start signals.

Detection of Products

DNA produced by the invention described herein can be detected by a variety of hybridization methods using homogeneous or non-homogeneous assays. DNA produced in tissues or cells, i.e., *in situ*, can be detected by any of the practiced methods for *in situ* hybridization. These include, but are not limited to, hybridization of the produced DNA with a nucleic acid probe labeled with a suitable chemical moiety, such as biotin. Probes used for the detection of produced DNA can be labeled with a variety of chemical moieties other than biotin. These include but are not limited to fluorescein, dinitrophenol, ethidium (see, for example, the disclosures of U.S. Patent Nos. 4,711,955; 5,241,060; and 4,707,440, *supra*).

The hybridized, labeled nucleic acid probe can be detected by a variety of means. These include but are not limited to reaction with complexes composed of biotin binding proteins, such as avidin or streptavidin, and color generating enzymes, such as horseradish peroxidase or alkaline phosphatase, which, in the presence of appropriate substrates and chromogens, yield colored products.

In accordance with this invention, DNA production from target sequences generally requires nucleic acid precursors, e.g., adenosine triphosphate, guanosine triphosphate, thymidine triphosphate and cytosine triphosphate, present in sufficient quantity and concentration in the reaction mixture. In other applications it may be advantageous to substitute one or more of the natural precursors with modified nucleotides. For example, when the invention described herein is being applied to the detection of specific nucleic acid sequences, immobilization of the produced DNA may be desirable. In such an instance, substitution of one or more of the natural nucleotide triphosphate precursors with a modified nucleotide, e.g., biotinylated deoxyuridine triphosphate, in place of thymidine triphosphate, would yield biotin-labeled DNA. The produced DNA could be separated by its affinity for a biotin binding protein, such as avidin, streptavidin or an anti-biotin antibody. A variety of nucleoside triphosphate precursors (U.S. Patent Nos. 4,711,955; 5,241,060; and 4,707,440, *supra*) labeled with chemical moieties which include, but are not limited to, dinitrophenol and fluorescein, and which can be bound by corresponding antibodies or by other binding proteins can be used in this manner. In other aspects of the invention, the produced DNA can be isotopically labeled by the inclusion of isotopically labeled deoxynucleotide precursors in the reaction mixture.

IDT Exhibit 1020.9038

Labeled DNA, produced by the invention described herein, can function as probe nucleic acid to be used to detect specific target nucleic acid sequences.

In certain detection formats the primers may be removed from the reaction mixture by capturing the product through direct capture (Brakel et al., U.S. Patent Application Serial No. 07/998,660, filed on December 23, 1992, the contents of which have been disclosed in European Patent Application 0 435 150 A2, published on July 3, 1991; and the contents of which are also incorporated by reference herein), or sandwich capture. (Engelhardt and Rabbani, allowed U.S. Patent Application Serial No. 07/968,706, *supra*), or by modifying the primers at the 3' end with biotin or imminobiotin without an arm or a very short arm such that the avidin will recognize only the unincorporated primers (single stranded form) but not the incorporated due to the double stranded form and the short length of the arm. Additionally, the primer may be labeled with ethidium or other intercalating moiety. In this condition, the ethidium or other intercalating moiety may be inactivated (Stavrianopoulos, U.S. Patent Application Serial No. 07/633,730, filed on December 24, 1990, published as European Patent Application Publication No. 0 492 570 A1 on July 1, 1992; the contents of which are incorporated by reference) in the unhybridized oligo and not in the hybridized oligo:target.

Another aspect of this invention herein described is to provide for a conjugate of a nucleic acid polymerizing enzyme (RNA polymerase) with a nucleic acid construct said nucleic acid construct contains an initiation site such as a promotor site for the corresponding RNA polymerase which is capable of producing nucleic acid both *in vivo* and *in vitro*. The enzyme could be linked directly to the nucleic acid or through a linkage group substantially not interfering with its function or the enzyme could be linked to the nucleic acid indirectly by a nucleic acid bridge or haptene receptor where the enzyme is biotinylated and the nucleic acid construct contains an avidin or *vice versa* or when the nucleic acid construct contains sequences for binding proteins such as a repressor and an enzyme linked to said nucleic acid binding protein (U.S. Patent No. 5,241,060, *supra*, and Pergolizzi, Stavrianopoulos, Rabbani, Engelhardt, Kline and Olsiewski, U.S. Patent Application Serial No. 08/032,769, filed on March 16, 1993, published as European Patent Application Publication No. 0 128 332 A1 on December 19, 1984, the latter having been "allowed" by the European Patent Office, and further incorporated by reference herein).

Further in regard to the just-described conjugate of the present invention, the protein in one aspect comprises an RNA polymerase or a subunit thereof and the nucleic acid construct contains the corresponding RNA polymerase promoter. The RNA polymerase can be selected from T7, T3 and SP6, or a combination of any of the foregoing. In another embodiment, the protein in the conjugate comprises DNA polymerase or reverse transcriptase and the nucleic acid construct contains at least one sequence complementary to an RNA molecule. The construct can take the form of double-stranded, single-stranded, or even partially single-stranded. Further, the nucleic acid construct in the conjugate may comprise at least one chemically modified nucleotide or nucleotide analog. The linkages of the protein to the construct are described in the preceding paragraph. The nucleic acid produced by or from this conjugate comprises deoxyribonucleic acid, ribonucleic acid, or combinations thereof, or it may be antisense or sense, or both.

As described in the summary of the invention, the above-described conjugate may be utilized in an *in vivo* process for producing a specific nucleic acid. In other aspects of this *in vivo* process, the construct is further characterized as comprising (independently) at least one promoter, at least one complementary sequence to a primer present in the cell, or codes for the protein in the conjugate, or for a protein other than the protein in the conjugate. The other protein may comprise a nucleic acid polymerase. In the instant where the polymerase comprises an RNA polymerase, the nucleic acid construct may comprise a promoter for the RNA polymerase. Further, the polymerase can be a DNA polymerase or reverse transcriptase.

(a) Direct Attachment of a Polymerase to the Construct

For example, if a construct containing a RNA polymerase linked directly or indirectly to a DNA construct or cassette is introduced into a cell, the RNA polymerase will transcribe the nucleic acid in the construct and is completely independent of any host RNA polymerases. Each molecule introduced into a cell will produce multiple copies of a segment of the construct. In the first iteration, the attached polymerase can produce the RNA for the target sequence itself. (See Figure 3 (A)). Alternatively, the promotor, specific for the attached polymerase, may be linked to two separate sequences, namely the polymerase gene and the target gene. See Figure 3 (B). In this instance, the amount of polymerase initiating at the promotor site will increase as the polymerase gene is transcribed and translated. Finally, the coding sequence transcribed by the T_7 promotor (or any specific first promotor) may produce any RNA polymerase (including T_7 polymerase or polymerase III or others), and this polymerase may transcribe off of another or second promotor (or off of a different strength T_7 or other first promotor) to produce the transcript of the target

IDT Exhibit 1020.9039

sequence. (See Figure 3 (C)).

Referring to the constructs or cassettes shown in Figure 4 (A-C), these can be derived by using standard recombinant DNA techniques. The appropriate piece of DNA can be isolated and covalently attached to the RNA polymerase under conditions whereby the RNA polymerase functions after being covalently attached to a solid matrix (Cook, P.R and Grove, F. Nuc. Acids Res, 20:3591-3598 (1992)). Methods of modifying the ends of DNA molecules for attachment of chemical moieties are well known (see, for example, U.S. Patent Application Serial No. 08/032,769, *supra*). The transcribed product can act per se as sense or antisense RNA or it can be translated into protein. The enzyme and/or nucleic acid constructs could be modified to facilitate transport and/or achieve resistance to degrading enzymes (U.S. Patent No. 5,241,060, *supra*).

(b) *In vivo* Amplification of Transcription

Constructs can be made that are dependent upon transcription or replication using a host polymerase. When such a construct contains a double promotor, the second promotor can be different than the first promotor or it can be a stronger or weaker version of the first promotor. Vectors can be chosen such that the constructs can either integrate into the chromosome, replicate autosomally or be replication-deficient and function only for transient expression. They can function in the nucleus or the cytoplasm if the target cell is eukaryotic. The figure below depicts various constructs or cassettes and is not limiting as to the possible variations contemplated by the present invention.

Referring to Figure 4 (A), the nucleic acid construct or cassette depicted in this figure contains a promotor that codes for the production of a polymerase that is not endogenous to the target cell. For example, an SV40 or RNA polymerase III promotor that codes for a T_7 RNA polymerase. Transcription and translation of these transcripts by cellular machinery results in the production of active T_7 RNA polymerase which will utilize the T_7 promotor to transcribe the target sequence (Fuerst, T.R. et al., Proc Nat Acad Sci USA 83:8122 (1986)) have shown high levels of transient expression using a dual construct system with the T_7 RNA polymerase gene on one construct and the target gene behind the T_7 promotor on the other construct. The simplest iteration of this construct is that the genes coding for the polymerase code for a polymerase that exists within the cell and therefore is not recognized by the host organism as a foreign protein and does not induce an immune response.

In Figure 4 (B), an additional autocatalytic cycle has been added whereby the extent of transcription of the target gene is enhanced by the production of T_7 RNA polymerase throughout the transient expression cycle.

In Figure 4 (C), the construct or cassette is similar to Figure 4 (B) with the additional element that there is a down regulation of the autocatalytic cascade by competition by a high efficiency promotor with a low efficiency transcriptional promotor.

Three Constructs with Promotors for Endogenous RNA Polymerase

As described in the summary, the present invention further provides a construct comprising a host promoter located on the construct such that the host transcribes a sequence in the construct coding for a different RNA polymerase which after translation is capable of recognizing its cognate promoter and transcribing from a DNA sequence of interest in the construct with the cognate promoter oriented such that it does not promote transcription from the construct of the different RNA polymerase. In one feature of this construct, the host promoter comprises a prokaryotic promoter, e.g., RNA polymerase, or eukaryotic promoter, e.g., Pol I, Pol II, Pol III, or combinations thereof, such prokaryotic or eukaryotic promoter being located upstream from the host promoter. The second RNA polymerase may be selected from various members, including T7, T3 and SP6, or combinations thereof. The DNA sequence of interest may comprise sense or antisense, or both, or it may comprise DNA or RNA, or still yet, it may encode a protein. The construct may further comprise at least one chemically modified nucleotide.

Additionally, promotors that will be read by polymerases within the target cell can be linked to the production of additional polymerase specific for that promotor or other promotors. The polymerases can be for example, T7 polymerase, RNA polymerase III, or any other polymerase. A second promotor keyed sequence can be in the construct such that a second polynucleotide can be synthesized from the construct. It can code for the production of antisense or sense RNA or DNA molecules. These constructs or cassettes can be created using standard recombinant DNA techniques.

The property and structure of all nucleic acid constructs provided in accordance with this invention is applicable to each other in combination, *in toto* or in part. That is to say, in the conjugate comprising a

IDT Exhibit 1020.9040

protein and a nucleic acid construct, the construct could include, for example, chemical modification and bubble structure or single-stranded regions for primer binding sites or RNA initiation sites. Other variations would be recognized by those skilled in the art in light of the detailed description of this invention.

One subject matter of the present invention is process A which is an *in vitro* process for producing more than one copy of a specific nucleic acid, said process being independent of a requirement for the introduction of an intermediate structure for the production of said specific nucleic acid, said process comprising the steps of:

(a) providing a nucleic acid sample containing or suspected of containing the sequence of said specific nucleic acid;

(b) contacting said sample with a mixture comprising:

(i) nucleic acid precursors,

(ii) one or more specific nucleic acid primers each of which is complementary to a distinct sequence of said specific nucleic acid, and

(iii) an effective amount of a nucleic acid producing catalyst; and

(c) allowing said mixture to react under isostatic conditions of temperature, buffer and ionic strength, thereby producing more than one copy of said specific nucleic acid.

Specific embodiments of this process A are the following items:

1. The process A wherein said specific nucleic acid is single-stranded or double-stranded.

2. The process A wherein said specific nucleic acid is selected from deoxyribonucleic acid, ribonucleic acid, a DNA.RNA hybrid or a polymer capable of acting as a template for a nucleic acid polymerizing catalyst.

3. The process A wherein said specific nucleic acid is in solution.

4. The process of item 3 further comprising the step of treating said specific nucleic acid with a blunt-end promoting restriction enzyme.

5. The process A wherein said specific nucleic acid is isolated or purified prior to the contacting step (b) or the reacting step (c).

6. The process of item 5 wherein said isolation or purification of said specific nucleic acid is carried out by means of sandwich or sandwich capture.

7. The process of item 6 further comprising the step of releasing said captured specific nucleic acid.

8. The process of item 7 wherein said releasing step is carried out by means of a restriction enzyme.

9. The process A wherein said nucleic acid precursors are selected from nucleoside triphosphates and nucleoside trisphosphate analogs, or a combination thereof.

10. The process of item 9 wherein said nucleoside triphosphates are selected from deoxyadenosine 5'-triphosphate, deoxyguanosine 5'-triphosphate, deoxythymidine 5'-triphosphate, deoxycytidine 5'-triphosphate, adenosine 5'-triphosphate, guanosine 5'-triphosphate, uridine 5'-triphosphate and cytidine 5'-triphosphate, or a combination of any of the foregoing.

11. The process of item 9 wherein said nucleoside triphosphate analogs are naturally occurring or synthetic, or a combination thereof.

12. The process of item 9 wherein at least one of said nucleoside triphosphates or nucleoside triphosphate analogs is modified on the sugar, phosphate or base.

13. The process A wherein said specific nucleic acid primers are selected from deoxyribonucleic acid, ribonucleic acid, a DNA.RNA copolymer, or a polymer capable of hybridizing or forming a base-specific pairing complex and initiating nucleic acid polymerization.

14. The process A wherein said specific nucleic acid primers comprise oligo- or polynucleotides.

15. The process A wherein said specific nucleic acid primers contain a 3'-hydroxyl group or an isosteric configuration of heteroatoms.

16. The process of item 15 wherein said heteroatoms are selected from nitrogen, sulfur, or both.

17. The process A wherein said specific nucleic acid primers are not substantially complementary to one another.

18. The process of item 17 wherein said specific nucleic acid primers contain no more than five complementary base-pairs in the sequences therein.

19. The process A wherein said specific nucleic acid primers comprise from about 5 to about 100 nucleotides.

20. The process of item 19 wherein said specific nucleic acid primers comprise from about 8 to about 20 nucleotides.

21. The process A wherein said specific nucleic acid primers comprise at least one non-complementary nucleotide or nucleotide analog base, or at least one sequence thereof.

IDT Exhibit 1020.9041

22. The process of item 21 wherein said specific nucleic acid primers further comprise from about 1 to about 200 noncomplementary nucleotide or nucleotide analogs.

23. The process of item 22 wherein said noncomplementary nucleotide or nucleotide analogs in said specific nucleic acid primers comprise from about 5 to about 20 nucleotides.

24. The process of item 21 wherein said noncomplementary base sequence or sequences are linked together by other than a phosphodiester bond.

25. The process A wherein said nucleic acid producing catalyst is selected from DNA polymerase and reverse transcriptase, or both.

26. The process A wherein said nucleic precursors or said specific primers or both are modified by at least one intercalating agent.

27. The process A further comprising the step (d) of detecting the product produced in step (c).

28. The process of item 27 wherein said detecting step (d) is carried out by means of incorporating into the product a labeled primer, a labeled precursor, or a combination thereof.

29. The process A further comprising the step of regenerating said one or more specific nucleic acid primers.

Another subject matter of the invention is process B which is an *in vitro* process for producing more than one copy of a specific nucleic acid, said products being substantially free of any primer-coded sequences, said process comprising the steps of:

(a) providing a nucleic acid sample containing or suspected of containing the sequence of said specific nucleic acid;

(b) contacting said sample with a mixture comprising:

(i) nucleic acid precursors,

(ii) one or more specific polynucleotide primers comprising at least one ribonucleic acid segment each of which primer is substantially complementary to a distinct sequence of said specific nucleic acid, and

(iii) an effective amount of a nucleic acid producing catalyst; and

(c) allowing said mixture to react under isostatic conditions of temperature, buffer and ionic strength, thereby producing at least one copy of said specific nucleic acid; and

(d) removing substantially or all primer-coded sequences from the product produced in step (c) to regenerate a primer binding site, thereby allowing a new priming event to occur and producing more than one copy of said specific nucleic acid.

Specific embodiments of this process B are the following items:

1. The process B wherein said step (d) removing is carried by digestion with an enzyme.

2. The process of item 1 wherein said enzyme comprises ribonuclease H.

3. The process B wherein said nucleic acid precursors are modified or unmodified.

4. The process B wherein said specific polynucleotide primers further comprise deoxyribonucleic acid.

5. The process B wherein said specific polynucleotide primers contain a 3'-hydroxyl group or an isosteric configuration of heteroatoms.

6. The process of item 5 wherein said heteroatoms are selected from nitrogen, sulfur, or both.

7. The process B wherein said specific polynucleotide primers further comprise from about 1 to about 200 noncomplementary nucleotide or nucleotide analogs.

A further subject matter of the present invention is process C which is an *in vitro* process for producing more than one copy of a specific nucleic acid, said products being substantially free of any primer-coded sequences, said process comprising the steps of:

(a) providing a nucleic acid sample containing or suspected of containing the sequence of said specific nucleic acid;

(b) contacting said sample with a mixture comprising:

(i) unmodified nucleic acid precursors,

(ii) one or more specific chemically-modified primers each of which primer is substantially complementary to a distinct sequence of said specific nucleic acid, and

(iii) an effective amount of a nucleic acid producing catalyst; and

(c) allowing said mixture to react under isostatic conditions of temperature, buffer and ionic strength, thereby producing at least one copy of said specific nucleic acid; and

(d) removing substantially or all primer-coded sequences from the product produced in step (c) to regenerate a primer binding site, thereby allowing a new priming event to occur and producing more than one copy of said specific nucleic acid.

Specific embodiments of process C are the following items:

1. The process C wherein said step (d) removing is carried by digestion with an enzyme.

IDT Exhibit 1020.9042

2. The process of item 1 wherein said enzyme comprises ribonuclease H.

3. The process C wherein said specific chemically modified primers are selected from ribonucleic acid, deoxyribonucleic acid, a DNA.RNA copolymer, and a polymer capable of hybridizing or forming a base-specific pairing complex and initiating nucleic acid polymerization, or a combination of any of the foregoing.

4. The process C wherein said specific chemically modified primers contain a 3'-hydroxyl group or an isosteric configuration of heteroatoms.

5. The process of item 4 wherein said heteroatoms are selected from nitrogen, sulfur, or both.

6. The process C wherein said specific chemically modified primers are selected from nucleoside triphosphates and nucleoside triphosphate analogs, or a combination thereof, wherein at least one of said nucleoside triphosphates or analogs is modified on the sugar, phosphate or base.

7. The process C wherein said specific chemically modified primers further comprise from about 1 to about 200 noncomplementary nucleotide or nucleotide analogs.

Another embodiment of the invention is process D which is an *in vitro* process for producing more than one copy of a specific nucleic acid, said products being substantially free of any primer-coded sequences, said process comprising the steps of:

(a) providing a nucleic acid sample containing or suspected of containing the sequence of said specific nucleic acid;

(b) contacting said sample with a mixture comprising:

(i) unmodified nucleic acid precursors,

(ii) one or more specific unmodified primers comprising at least onesegment each of which primer comprises at least one non-complementary sequence to a distinct sequence of said specific nucleic acid, such that upon hybridization to said specific nucleic acid at least one loop structure is formed, and

(iii) an effective amount of a nucleic acid producing catalyst; and

(c) allowing said mixture to react under isostatic conditions of temperature, buffer and ionic strength, thereby producing at least one copy of said specific nucleic acid; and

(d) removing substantially or all primer-coded sequences from the product produced in step (c) to regenerate a primer binding site, thereby allowing a new priming event to occur and producing more than one copy of said specific nucleic acid.

Specific embodiments of process D are the following items:

1. The process D wherein said step (d) removing is carried by digestion with an enzyme.

2. The process of item 1 wherein said enzyme comprises ribonuclease H.

3. The process D wherein said specific unmodified primers are selected from ribonucleic acid, deoxyribonucleic acid, a DNA.RNA copolymer, and a polymer capable of hybridizing or forming a base-specific pairing complex and initiating nucleic acid polymerization, or a combination of any of the foregoing.

4. The process D wherein said specific unmodified primers further comprise from about 1 to about 200 noncomplementary nucleotide or nucleotide analogs.

Another embodiment of the invention is construct A which is a promoter-independent non-naturally occurring nucleic acid construct which when present in a cell produces a nucleic acid without the use of any gene product coded by said construct.

Specific embodiments of the construct A of the invention are the following items:

1. The construct A comprising double-stranded and single-stranded nucleic acid regions.

2. The construct A wherein said nucleic acid comprises deoxyribonucleic acid, ribonucleic acid, a DNA.RNA copolymer, or a polymer capable of hybridizing or forming a base-specific pairing complex and initiating nucleic acid polymerization.

3. The construct A comprising at least one modified nucleotide or nucleotide analog.

4. The construct A comprising at least one single-stranded region.

5. The construct of item 4 wherein said single-stranded region comprises a bubble.

6. The construct of item 5 wherein said bubble comprises at least one complementary sequence to a nucleic acid present in the cell.

7. The construct of item 5 wherein said bubble comprises at least one polyT sequence.

Another embodiment of the invention is a conjugate which is a conjugate comprising a protein-nucleic acid construct, said nucleic acid construct not coding for said protein, and which conjugate produces a nucleic acid when present in a cell.

Specific embodiments of the conjugate of the invention are the following items:

IDT Exhibit 1020.9043

1. The conjugate wherein said protein comprises an RNA polymerase or a subunit thereof and the nucleic acid construct contains the corresponding RNA polymerase promoter.

2. The conjugate of item 1 wherein said RNA polymerase is selected from T7, T3 and SP6, or a combination of any of the foregoing.

3. The conjugate wherein said protein comprises DNA polymerase or reverse transcriptase and said nucleic acid construct contains at least one sequence complementary to an RNA molecule.

4. The conjugate wherein said nucleic acid construct is double-stranded, single-stranded, or partially single-stranded.

5. The conjugate wherein said nucleic acid construct comprises at least one chemically modified nucleotide or nucleotide analog.

6. The conjugate wherein said protein is linked to said nucleic acid construct by means of a covalent linkage.

7. The conjugate wherein said protein is linked to said nucleic acid construct by means of base-pairing of complementary nucleic acid sequences.

8. The conjugate wherein said protein is linked to said nucleic acid construct by means of a nucleic acid binding protein.

9. The conjugate of item 8 wherein said nucleic acid binding protein comprises a repressor protein bound to an enzyme.

10. The conjugate wherein said protein is linked to said nucleic acid construct by means of ligand receptor binding.

11. The conjugate wherein the nucleic acid produced is deoxyribonucleic acid, ribonucleic acid, or a combination thereof.

12. The conjugate wherein the nucleic acid produced is sense or antisense, or both.

A further emobidment of the invention is process E which is an *in vivo* process for producing a specific nucleic acid, said process comprising the steps of:

(a) providing a conjugate comprising a protein-nucleic acid construct, said conjugate being capable of producing a nucleic acid when present in a cell; and

(b) introducing said conjugate into a cell, thereby producing said specific nucleic acid.

Specific embodiments of process E are the following items:

1. The process E wherein said construct comprises at least one promoter.

2. The process E wherein said construct comprises at least one complementary sequence to a primer present in the cell.

3. The process E wherein said nucleic acid construct codes for the protein in said conjugate.

4. The process E wherein said nucleic acid construct codes for a protein other than the protein in said conjugate.

5. The process of item 4 wherein said other protein comprises a nucleic acid polymerase.

6. The process of item 5 wherein said polymerase comprises an RNA polymerase and said nucleic acid construct comprises a promoter for said RNA polymerase.

7. The process of item 5 wherein said polymerase comprises a DNA polymerase or reverse transcriptase.

A further embodiment of the invention is construct B which is a construct comprising a host promoter located on the construct such that the host transcribes a sequence in the construct coding for a different RNA polymerase which after translation is capable of recognizing its cognate promoter and transcribing from a DNA sequence of interest in the construct with said cognate promoter oriented such that it does not promote transcription from the construct of said different RNA polymerase.

Specific embodiments of the construct B of the invention are the following items:

1. The construct B wherein said host promoter comprises a prokaryotic or eukaryotic promoter upstream from the host promoter.

2. The construct B wherein said host promoter and the promoter for the second RNA polymerase are located on opposite strands.

3. The construct of item 1 wherein said prokaryotic promoter comprises a RNA polymerase.

4. The construct of item 1 wherein said eukaryotic promoter is selected from Pol I, Pol II and Pol III, or a combination of any of the foregoing.

5. The construct B wherein said second RNA polymerase is selected from T7, T3 and SP6, or a combination of any of the foregoing.

6. The construct B wherein said DNA sequence of interest comprises sense or antisense, or both.

7. The construct B wherein said DNA sequence of interest comprises deoxyribonucleic acid or ribonucleic acid.

IDT Exhibit 1020.9044

8. The construct B wherein said DNA sequence of interest encodes a protein.
9. The construct B comprising at least one chemically modified nucleotide.

The examples which follow are set forth to illustrate various aspects of the present invention but are not intended to limit in any way the scope as more particularly set forth in the claims below.

EXAMPLES

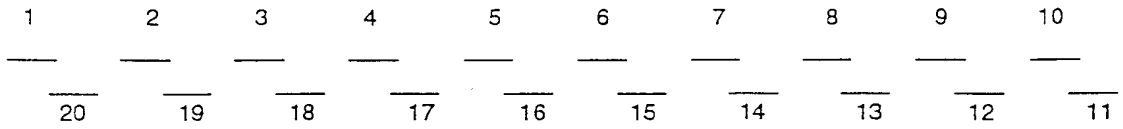
Example 1

Primers

A set of twenty single stranded oligonucleotide primers, fifteen nucleotides long, were chemically synthesized.

The first set of 10 primers was complementary to one strand of M13mp18 replicative form (RF) starting at base 650 and extending to base 341. An interval of 15 nucleotides separated successive primers. The second set of 10 primers contained sequences identical to the single-stranded M13mp18 phage genome starting at base 351 and extending to base 635, again with 15 nucleotide gaps separating successive primers. There is a complementarity of 5 bases between opposing primers, but at an ionic concentration of 0.08M NaCl and 45 °C these primers will not hybridize to each other. The sequences of the primers are shown in Figure 6.

ARRANGEMENT OF OLIGONUCLEOTIDE PRIMERS IN AMPLIFICATION REACTION



Primer 1 begins at base 650 and primer 11 begins at base 351.

Example 2

Amplification Target

The target of amplification was the M13mp18 RF. This was digested with either *Taq*1 or a combination of *Bam*H1 and *Eco*R1. *Eco*R1 and *Bam*H1 cut at sites close to each other and digestion with either enzyme alone would transform the circular RF molecule into a linear DNA molecule. The *Taq*1 enzyme digests M13mp18 RF yielding 12 fragments. The sequence to be amplified (nucleotides 351 to 650) was flanked in the *Bam*H1/*Eco*R1 digested RF by two regions, 1,371 bases and 5,601 bases, and *Taq*1-digested M13mp18 RF was flanked by regions of 15 and 477 nucleotides (see Figure 7).

In amplification experiments, the restriction digests were used without any further purification. For amplification, a control of irrelevant DNA (calf thymus) was employed.

The precursors were added in 50µl aliquots. One 10µl aliquot of the precursors was mixed with 90µl H_2O and loaded on a glass fiber filter, dried and counted. The counts were multiplied by 5 and divided by 160 (nmoles in the incubation mix). Specific activity is the cpm/nmoles of nucleotides.

Amplification is measured as follows. The total counts are determined and this number is divided by the specific activity of the precursors to determine the number of nanomoles of incorporation. The target (in n grams) is divided by 330 (average molecular weight of nucleotide) to determine the nanomoles of input target phosphate. The amplification is then calculated by dividing the nanomoles of product by the nanomoles of input target.

IDT Exhibit 1020.9045

Example 3

The Effect of Primer Concentration on the Amplification of Target DNA.

An incubation mixture of 130 µl contained the following reaction components: 40 mM sodium phosphate, pH 7.5, 400 µM each of the four deoxynucleotide triphosphates, 5 mM dithiothreitol, 40 ng of Taq1-digested M13mp18 RF (containing 3.5 ng of the *Taq*1 fragment to be amplified), and all 20 primers (at 0.04 OD/ml, 0.4 OD/ml or 0.8 OD/ml) and 15 units of Klenow fragment of DNA polymerase. The mixture was left at room temperature for 20 minutes in order to allow the enzyme to cover all of the initiation sites on the template. The polymerization was then initiated by the addition of Mg^{++}, 7 mM final concentration, and the tubes were placed in a 45 °C bath. After 1 hour an additional 15 units of the enzyme were added, and the incubation was continued for another hour. The reaction was stopped with 100 µmoles of EDTA, 100 µg sonicated calf thymus DNA were added, and the nucleic acids were precipitated with 1.0 ml of cold 10% TCA for 60 minutes at 0 °C. The mixture was filtered through a glass fiber filter, the filter was washed 3 times with cold 5% TCA, then twice with ethanol, dried and counted in a Beckman liquid scintillation counter.

The specific activity of the nucleotide precursors was 9,687 cpm/nmole. The tagged *Taq1* DNA fragment contained 0.0106 nmoles of nucleotides.

Primer Concentration	Incorporation (cpm)	Incorporation (nmoles nucleotide)	Amplification
0.04 OD/ml	32,482	3.35	316
0.4OD/ml	366,260	37.8	3566
0.8 OD/ml	512,260	52.88	4988

Example 4

The Random Priming Activity of the Primers on Calf Thymus DNA.

To test for the effect of the primers on the background, an assay was performed, as described in the preceding example (Example 3 above), in which background was determined with and without primers as well as with and without melting of the calf thymus DNA.

The amplification conditions were the same as in Example 1 except that only 5 ug (15.0 nmoles) calf thymus DNA were used as a target. The DNA employed was double stranded or heated at 100 °C for 10 minutes in the presence or absence of primers (0.4 OD/ml each) before being chilled on ice.

Double Stranded DNA	Melted DNA	Primers	Incorporation cpm	Incorporation umoles	Amplification
+			239,100	24.68	1.64
+		+	276,540	28.54	1.90
	+	+	556,560	57.45	3.83
	+		28,432	2.93	0.19

This experiment suggests that the random priming activity of the primers is not substantial, that incorporation on double stranded DNA is due to the nicks on the DNA molecules, and that melting abolishes to a large extent the priming positions on the irrelevant DNA.

Example 5

Amplification of the M13 Fragment in the presence of a large Excess (1500-Fold) of Irrelevant DNA

The amplification conditions were the same as in Example 1. Primers (0.4 OD/ml), 5 ug calf thymus DNA and 40 ng M13mp18 DNA containing 3.5 ng of fragment were employed in this example.

IDT Exhibit 1020.9046

Calf Thymus DNA	M13mp18 DNA	Incorporation cpm	Incorporation nmoles	Amplification
+		575,440	59.4	3.96x
	+	338,900	35.0	3,300x
+	+	713,440	73.6	

The experimental results above show that the target can be amplified in the presence of large amounts of irrelevant DNA. The net amplification was 1,343 even though in this case the target DNA inhibits the amplification of the irrelevant DNA by competing for initiation points. It is possible that the amplification was even larger.

These experimental results were also analyzed by running the samples on a 2% agarose gel. In Figure 8 it can be seen that the calf thymus template (lane 3) only gives high molecular weight DNA bands composed of a mixture of input DNA as well as DNA synthesized by random priming (as seen in the incorporation figures in the Table above given for this example). On the other hand, it can be seen that the mp18 template (lane 2) gives a distinct pattern of lower molecular weight bands, and in lane 1, similar bands are observed when the mp18 template was mixed with 1500 times as much calf thymus DNA demonstrating that the foreign DNA did not significantly affect the amplification of DNA from the mp18 template.

Example 6

Amplification of Different Restriction Digests

The incubation conditions were the same as in Example 4. Forty nanograms of total M13mp18 DNA were used in each experiment with 0.4 OD/ml primers. In one case, the M13mp18 DNA was cut in only one position (using *Eco*R1) leaving the fragment to be amplified flanked by two large pieces. In the other case where the RF was treated with *Taq*1, the fragment was contained in one 639 base pair fragment. The specific activity of the precursors was 8.385 cpm/nmole.

	Incorporation cpm	Incorporation nmoles
Large Fragment	393,480	46.92
Small Fragment	262,808	31.34

These experimental results show that the enzyme does not extend polymerization very far from the region where the primers hybridize, otherwise a much larger incorporation using the large piece would have been otherwise expected because the elongation of the primers by the enzyme can extend in both directions.

Example 7

A Comparison of Synchronized and Unsynchronized Reactions

In all of the preceding experiments, the enzyme was preincubated with the target-primer mixture to allow binding of the enzyme at the 3' end of the hybridized primers on the target, followed by the addition of magnesium to initiate polymerization. The conditions were the same as in Example 1.

To assay the effect of this synchronization on the extent of the reaction, the incorporation in a synchronized reaction was compared to an unsynchronized reaction initiated by adding magnesium to the complete reaction mix before enzyme addition. The reaction conditions are described in Example 3. The specific activity was 9687 cpm/nmole.

	Incorporation cpm	Incorporation nmoles	Amplification
Synchronized	495,080	51.1	4818
Unsynchronized	416,380	42.9	4052

IDT Exhibit 1020.9047

The results above demonstrate that synchronization of the reaction is not essential for the amplification reaction.

Example 8

The Effect of Variations of the Reaction Conditions on the Product Produced by the Procedure of Example 3

A reaction was performed according the the reaction conditions of Example 3 in which twenty primers were added to the reaction mixture as well as the *Taq 1* fragments (40 nanograms, i.e., 3.5 nanograms of insert that will hybridize with the primers) described in Example 3 with the exception that the buffer was altered. In the first lane of the gel shown in Figure 9, the reaction was performed without target DNA added. In lane 2 the reaction was performed in a phosphate buffer (0.04 M, pH 7.5). Lane 3 contains the molecular weight buffers of *Msp I* digestion of pBR322 DNA. In the fourth lane the reaction was performed in which the phosphate buffer was substituted by MOPS buffer at 0.1 M and pH 7.5 (measured 25 °C). It can be seen that the reaction in the phosphate buffer produced an agglomeration of DNAs that when dissociated by heat or other double helix disrupting agents lead to an number of products of a size smaller than the agglomeration structures. The size distribution of the products in the MOPS-buffered reaction corresponds to the distances between certain of the oligo primer binding sites. The smallest is approximately 76 nucleotide pairs in size which is approximately the distance between the closest specific oligo primer binding sites.

Example 9

Effect of Various Buffers on the Amplification Reaction.

The buffer used for the amplification reaction can have significant effects upon the degree of amplification. In the following example, phosphate buffer (which was employed in Example 7) was compared with the following zwitterion buffers:
4-morpholinoethyl sulfonate (MES),
4-morpholinopropionyl sulfonate (MOPS),
N-dimethylaminobutyric acid (DMAB), and
N-dimethylglycine (DMG).

Trizma base was used to adjust MES or MOPS to pH 7.5, DMAB to pH 7.8, and DMG to pH 8.6. In the previous experiments, 4.0 ng of mp18 (containing 3.5 ng of the target fragment) was used as a template. In this experiment, the amount of template was reduced ten-fold compared to those experiments (4 ng of mp18; 350 pg of target fragment). Other changes in the experimental procedure was the omission of DTT and the use of a single addition of 10 units of Klenow polymerase. Mg^{++} and dNTP concentrations (7.5 mM and 400 µM each dNTP) were as described previously.

As before, reactions were preincubated at room temperature for 30 minutes prior to the addition of the Mg^{++}. After addition of Mg^{++}, reactions were immediately transferred to a 45 °C water bath and incubated for 4 hours. The reaction was stopped by the addition of 5µl of 500 mM EDTA to give a final concentration of approximately 20 mM.

For evaluation of the extent of polymerization, an aliquot of 40 µl (out of a 120 µl incubation mix) was mixed with 50 µg of sonicated calf thymus DNA and precipitated on ice with 1 ml of 10% TCA. After one hour, the precipitate was collected on glass fiber filters, washed 3 times with 5% of cold TCA, 2 times with 95% ETOH, dried and counted in a liquid scintillation counter. The input was measured by the addition of radioactive precursor onto a filter without precipitation with TCA and counted as before. The results are given in the table below. As controls, the reactions were also carried out without the addition of any target template.

IDT Exhibit 1020.9048

Buffer	Incorporation From Template (in cpm)	No Template Control (in cpm)	Template-Specific Incorporation (in cpm)	Net Synthesis (nanomoles)	Amplification Factor
Phosphate	4,008	2,628	1,380	0.255	240
MES	299,367	212,778	86,589	16.03	15,123
MOPS	184,500	49,521	114,979	21.28	20,075
DMAB	207,239	5,859	211,380	39.13	36,915
DMG	182,655	32,012	150,643	27.89	26,311

Compared to the no template control, the highest efficiency of amplification was obtained with the DMAB buffer. The results of this experiment demonstrated that an amplification of the target region approaching 37,000 fold could be obtained. It should be noted that another buffer, MES, gave higher incorporation, but the no template control demonstrated that there was non-specific polymerization leading to a net amplification of only 20,000 fold. The next best buffer system was DMG where net amplification was over 26,000 fold, followed by MOPS with 20,000 fold amplification.

The results of this experiment were also analyzed qualitatively by ethanol precipitating the remaining 80 µl of the reaction mixtures, resuspending them in 80 µl of TE buffer and running 10 µl aliquots on 2% agarose gels. These results are shown in Figure 10 and agree with the results shown in the table above.

Example 10

Incorporation of radioactive precursors measures total synthesis of DNA including both specific as well as template-independent DNA synthesis. Oligos No. 1,3,5,7,9,12,14,16,18 and 20 from Example 1 were employed in a series of amplification reactions. This limited number was chosen such that there would be a region within the amplicon that does not correspond to any of the primers, allowing the use of a 30 base probe (bases 469 to 498) labeled with biotin that corresponds to this open region.

The experimental design was to use DMAB and DMG buffers. Example 9 had previously shown little or no template-independent synthesis with DMAB and substantial template-independent synthesis with DMG. Reactions with and without Taq digested mp18 were carried out. The reaction mixtures were precipitated with ethanol, resuspended in TE buffer and aliquots were electrophoresed through a 2% agarose gel. A southern blot was made from this gel and probed with 200 ng/ml labeled oligo in 31% formamide/ 2X SSC at 25 °C for 2 hours and washed 3 X with 0.1X SSC/0.1% Triton X-100 for 5 minutes each at 37 °C. Membranes were developed using an alkaline phosphate detection system obtained from Enzo Biochem, Inc.

As seen in Figure 11, this set of experiments demonstrates that the product of the amplification is strongly dependent upon the specific buffer used in the reaction. The best results were obtained with DMAB buffer where essentially no incorporation (data not shown) or hybridization (Figure 11, lane 1) with the reaction mixture from the no template control sample. The template dependent synthesis with DMAB (Figure 11, lane 2) produced DNA that hybridized with the labeled probe.

The nature and origin of the non-template derived synthesis achieved with DMG buffer (Figure 11, lane 3) is still under current study.

Example 11

Determination of the Nature of the Ends of the Amplified Product

If the product strands act as the template for polymerization of nucleic acid then the products should have blunt ends. One method of assaying for the presence of blunt ends is based on the notion that these molecules will undergo blunt end ligation. Molecules with 'ragged' ends (single stranded tails on the 3' or 5' end) will not participate in the ligation reaction.

Because the amplified product is initiated using chemically synthesized primer molecules, the 5' ends will not under phosphorylation. The first step of this proof will be to phosphorylate the 5' ends of both single stranded and double stranded molecules. These 5' phosphorylated molecules will then be ligated using the

IDT Exhibit 1020.9049

T4 DNA ligase. The unamplified DNA will act as the negative control and a PCR-generated amplicon will act as the positive control.

As can be seen in the gel reported in Figure 12, T4 ligase treatment increases the molecular weight of the amplified product selectively. This is most apparant in lane 4 of Figure 12, where there is an appreciable increase in size observed as a result of the completed ligation reaction.

The positive control with the PCR-generated amplicon (primed by oligos initiating at nucleotide 381 and from nucleotide 645 of the template which predicts an amplicon of 264 nucleotides) also show a shift in position after ligation (lane 7). Because there was not much DNA included in this reaction, the appearance of a spectrum of multimers of the amplicon was not observed, but the loss of material from the position of the unligated material (lanes 5 and 6) was noted. Some material left at the position of the untreated amplicon in lane 7 was also noted. It is possible that this material does not participate in the ligation reaction because of the addition of A to the 3' end of the amplicon which is a property of the Taq polymerase.

Example 12

Amplification from non-denatured template

To explain the high level of amplification in this system, it was assumed that some of the primers might be able to initiate DNA synthesis by inverting the ends of double-stranded DNA products synthesized during amplification. This "breathing" was demonstrated in the following experiment. The template was a blunt-ended double-stranded DNA molecule obtained from Dr. Christine L. Brakel, the blunt ends extending from bases 371 to 645 in the mp18 DNA sequence. These ends exactly match primers Nos. 1 and 12 (described in Example 1). In this experiment, no radioactive precursors were used. Analysis was performed by gel electrophoresis through 2% agarose. Reagent conditions were the same as Example 10 where DMG was used as the buffer, but only 2 primers, No. 1 and No. 2 were used and no denaturation of the starting template was performed. In Figure 13, for comparison purposes, the same amount of DNA was used as the input on the gel (lane 1). In lane 2, no template was added. In lane 3, the complete reaction mix is shown. In lane 4, 12 times as much DNA as the template input in either lanes 1 or 3 has been included as a size marker. In both lanes 2 and 3, some non-specific synthesis can be seen, but the specific copying of the template can clearly be distinguished in lane 3. Therefore, as lane 3 indicates, newly synthesized DNA of the same size as the input was formed using non-denatured double-stranded DNA, proving that the double-stranded blunt ends can serve as initiation points for replication.

Example 13

Amplication from RNA template

Although it has been demonstrated by the present invention that DNA can be amplified, it would be useful, however, to also be able to employ RNA as a potential template. Accordingly, the double-stranded DNA molecule used in Example 12 was ligated into the Sma I site of a vector pIBI31 (obtained from International Biotechnology Corp) that contains a promotor for T7 RNA polymerase. Using standard methodologies, an RNA transcript was synthesized encompassing the same sequences used in example 12. This transcript was amplified using standard conditions with all 20 primers in DMG buffer. Taq digested mp18 DNA was used as a control. As seen in Figure 14 there was extensive synthesis. There are characteristic bands that appear in lane 4, the reaction with the RNA template, as well as in lane 2, the reaction with the DNA template that do not appear in the template independent synthesis seen in lanes 1 and 5.

This demonstrates that the system is capable of utilizing the RNA transcript as a template without the introduction of any other enzyme besides the Klenow, thus proving that the Klenow enzyme can be used as a reverse transcriptase as indicated in the disclosure. This result was studied further by making a Southern blot of the gel seen in Figure 14 and probing with nick-translated biotinylated mp18 DNA using a nick translation kit obtained from Enzo Biochem, Inc. As seen in Figure 15, there was little or no hybridization of the probe to the reaction product of the template independent reactions (lanes 1 and 5) whereas extensive hybridization was observed with the RNA derived reaction (lane 4) as well as the DNA template derived reaction (lane 2), as expected.

IDT Exhibit 1020.9050

Example 14

Strand Displacement Using Ethidium-Labeled Oligonucleotides

Three oligonucleotides with the sequence listed in Figure 16 were prepared and labeled F1, F1C and F3. The unlabeled complement of F1C was hybridized to unlabeled F1. The ratio of F1C: F1 for the hybridization was 1:2. (F1C at a concentration of 0.13 O.D/ml and F1 at a concentration of 0.26 O.D./ml.) Hybridization was performed in 1X SSC for two hours at 45 ° C.

Aliquots of the hybrid were mixed with different amounts of ethidium-labeled F1 (F1E) in 1X SSC and incubated for 18 hours either at 43 ° C. or at 37 ° C. The ratios of F1E oligo to the unlabeled oligo F1C was 1:1, 2:1, 3:1 and 4:1. (The 1:1 reaction contained 0.0325 O.D of the F1E, 0.065 O.D. of F1 and 0.0325 O.D. of F1C.) At the end of the incubation period, 50μl of each mixture was incubated with 50μl of diazonium mixture for 5 minutes at room temperature. To prepare the diazonium mixture, 10 μl of the diazonium stock solution, (50 mM in 1M HCl), was added to 100 μl of cold dilution buffer, (1 X SSC and 0.2 M $KHCO_3$, prepared fresh). The diazonium stock solution is stored at -20 ° C.

Under these conditions the diazonium will destroy the fluorescence associated with the ethidium in single stranded oligonucleotides. See, e.g., European Patent Application Publication No. 0 492 570 A1, published on July 1, 1992, based on priority document, U.S. Patent Application Serial No. 07/633,730, filed on December 24, 1990, the contents of which are incorporated by reference. But the diazonium will not destroy the fluorescence associated with the ethidium that has intercalated into the double stranded DNA. The survival of the ethidium, under these reaction conditions, is a measure of the extent of formation of a double helix by the ethidium-labeled oligonucleotides, thus indicating displacement of the non-ethidium containing strand by that of the ethidium labeled. This property of the ethidium labeled oligonucleotides by primers can be usefully employed to facilitate initiation of polymerization on double stranded templates. As seen in the figure in Figure 17, the ethidium-labeled oligo displaces the non-ethidium-labeled oligo better at 43 ° C than at 37 ° C.

Example 15

T7 Promoter Oligonucleotide 50 Mer Labeled with Ethidium

An oligonucleotide 50-mer including the T7 promoter region of IBI 31 plasmid constructs (plasmid sequences derived from manufacturer, International Biotechnology, Inc.) was synthesized. Its sequence is as follows:

3'-TAC T*AA T*GC GGT* CT*A T*AG T*T--AA TCA TGA AT--T AAT* TAT* GCT* GAG T*GA T*AT* C-5',

where T* represents allylamine dU, and therefore ethidium modification and the 10 base sequence set off by dashes (--AA TCA TGA AT--) was introduced to provide a restriction enzyme site.

Example 16

Use of the Oligonucleotide 50-Mer to Regulate RNA Synthesis *In Vitro*

This nucleotide is complementary to the ATG strand of the lac z gene of IBI 31, and also contains a 10-base sequence for use in restriction enzyme digestion. The oligonucleotide 50-mer also contains sequences overlapping the T7 promotor in the IBI 31 plasmid constructs. Thus, it might be expected to interfere with *in vitro* transcription by T7 RNA polymerase even though the sequences in this oligo are entirely upstream of the start of transcription by T7 RNA polymerase. Because the plasmid constructs contain opposing T7 and T3 promotors, this also means that the oligo 50-mer is identical in sequence to the RNA that is made by the T3 RNA polymerase *in vitro*.

The effect of this oligonucleotide on *in vitro* transcription by T7 and T3 polymerases from an IBI 31 plasmid construct (pIBI 31-BH5-2) and from a BlueScript II plasmid construct (pBSII/HCV) was studied. See Figure 18 which contains the same target sequences, but in a "split" arrangement. The polylinker sequences of these plasmids are given in Figure 18. Comparing the effect of the oligo on these two

IDT Exhibit 1020.9051

different target template serves to partially control for the possible non-specific inhibitory effects of ethidium groups on the RNA polymerases because the oligonucleotide would be expected to inhibit transcription from any template containing an appropriate promotor regardless of the "split" if the effect were due to the oligo's interaction with the polymerase rather than with the template.

At a concentration of 60-fold excess of oligonucleotide (0.6μM final) over plasmid with either the allylamine labeled oligonucleotide of the ethidium labeled oligonucleotide in a transcription reaction mixture, the following results were obtained:

Plasmid Transcribed	Polymerase Used	Oligo Used	nanomoles Incorporated	% of control
pIBI 31-BH5-2	T3	None	236	100
pIBI 31-BH5-2	T3	Allylamine labeled	233	99
pIBI 31-BH5-2	T3	Ethidium labeled	87	37
pIBI 31-BH5-2	T7	None	208	100
pIBI 31-BH5-2	T7	Allylamine labeled	198	95
pIBI 31BH-5-2	T7	Ethidium labeled	3	1.4
pBSII/HCV	T3	None	112	100
pBSII/HCV	T3	Allylamine labeled	158	>100
pBSII/HCV	T3	Ethidium labeled	185	>100
pBSII/HCV	T7	None	125	100
pBSII/HCV	T7	Allylamine labeled	154	>100
pBSII/HCV	T7	Ethidium labeled	62	50

These results indicate that the ethidium-modified oligo sequence is capable of specifically inhibiting transcription by the T7 polymerase from the T7 promotor region provided that the promoter region is not interrupted by the multiple cloning region and inserted DNA. Thus, the effect is dependent on the template DNA and is not merely the result of inhibition of the T7 polymerase by the ethidium groups.

Many obvious variations will be suggested to those of ordinary skill in the art in light of the above detailed description of the invention. All such variations are fully embraced by the scope and spirit of the present invention as set forth in the claims which follow.

Claims

1. An *in vitro* process for producing more than one copy of a specific nucleic acid, said process being independent of a requirement for the introduction of an intermediate structure for the production of said specific nucleic acid, said process comprising the steps of:

(a) providing a nucleic acid sample containing or suspected of containing the sequence of said specific nucleic acid;

(b) contacting said sample with a mixture comprising:

(i) nucleic acid precursors,

(ii) one or more specific nucleic acid primers each of which is complementary to a distinct sequence of said specific nucleic acid, and

(iii) an effective amount of a nucleic acid producing catalyst; and

(c) allowing said mixture to react under isostatic conditions of temperature, buffer and ionic strength, thereby producing more than one copy of said specific nucleic acid.

2. An *in vitro* process for producing more than one copy of a specific nucleic acid, said products being substantially free of any primer-coded sequences, said process comprising the steps of:

(a) providing a nucleic acid sample containing or suspected of containing the sequence of said specific nucleic acid;

(b) contacting said sample with a mixture comprising:

(i) nucleic acid precursors,

(ii) one or more specific polynucleotide primers comprising at least one ribonucleic acid segment each of which primer is substantially complementary to a distinct sequence of said specific nucleic acid, and

(iii) an effective amount of a nucleic acid producing catalyst; and

IDT Exhibit 1020.9052

(c) allowing said mixture to react under isostatic conditions of temperature, buffer and ionic strength, thereby producing at least one copy of said specific nucleic acid; and
(d) removing substantially or all primer-coded sequences from the product produced in step (c) to regenerate a primer binding site, thereby allowing a new priming event to occur and producing more than one copy of said specific nucleic acid.

3. An *in vitro* process for producing more than one copy of a specific nucleic acid, said products being substantially free of any primer-coded sequences, said process comprising the steps of:
(a) providing a nucleic acid sample containing or suspected of containing the sequence of said specific nucleic acid;
(b) contacting said sample with a mixture comprising:
(i) unmodified nucleic acid precursors,
(ii) one or more specific chemically-modified primers each of which primer is substantially complementary to a distinct sequence of said specific nucleic acid, and
(iii) an effective amount of a nucleic acid producing catalyst; and
(c) allowing said mixture to react under isostatic conditions of temperature, buffer and ionic strength, thereby producing at least one copy of said specific nucleic acid; and
(d) removing substantially or all primer-coded sequences from the product produced in step (c) to regenerate a primer binding site, thereby allowing a new priming event to occur and producing more than one copy of said specific nucleic acid.

4. An *in vitro* process for producing more than one copy of a specific nucleic acid, said products being substantially free of any primer-coded sequences, said process comprising the steps of:
(a) providing a nucleic acid sample containing or suspected of containing the sequence of said specific nucleic acid;
(b) contacting said sample with a mixture comprising;
(i) unmodified nucleic acid precursors,
(ii) one or more specific unmodified primers comprising at least onesegment each of which primer comprises at least one non-complementary sequence to a distinct sequence of said specific nucleic acid, such that upon hybridization to said specific nucleic acid at least one loop structure is formed, and
(iii) an effective amount of a nucleic acid producing catalyst; and
(c) allowing said mixture to react under isostatic conditions of temperature, buffer and ionic strength, thereby producing at least one copy of said specific nucleic acid; and
(d) removing substantially or all primer-coded sequences from the product produced in step (c) to regenerate a primer binding site, thereby allowing a new priming event to occur and producing more than one copy of said specific nucleic acid.

5. A promoter-independent non-naturally occurring nucleic acid construct which when present in a cell produces a nucleic acid without the use of any gene product coded by said construct.

6. A conjugate comprising a protein-nucleic acid construct, said nucleic acid construct not coding for said protein, and which conjugate produces a nucleic acid when present in a cell.

7. An *in vivo* process for producing a specific nucleic acid, said process comprising the steps of:
(a) providing a conjugate comprising a protein-nucleic acid construct, said conjugate being capable of producing a nucleic acid when present in a cell; and
(b) introducing said conjugate into a cell, thereby producing said specific nucleic acid.

8. A construct comprising a host promoter located on the construct such that the host transcribes a sequence in the construct coding for a different RNA polymerase which after translation is capable of recognizing its cognate promoter and transcribing from a DNA sequence of interest in the construct with said cognate promoter oriented such that it does not promote transcription from the construct of said different RNA polymerase.

IDT Exhibit 1020.9053

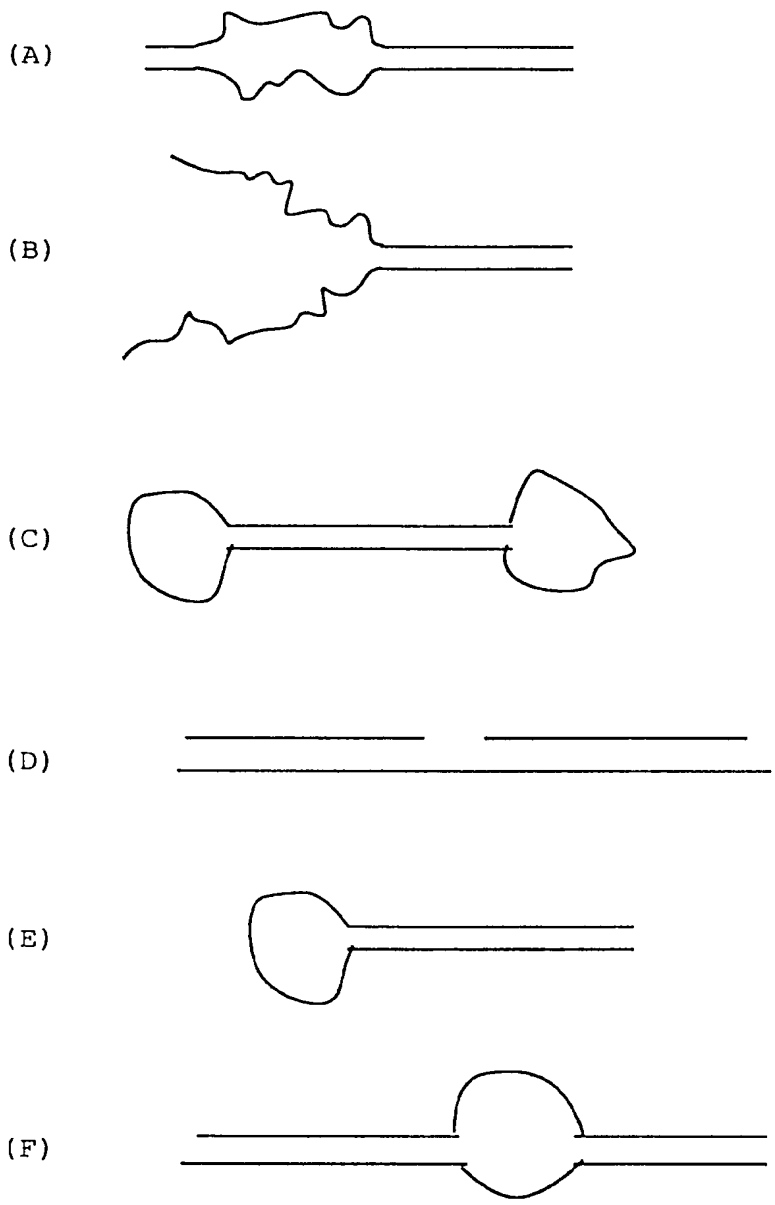


Figure 1 (A-F)

Construct Forms Comprising at Least one Single-Stranded Region

IDT Exhibit 1020.9054

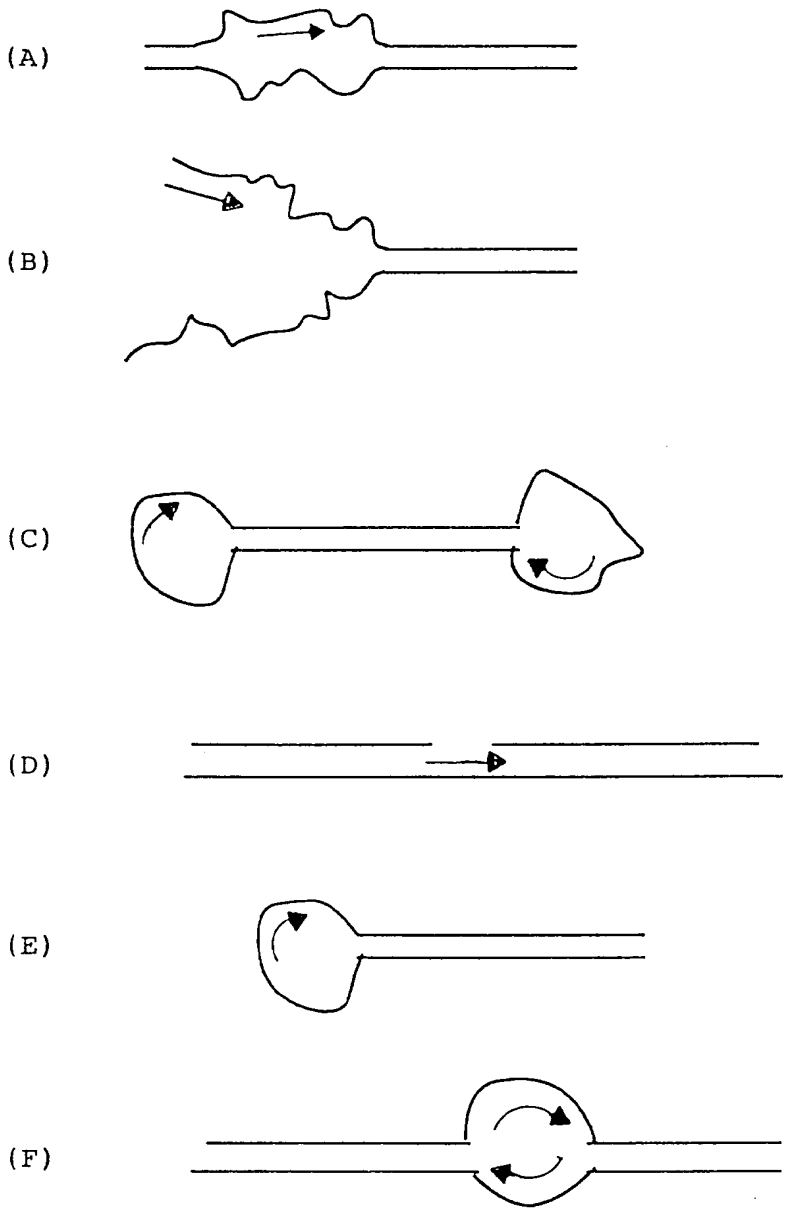


Figure 2 (A-F)

Functional Forms of the Construct

IDT Exhibit 1020.9055

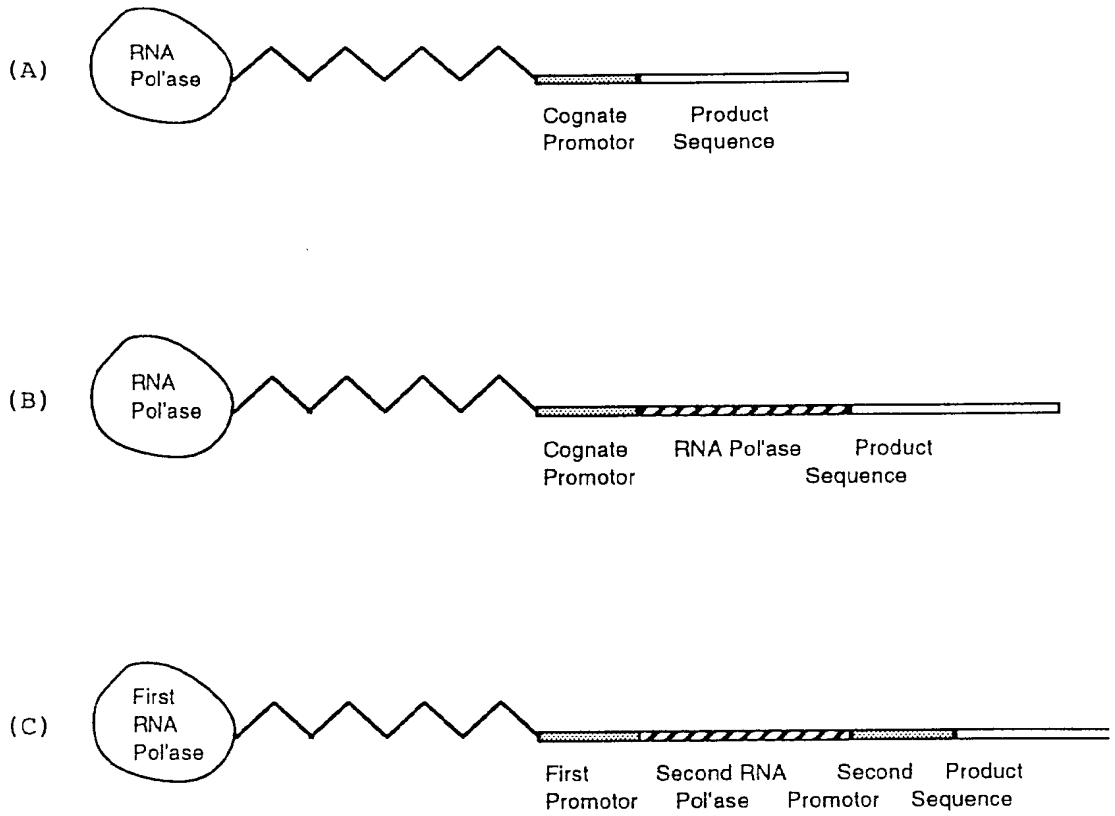


Figure 3 (A-C)

Three Constructs with an RNA Polymerase Covalently Attached to a Transcribing Cassette

IDT Exhibit 1020.9056

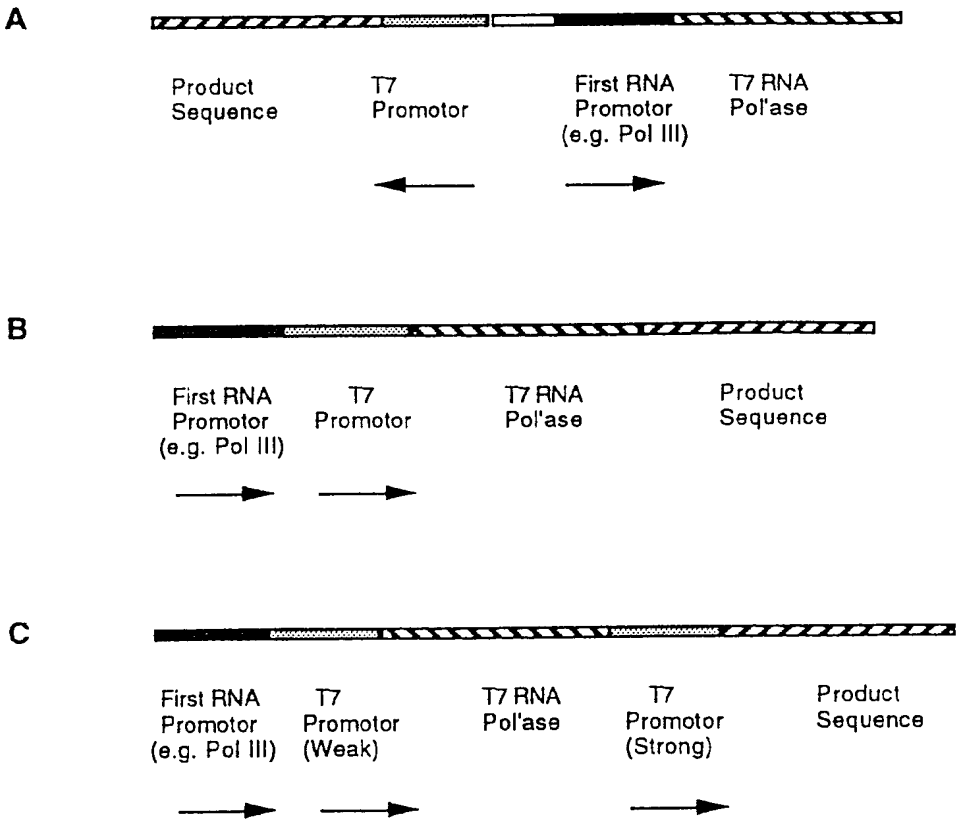


Figure 4 (A-C)

Three Constructs with Promoters for Endogenous RNA Polymerase

IDT Exhibit 1020.9057

M13mp18. Seq Length: 7250

1.	AATGCTACTA	CTATTAGTAG	AATTGATGCC	ACCTTTTCAG	CTCGCGCCCC
51.	AAATGAAAAT	ATAGCTAAAC	AGGTTATTGA	CCATTTGCGA	AATGTATCTA
101.	ATGGTCAAAC	TAAATCTACT	CGTTCGCAGA	ATTGGGAATC	AACTGTTACA
151.	TGGAATGAAA	CTTCCAGACA	CCGTACTTTA	GTTGCATATT	TAAAACATGT
201	TGAGCTACAG	CACCAGATTC	AGCAATTAAG	CTCTAAGCCA	TCCGCAAAAA
251	TGACCTCTTA	TCAAAAGGAG	CAATTAAAGG	TACTCTCTAA	TCCTGACCTG
301.	TTGGAGTTTG	CTTCCGGTCT	GGTTCGCTTT	GAAGCTCGAA	TTAAAACGCG
351.	ATATTTGAAG	TCTTTCGGGC	TTCCTCTTAA	TCTTTTTGAT	GCAATCCGCT
401.	TTGCTTCTGA	CTATAATAGT	CAGGGTAAAG	ACCTGATTTT	TGATTTATGG
451.	TCATTCTCGT	TTTCTGAACT	GTTTAAAGCA	TTTGAGGGGG	ATTCAATGAA
501.	TATTTATGAC	GATTCCGCAG	TATTGGACGC	TATCCAGTCT	AAACATTTTA
551.	CTATTACCCC	CTCTGGCAAA	ACTTCTTTTG	CAAAAGCCTC	TCGCTATTTT
601.	GGTTTTTATC	GTCGTCTGGT	AAACGAGGGT	TATGATAGTG	TTGCTCTTAC
651.	TATGCCTCGT	AATTCCTTTT	GGCGTTATGT	ATCTGCATTA	GTTGAATGTG
701.	GTATTCCTAA	ATCTCAACTG	ATGAATCTTT	CTACCTGTAA	TAATGTTGTT
751.	CCGTTAGTTC	GTTTTATTAA	CGTAGATTTT	TCTTCCCAAC	GTCCTGACTG
801.	GTATAATGAG	CCAGTTCTTA	AAATCGCATA	AGGTAATTCA	CAATGATTAA
851.	AGTTGAAATT	AAACCATCTC	AAGCCCAATT	TACTACTCGT	TCTGGTGTTC
901.	TCGTCAGGGC	AAGCTTATT	CACTGAATGA	GCAGCTTTGT	TACGTTGATT
951.	TGGGTAATGA	ATATCCGGTT	CTTGTCGAAG	ATTACTCTTG	ATGAAGGTCA
1001	GCCAGCCTAT	GCGCCTGGTC	TGTACACCGT	TCATCTGTCC	TCTTTCAAAG
1051	TTGGTCAGTT	CGGTTCCCTT	ATGATTGACC	GTCTGCGCCT	CGTTCCGGCT
1101	AAGTAACATG	GAGCAGGTCG	CGGATTTCGA	CACAATTTAT	CAGGCGATGA
1151	TACAAATCTC	CGTTGTACCTT	TGTTTCGCGC	TTGGTATAAT	CGCTGGGGGT
1201	CAAAGATGAG	TGTTTTAGTG	TATTCTTTCG	CCTCTTTCGT	TTTAGGTTGG

Figure 5

M13mp18 Nucleic Acid Sequence

IDT Exhibit 1020.9058

1251	TGCCTTCGTA	GTGGCATTAC	GTATTTTACC	CGTTTAATGG	AAACTTCCTC
1301	ATGAAAAAGT	CTTTAGTCCT	CAAAGCCTCT	GTAGCCGTTG	CTACCCTCGT
1351	TCCGATGCTG	TCTTTCGCTG	CTGAGGGTGA	CGATCCCGCA	AAAGCGGCCT
1401	TTAACTCCCT	GCAAGCCTCA	GCGACCGAAT	ATATCGGTTA	TGCGTGGGCG
1451	ATGGTTGTTG	TCATTGTCGG	CGCAACTATC	GGTATCAAGC	TGTTTAAGAA
1501	ATTCACCTCG	AAAGCAAGCT	GATAAACCGA	TACAATTAAA	GGCTCCTTTT
1551	GGAGCCTTTT	TTTTTGGAGA	TTTTCAACGT	GAAAAAATTA	TTATTCGCAA
1601	TTCCTTTAGT	TGTTCCTTTC	TATTCTCACT	CCGCTGAAAC	TGTTGAAAGT
1651	TGTTTAGCAA	AACCCCATAC	AGAAAATTCA	TTTACTAACG	TCTGGAAAGA
1701	CGACAAAACT	TTAGATCGTT	ACGCTAACTA	TGAGGGTTGT	CTGTGGAATG
1751	CTACAGGCGT	TGTAGTTTGT	ACTGGTGACG	AAACTCAGTG	TTACGGTACA
1801	TGGGTTCCTA	TTGGGCTTGC	TATCCCTGAA	AATGAGGGTG	GTGGCTCTGA
1851	GGGTGGCGGT	TCTGAGGGTG	GCGGTTCTGA	GGGTGGCGGT	ACTAAACCTC
1901	CTGAGTACGG	TGATACACCT	ATTCCGGGCT	ATACTTATAT	CAACCCTCTC
1951	GACGGCACTT	ATCCGCCTGG	TACTGAGCAA	AACCCGCTA	ATCCTAATCC
2001	TTCTCTTGAG	GAGTCTCAGC	CTCTTAATAC	TTTCATGTTT	CAGAATAATA
2051	GGTTCCGAAA	TAGGCAGGGG	GCATTAACTG	TTTATACGGC	CACTGTTACT
2101	CAAGGCACTG	ACCCCGTTAA	AACTTATTAC	CAGTACACTC	CTGTATCATC
2151	AAAAGCCATG	TATGACGCTT	ACTGGAACGG	TAAATTCAGA	GACTGCGCTT
2201	CAAGGCACTG	ACCCCGTTAA	AACTTATTAC	CAGTACACTC	CTGTATCATC
2151	AAAAGCCATG	TGCCTCAACC	TCCTGTCAAT	GCTGGCGCGG	GCTCTGGTGG
2201	TCCATTCTGG	CTTTAATCAA	GATCCATTCG	TTTGTGAATA	TCAAGGCCAA
2251	TCGTCTGACC	TGCCTCAACC	TCCTGTCAAT	GCTGGCGCGG	GCTCTGGTGG
2301	TGGTTCTGGT	GGCGGCTCTG	AGGGTGGTCG	CTCTGAGGGT	GGCGGTTCTG
2351	AGGGTGGCGG	CTCTGAGGGA	GGCGGTTCCG	GTGGTGGCTC	TGGTTCCGGT
2401	GATTTTGATT	ATGAAAAGAT	GGCAAACGCT	AATAAGGGGG	CTATGACCGA
2451	AAATGCCGAT	GAAAACGCGC	TACAGTCTGA	CGCTAAAGGC	AAACTTGATT

Figure 5

M13mp18 Nucleic Acid Sequence

IDT Exhibit 1020.9059

2501 CTGTCGCTAC TGATTACGGT GCTGCTATCG ATGGTTTCAT TGGTGACGTT
2551 TCCGGCCTTG CTAATGGTAA TGGTGCTACT GGTGATTTTG CTGGCTCTAA
2601 TTCCCAAATG GCTCAAGTCG GTGACGGTGA TAATTCACCT TTAATGAATA
2651 ATTTCCGTCA ATATTTACCT TCCCTCCCTC AATCGGTTGA ATGTCGCCCT
2701 TTTGTCTTTA GCGCTGGTAA ACCATATGAA TTTTCTATTG ATTGTGACAA
2751 AATAAACTTA TTCCGTGGTG TCTTTGCGTT TCTTTTATAT GTTGCCACCT
2801 TTATGTATGT ATTTTCTACG TTTGCTAACA TACTGCGTAA TAAGGAGTCT
2851 TTATCATGCC AGTTCTTTTG GGTATTCCGT TATTATTGCG TTTCCTCGGT
2901 TTCCTTCTGG TAACTTTGTT CGGCTATCTG CTTACTTTTC TTAAAAAGGG
2951 CTTCGGTAAG ATAGCTATTG CTATTTCATT GTTTCTTGCT CTTATTATTG
3001 GGCTTAACTC AATTCTTGTG GGTTATCTCT CTGATATTAG CGCTCAATTA
3051 CCCTCTGACT TTGTTCAGGG TGTTCAGTTA ATTCTCCCGT CTAATGCGCT
3101 TCCCTGTTTT TATGTTATTC TCTCTGTAAA GGCTGCTATT TTCATTTTTG
3151 ACGTTAAACA AAAAATCGTT TCTTATTTGG ATTGGGATAA ATAATATGGC
3201 TGTTTATTTT GTAACTGGCA AATTAGGCTC TGGAAAGACG CTCGTTAGCG
3251 TTGGTAAGAT TCAGGATAAA ATTGTAGCTG GGTGCAAAAT AGCAACTAAT
3301 CTTGATTTAA GGCTTCAAAA CCTCCCGCAA GTCGGGAGGT TCGCTAAAAC
3351 GCCTCGCGTT CTTAGAATAC CGGATAAGCC TTCTATATCT GATTTGCTTG
3401 CTATTGGGCG CGGTAATGAT TCCTACGAATG AAAATAAAAA CGGCTTGCTT
3451 GTTCTCGATG AGTGCGGTAC TTGGTTTAAT ACCCGTTCTT GGAATGATAA
3501 GGAAAGACAG CCGATTATTG ATTGGTTTCT ACTGCTCGT AAATTAGGAT
3551 GGGATATTAT TTTTCTTGTT CAGGACTTAT CTATTGTTGA TAAACAGGCG
3601 CGTTCTGCAT TAGCTGAACA TGTTGTTTAT TGTCGTCGTC TGGACAGAAT
3651 TACTTTACCT TTTGTCGGTA CTTTATATTC TCTTATTACT GGCTCGAAAA
3701 TGCCTCTGCC TAAATTACAT GTTGGCGTTG TTAAATATGG CGATTCTCAA
3751 TTAAGCCCTA CTGTTGAGCG TTGGCTTTAT ACTGGTAAGA ATTTGTATAA
3801 CGCATATGAT ACTAAACAGG CTTTTTCTAG TAATTATGAT TCCGGTGTTT

Figure 5

M13mp18 Nucleic Acid Sequence

IDT Exhibit 1020.9060

3851	ATTCTTATTT	AACGCCTTAT	TTATCACACG	GTCGGTATTT	CAAACCATTA
3901	AATTTAGGTC	AGAAGATGAA	ATTAACTAAA	ATAATATTGA	AAAAGTTTTC
3951	TCGCGTTCTT	TGTCTTGCGA	TTGGATTTGC	ATCAGCATTT	ACATATAGTT
4001	ATATAACCCA	ACCTAAGCCG	GAGGTTAAAA	AGGTAGTCTC	TCAGACCTAT
4051	GATTTTGATA	AATTCACTAT	TGACTCTTCT	CAGCGTCTTA	ATCTAAGCTA
4101	TCGCTATGTT	TTCAAGGATT	CTAAGGGAAA	ATTAATTAAT	AGCGACGATT
4151	TACAGAAGCA	AGGTTATTCA	CTCACATATA	TTGATTTATG	TACTGTTTCC
4201	ATTAAAAAAG	GTAATTCAAA	TGAAATTGTT	AAATGTAATT	AATTTTGTTT
4251	TCTTGATGTT	TGTTTCATCA	TCTTCTTTTG	CTCAGGTAAT	TGAAATGAAT
4301	AATTCGCCTC	TGCGCGATTT	TGTAACTTGG	TATTCAAAGC	AATCAGGCGA
4351	AATCCGTTATT	GTTTCTCCCG	ATGTAAAAGG	TACTGTTACT	GTATATTCAT
4401	CTGACGTTAA	ACCTGAAAAT	CTACGCAATT	TCTTTATTTC	TGTTTTACGT
4451	GCTAATAATT	TTGATAATGGT	TGGTTCAATT	CCTTCCATAA	TTCAGAAGTA
4501	TAATCCAAAC	AATCAGGATT	ATATTGATGA	ATTGCCATCA	TCTGATAATC
4551	AGGAATATGA	TGATAATTCC	GCTCCTTCTG	GTGGTTTCTT	TGTTCCGCAA
4601	AATGATAATG	TTACTCAAAC	TTTTAAAATT	AATAACGTTC	GGGCAAAGGA
4651	TTTAATACGA	GTTGTCGAAT	TGTTTGTAAA	GTCTAATACT	TCTAAATCCT
4701	CAAATGTATT	ATCTATTGAC	GGCTCTAATC	TATTAGTTGT	TAGTGCTCCT
4751	AAAGATATTT	TAGATAACCT	TCCTCAATTC	CTTTCTACTG	TTGATTTGCC
4801	AACTGACCAG	ATATTGATTG	AGGGTTTGAT	ATTTGAGGTT	CAGCAAGGTG
4851	ATGCTTTAGA	TTTTTCATTT	GCTGCTGGCT	CTCAGCGTGG	CACTGTTGCA
4901	GGCGGTGTTA	ATACTGACCG	CCTCACCTCT	GTTTTATCTT	CTGCTGGTGG
4951	TTCGTTCGGT	ATTTTTAATG	GCGATGTTTT	AGGGCTATCA	GTTCGCGCAT
5001	TAAAGACTAA	TAGCCATTCA	AAAATATTGT	CTGTGCCACG	TATTCTTACG
5051	CTTTCAGGTC	AGAAGGGTTC	TATCTCTGTT	GGCCAGAATG	TCCCTTTTAT
5101	TAAAGACTAA	TAGCCATTCA	AAAATATTGT	CTGTGCCACG	TATTCTTACG
5151	CGATTGAGCG	TCAAAATGTA	GGTATTTCCA	TGAGCGTTTT	TCCTGTTGCA

Figure 5

M13mp18 Nucleic Acid Sequence

IDT Exhibit 1020.9061

5201 ATGGCTGGCG GTAATATTGT TCTGGATATT ACCAGCAAGG CCGATAGTTT
5251 GAGTTCTCT ACTCAGGCAA GTGATGTTAT TACTAATCAA AGAAGTATTG
5301 CTACAACGGT TAATTTGCGT GATGGACAGA CTCTTTTACT CGGTGGCCTC
5351 ACTGATTATA AAAACACTTC TCAAGATTCT GGCGTACCGT TCCTGTCTAA
5401 AATCCCTTTA ATCGGCCTCC TGTTTAGCTC CCGCTCTGAT TCCAACGAGG
5451 AAAGCACGTT ATACGTGCTC GTCAAAGCAA CCATAGTACG CGCCCTGTAG
5501 CGGCGCATTA AGCGCGGCGG GTGTGGTGGT TACGCGCAGC GTGACCGCTA
5551 CACTTGCCAG CGCCCTAGCG CCCGCTCCTT TCGCTTTCTT CCCTTCCTTT
5601 CTCGCCACGT TCGCCGGCTT TCCCCGTCAA GCTCTAAATC GGGGGCTCCC
5651 TTTAGGGTTC CGATTTAGTG CTTTACGGCA CCTCGACCCC AAAAAACTTG
5701 ATTTGGGTGA TGGTTCACGT AGTGGGCCAT CGCCCTGATA GACGGTTTTT
5751 CGCCCTTTGA CGTTGGAGTC CACGTTCTTT AATAGTGGAC TCTTGTTCCA
5801 AACTGGAACA ACACTCAACC CTATCTCGGG CTATTCTTTT GATTTATAAG
5851 GGATTTTGCC GATTTCGGAA CCACCATCAA ACAGGATTTT CGCCTGCTGG
5901 GGCAAACCAG CGTGGACCGC TTGCTGCAAC TCTCTCAGGG CCAGGCGGTG
5951 AAGGGCAATC AGCTGTTGCC CGTCTCGCTG GTGAAAAGAA AAACCACCCT
6001 GGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCATTAA
6051 TGCAGCTGGC ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA
6101 CGCAATTAAT GTGAGTTAGC TCACTCATTA GGCACCCCAG GCTTTACACT
6151 TTATGCTTCC GGCTCGTATG TTGTGTGGAA TTGTGAGCGG ATAACAATTT
6201 CACACAGGAA ACAGCTATGA CCATGATTAC GAATTCGAGC TCGGTACCCG
6251 GCGATCCTCT AGAGTCGACC TGCAGGCATG CAAGCTTGGC ACTGGCCGTC
6301 GTTTTACAAC GTCGTGACTG GGAAAACCCT GGCGTTACCC AACTTAATCG
6351 CCTTGCAGCA CAATCCCCTT TCGCCAGCTG GCGTAATAGC GAAGAGGCCC
6401 GCACCGATCG CCCTTCCCAA CAGTTGCGCA GCCTGAATGG CGAATGGCGC
6451 TTTGCCTGGT TTCCGGCACC AGAAGCGGTG CCGGAAAGCT GGCTGGAGTG
6501 CGATCTTCCT GAGGCCGATA CGGTCGTCGT CCCCTCAAAC TGGCAGATGC

Figure 5

M13mp18 Nucleic Acid Sequence

IDT Exhibit 1020.9062

6551	ACGGTTACGA	TGCGCCCATC	TACACCAACG	TAACCTATCC	CATTACGGTC
6601	AATCCGCCGT	TTGTTCCCAC	GGAGAATCCG	ACGGGTTGTT	ACTCGCTCAC
6651	ATTTAATGTT	GATGAAAGCT	GGCTACAGGA	AGGCCAGACG	CGAATTATTT
6701	TTGATGGCGT	TCCTATTGGT	TAAAAAATGA	GCTGATTTAA	CAAAAATTTA
6751	ACGCGAATTT	TAACAAAATA	TTAACGTTTA	CAATTTAAAT	ATTTGCTTAT
6801	ACAATCTTCC	TGTTTTTGGG	GCTTTTCTGA	TTATCAACCG	GGGTACATAT
6851	GATTGACATG	CTAGTTTTAC	GATTACCGTT	CATCGATTCT	CTTGTTTGCT
6901	CCAGACTCTC	AGGCAATGAC	CTGATAGCCT	TTGTAGATCT	CTCAAAAATA
6951	GCTACCCTCT	CCGGCATGAA	TTTATCAGCT	AGAACGGTTG	AATATCATAT
7001	TGATGGTGAT	TTGACTGTCT	CCGGCCTTTC	TCACCCTTTT	GAATCTTTAC
7051	CTACACATTA	CTCAGGCATT	GCATTTAAAA	TATATGAGGG	TTCTAAAAAT
7101	TTTTATCCTT	GCGTTGAAAT	AAAGGCTTCT	CCCGCAAAAG	TATTACAGGG
7151	TCATAATGTT	TTTGGTACAA	CCGATTTAGC	TTTATGCTCT	GAGGCTTTAT

Figure 5

M13mp18 Nucleic Acid Sequence

IDT Exhibit 1020.9063

COMPLEMENTARY TO M_{13}

POSITION	5' * 3'	POSITION	
645	AGCAACACTATCATA	631	$M_{13}/1$
615	ACGACGATAAAAACC	601	$M_{13}/2$
585	TTTTGCAAAAGAAGT	571	$M_{13}/3$
555	AATAGTAAAATGTTT	541	$M_{13}/4$
525	CAATACTGCGGAATG	511	$M_{13}/5$
495	TGAATCCCCCTCAAA	481	$M_{13}/6$
465	AGAAAACGAGAATGA	451	$M_{13}/7$
435	CAGGTCTTTACCCTG	421	$M_{13}/8$
405	AGGAAAGCGGATTGC	391	$M_{13}/9$
375	AGGAAGCCCGAAAGA	361	$M_{13}/10$

COMPLEMENTARY TO SS PHAGE DNA

POSITION	5' * * 3'	POSITION	
351	ATATTTGAAGTCTTT	366	$M_{13}/11$
371	TCTTTTTGATGCAAT	386	$M_{13}/12$
391	CTATAATACTCAGGG	406	$M_{13}/13$
411	TGATTTATGGTCATT	426	$M_{13}/14$
431	GTTTAAAGCATTTGA	446	$M_{13}/15$
451	TATTTATGACGATTC	466	$M_{13}/16$
471	TATCCAGTCTAAACA	486	$M_{13}/17$
491	CTCTGGCAAAACTTC	506	$M_{13}/18$
511	TCGCTATTTTGGTTT	526	$M_{13}/19$
531	AAACGAGGGTTATGA	546	$M_{13}/20$

Figure 6

Primers for Nucleic Acid Production Derived from M13mp18 Sequence

IDT Exhibit 1020.9064

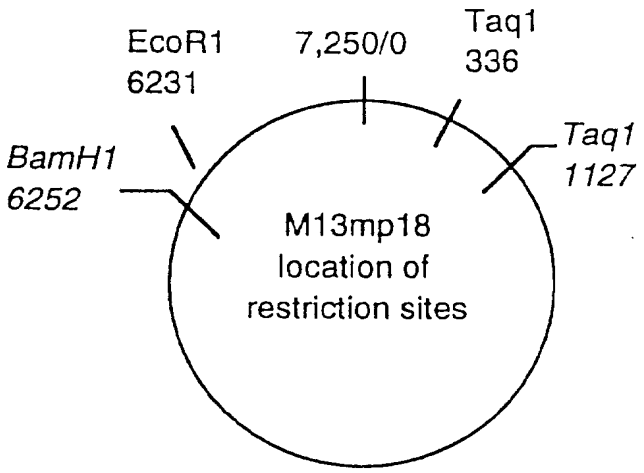
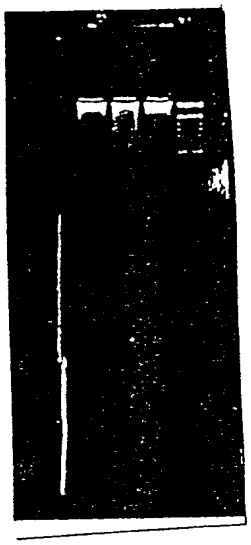


Figure 7

Appropriate M13mp18 Restriction Sites

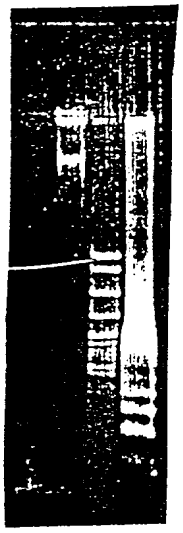
IDT Exhibit 1020.9065



Lane 1: from calf thymus + Taq digested mp18 amplification reaction
Lane 2: from Taq digested mp18 amplification reaction
Lane 3: from calf thymus amplification reaction
Lane 4: øX174 Hinf1 size marker

Figure 8

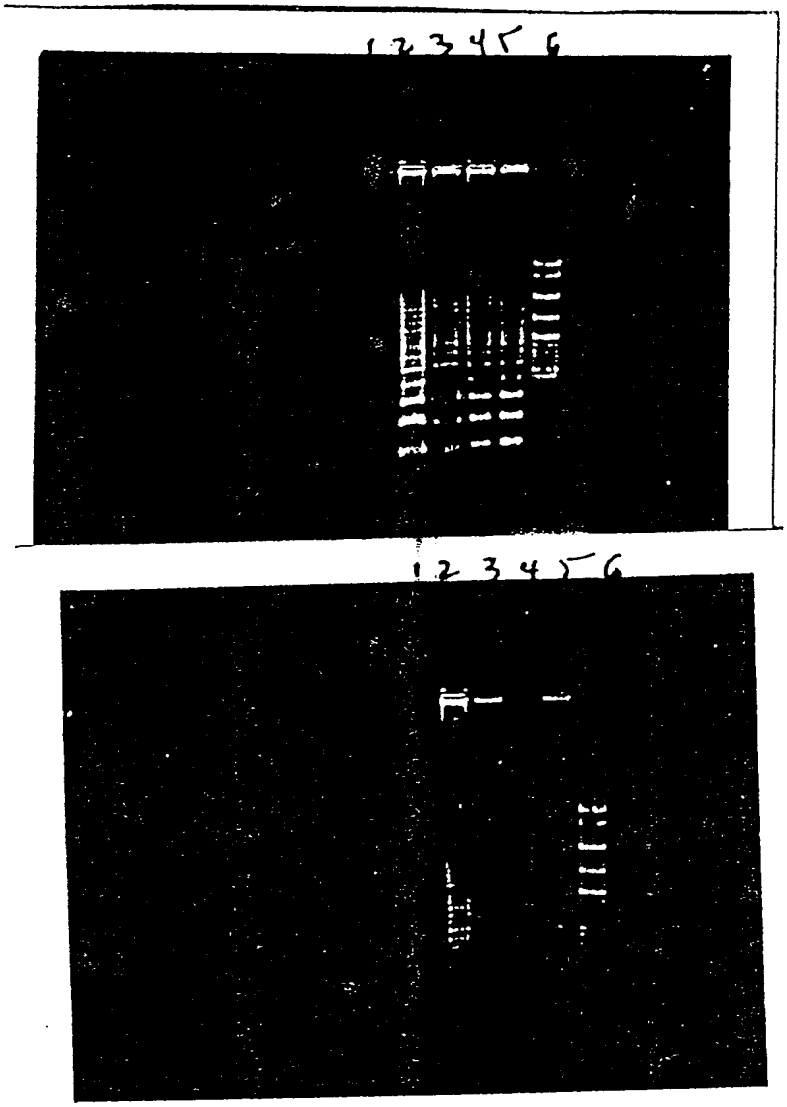
IDT Exhibit 1020.9066



Lane 1: no template
Lane 2: mp18 template, phosphate buffer
Lane 3: MspI/pBR322 size marker
Lane 4: mp18 template, MOPS buffer

Figure 9

IDT Exhibit 1020.9067

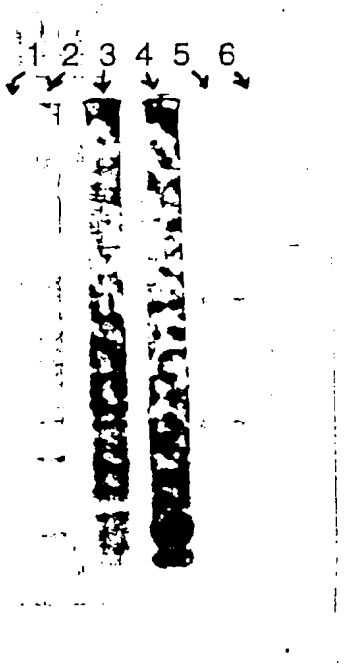


Top= (+) Template
Bottom= (-) Template

Lane 1: phosphate buffer
Lane 2: MES
Lane 3: MOPS
Lane 4: DMAB
Lane 5: DMG
Lane 6: pBR322/MspI size marker

Figure 10

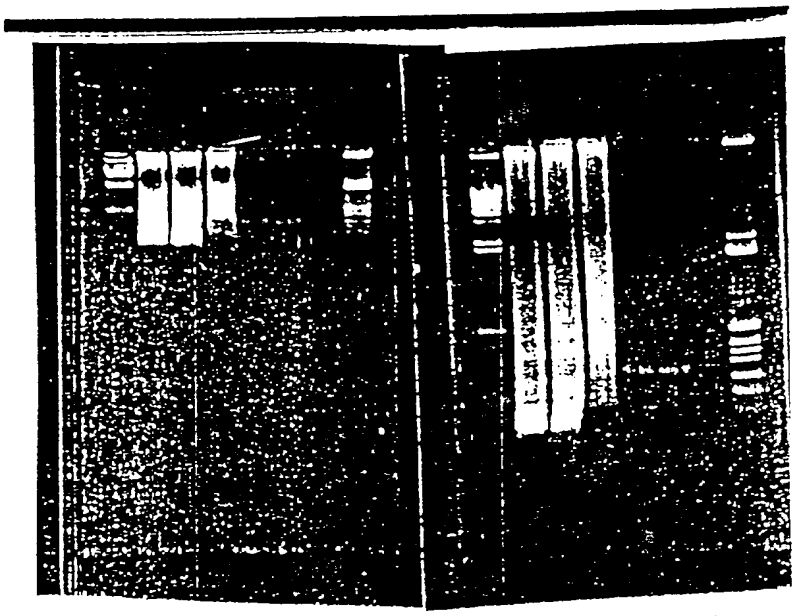
IDT Exhibit 1020.9068



Lane 1: DMAB buffer, no template
Lane 2: DMAB buffer, mp18 template
Lane 3: DMG buffer, no template
Lane 4: DMG buffer, mp18 template
Lane 5: No reaction
Lane 6: 200 ng Taq I digested mp18 size marker/positive control

Figure 11

IDT Exhibit 1020.9069



First Time Interval Second Time Interval

Agarose Gel Analysis

Lane 1: lambda Hind III marker
Lane 2: Amp/Untreated
Lane 3: Amp/Kinased
Lane 4: Amp/Kinased/Ligated
Lane 5: PCR/Untreated
Lane 6: PCR/Kinased
Lane 7: PCR/Kinased/Ligated
Lane 8: øX174/Hinf1 marker

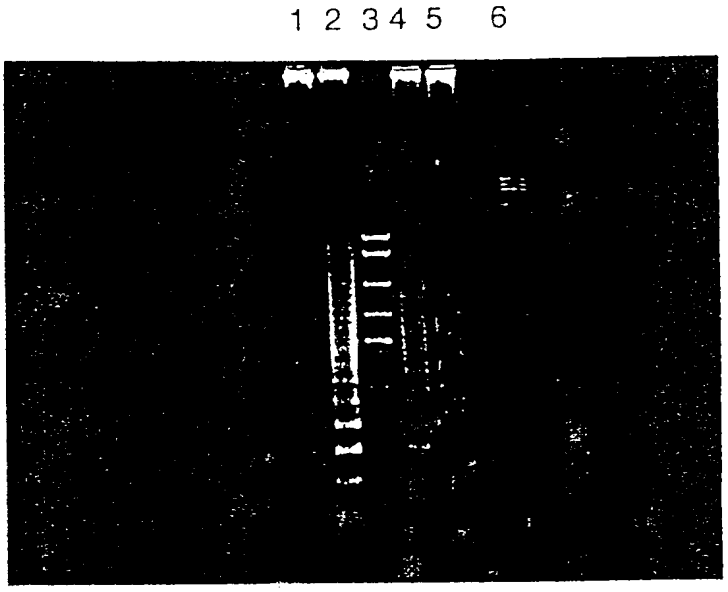
Figure 12

IDT Exhibit 1020.9070



Figure 13

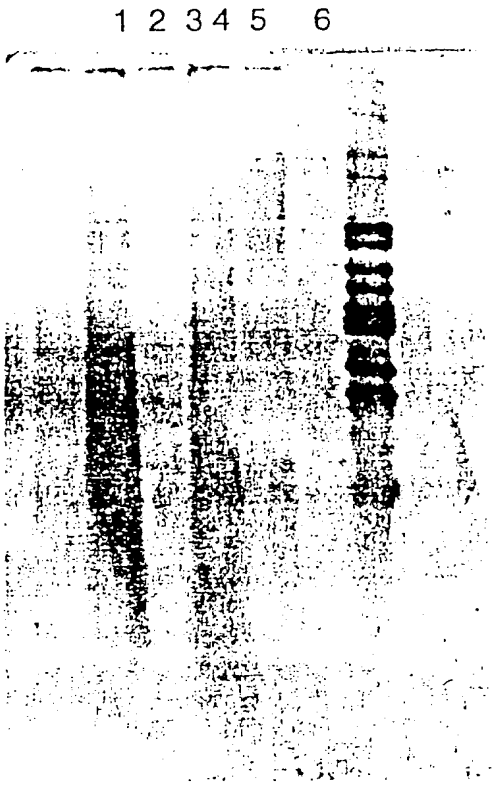
IDT Exhibit 1020.9071



Lane 1: Primers alone
Lane 2: Primers + taq digested M13 DNA
Lane 3: Molecular weight markers
Lane 4: Primers + RNA
Lane 5: Primers alone
Lane 6: M13 digested DNA
Buffer was dimethyl amino glycine, pH 8.6

Figure 14

IDT Exhibit 1020.9072



Lane 1: Primers alone
Lane 2: Primers + taq digested M13 DNA
Lane 3: Molecular weight markers
Lane 4: Primers + RNA
Lane 5: Primers alone
Lane 6: M13 digested DNA
Buffer was dimethyl amino glycine, pH 8.6

Figure 15

IDT Exhibit 1020.9073

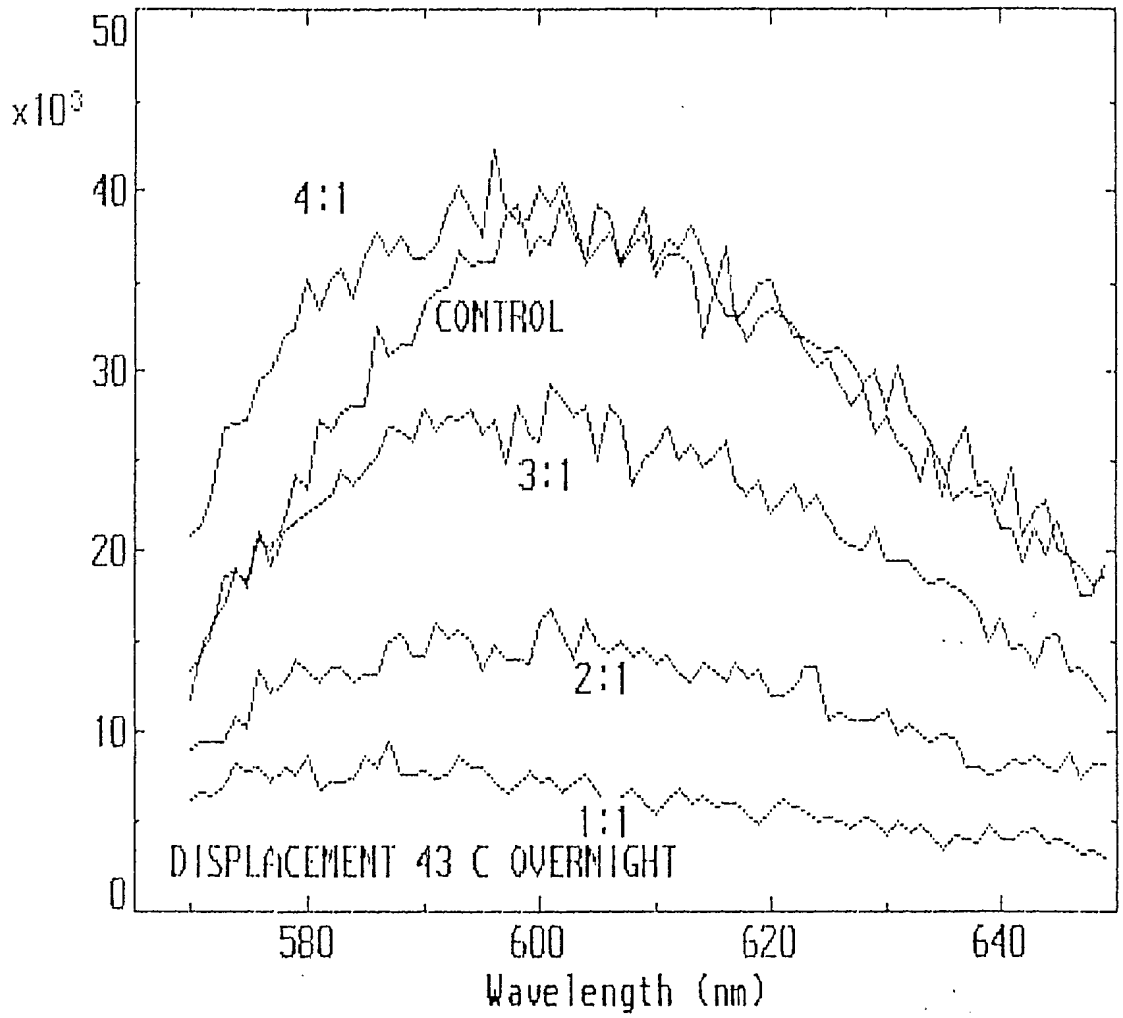


Figure 16

IDT Exhibit 1020.9074

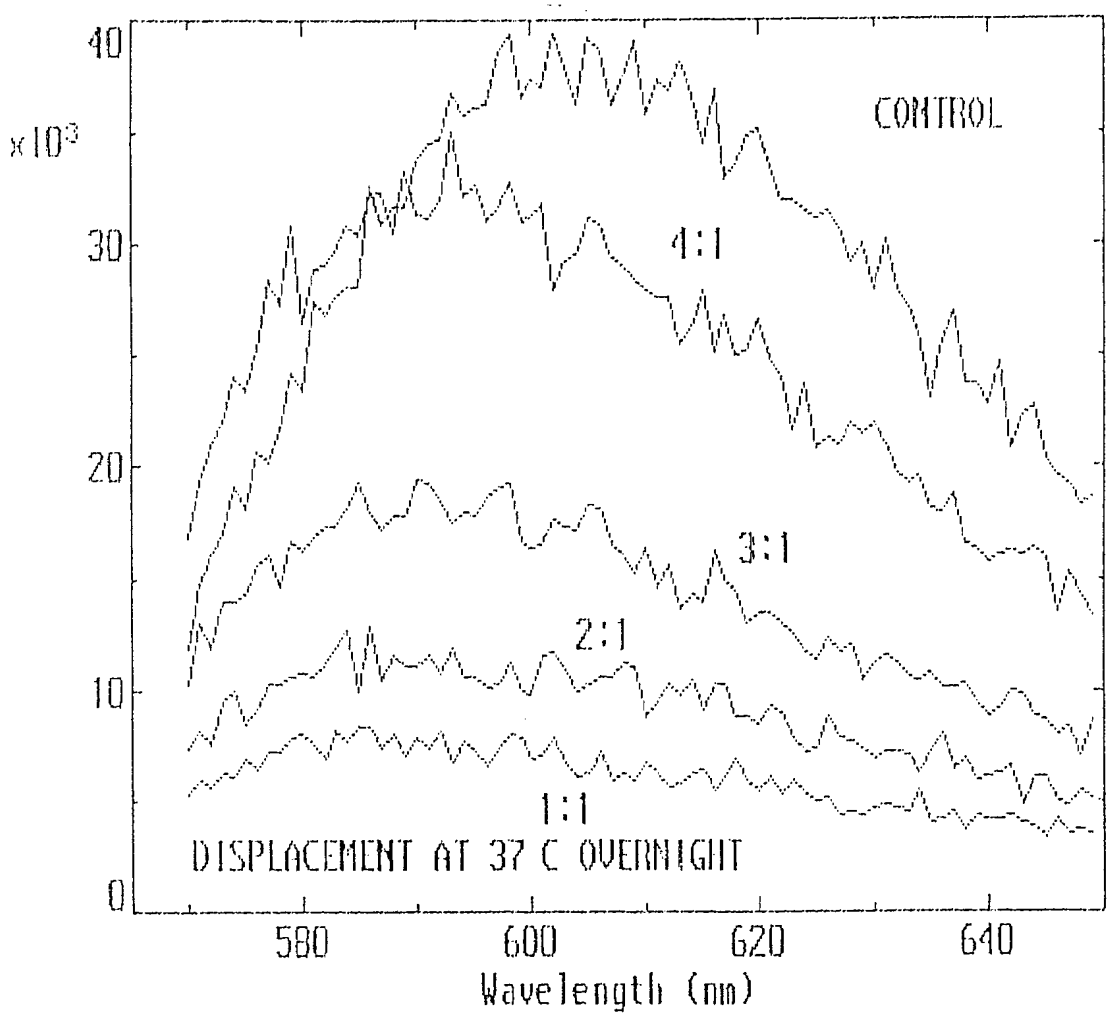


Figure 17

IDT Exhibit 1020.9075

pIBI 31-BH5-2

fmet AUG of Lac z {T7 Promotor region····
LAC PROMOTOR..ATG ACC ATG ATT ACG CCA GAT ATC AAA TTA ATA CGA CTC ACT ATA

oligo 50-mer 3'- tac t*aa t*gc ggt* ct*a t*ag t*Vt aat* tat* gct* gag t*ga t*at* c-5'
10 base insert

T7 RNA Start {«« T3 Promotor Region }
IGGG CTC ICCT TTA GTG ACG GTT AAT
···»»} «- T3 Start Signal

pIBI 31 BSII/HCV

fmet AUG of Lac z {T3 Promotor region -»} T3 RNA Start
LAC PROMOTOR .ATG ACC ATG ATT ACG CCA AGC TCG AAA TTA ACC CTC ACT AAA /GGG
oligo 50-mer 3'- tac t*aa t*ac t*aa t*gc ggt* t*V--10 base insert--.....

{«- T7 Promotor Region }
MULTIPLE CLONING SITE + 390 BASE INSERT CTA /TAG TGA GTC CGT ATT AAT....
«- T7 Start Signal
5'-ct*a t*ag t*ga gt*c gt*a tt*a at*.....

Figure 18

IDT Exhibit 1020.9076

Espacenet

Bibliographic data: WO0055364 (A2) — 2000-09-21

GENETIC ANALYSIS

Inventor(s): REEVE MICHAEL ALAN [GB]; WORKMAN NICHOLAS IAN [GB]; MARTIN-PARRAS LUIS [GB] ± (REEVE, MICHAEL, ALAN, ; WORKMAN, NICHOLAS, IAN, ; MARTIN-PARRAS, LUIS)

Applicant(s): AMERSHAM PHARM BIOTECH UK LTD [GB]; REEVE MICHAEL ALAN [GB]; WORKMAN NICHOLAS IAN [GB]; MARTIN PARRAS LUIS [GB] ± (AMERSHAM PHARMACIA BIOTECH UK LIMITED, ; REEVE, MICHAEL, ALAN, ; WORKMAN, NICHOLAS, IAN, ; MARTIN-PARRAS, LUIS)

Classification: **- international:** ***C12N15/09; C12Q1/68; G01N1/28; G01N33/53; G01N33/566; G01N37/00;*** (IPC1-7): C12Q1/68
- cooperative: **C12Q1/683**; **C12Q1/6837**

Application number: WO2000GB00916 20000310 Global Dossier

Priority number(s): EP19990301933 19990312

Also published as: WO0055364 (A3) AU3178200 (A) CA2362771 (A1) EP1173609 (A2) JP2002538837 (A)

Abstract of WO0055364 (A2)

A method is described for use in whole genome analysis. The method - termed inter-population perfectly matched duplex depletion - can overcome many of the limitations of current approaches based upon SNPs and linkage disequilibrium within isolated populations. Inter-population perfectly matched duplex depletion isolates a fragment (or fragments) containing differences between the "affected" and "unaffected" populations or cells. A convenient method - terminal restriction site profiling arrays (TRSPAs) - is described for the analysis of such fragments. A totally diagnostic internal control DNA is also described which allows both the extent and exact nature of any partial digestion to be unambiguously determined for inter-population perfectly matched duplex depletion or TRSPA restriction.

IDT Exhibit 1020.9077

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 00/55364**

(43) International Publication Date: 21 September 2000 (21.09.00)

(21) International Application Number: PCT/GB00/00916

(22) International Filing Date: 10 March 2000 (10.03.00)

(30) Priority Data:
99301933.0 12 March 1999 (12.03.99) EP

(71) Applicant *(for all designated States except US)*: AMERSHAM PHARMACIA BIOTECH UK LIMITED [GB/GB]; Amersham Place, Little Chalfont, Buckinghamshire HP7 9NA (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: REEVE, Michael, Alan [GB/GB]; 118 St Andrews Road, Henley–on–Thames, Oxfordshire RG9 1PL (GB). WORKMAN, Nicholas, Ian [GB/GB]; 4 Canford Court, Aylesbury, Buckinghamshire HP21 7NZ (GB). MARTIN–PARRAS, Luis [ES/GB]; 7 Gladstone Rise, High Wycombe, Buckinghamshire HP13 7NW (GB).

(74) Agents: PENNANT, Pyers et al.; Stevens Hewlett & Perkins, Halton House, 20/23 Holborn, London EC1N 2JD (GB).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, 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, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, 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), 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: GENETIC ANALYSIS

(57) Abstract

A method is described for use in whole genome analysis. The method – termed inter–population perfectly matched duplex depletion – can overcome many of the limitations of current approaches based upon SNPs and linkage disequilibrium within isolated populations. Inter–population perfectly matched duplex depletion isolates a fragment (or fragments) containing differences between the "affected" and "unaffected" populations or cells. A convenient method – terminal restriction site profiling arrays (TRSPAs) – is described for the analysis of such fragments. A totally diagnostic internal control DNA is also described which allows both the extent and exact nature of any partial digestion to be unambiguously determined for inter–population perfectly matched duplex depletion or TRSPA restriction.

IDT Exhibit 1020.9078

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 1020.9079

GENETIC ANALYSIS

INTRODUCTION

Limitations of the current approaches

There are a number of limitations to carrying out association studies using single nucleotide polymorphisms (SNPs) and linkage disequilibrium within human populations (see Science, Vol 278, p1580, (1997) for a review of such methods). We have no control over recombination frequency around a given locus or over past human genetic crossing. Some mutations will be closely correlated with nearby SNPs and others will not.

The need for whole genome analysis

With the SNP and linkage disequilibrium approach (and many others), markers are essentially used as a surrogate for sequencing - the more markers, the better. The logical endpoint of the above argument is to look at every base in the human genome - and carry out what could be termed a whole genome association study. In essence, the sequence at every base would be determined for the genome of each member of a phenotypically 'affected' and a phenotypically 'unaffected' population. Statistical correlations (associations) could then be drawn between sequence differences and phenotype. Such associations would have future predictive values for the phenotype of interest, knowing the genotype and could be of great value in medicine and pharmacogenetics.

The current invention selectively enriches for DNA fragments that determine phenotype in the 'affected' population and thus makes the prospect of carrying out whole genome association studies for humans and other species a very real possibility.

Definition of terms used with the current invention

Within the scope of the current invention, the individuals chosen for whole genome analysis may be human, animal or plant and they may be eukaryotic, prokaryotic or archaebacterial in origin.

The terms 'affected' and 'unaffected' are used without limitation in order to categorise individuals into two groups – those that possess a defined phenotype of interest ('affected' individuals) and those that do not possess the phenotype of interest ('unaffected' individuals). The phenotype common to the 'affected' individuals may be either beneficial (e.g. for these individuals, a particular pharmaceutical entity might show high efficacy in a phase II clinical trial) or detrimental (e.g. for these individuals, a particular pharmaceutical entity might show adverse toxicology in a phase I clinical trial).

The 'affected' population may comprise one or more individuals and the 'unaffected' population may similarly comprise one or more individuals according to the particular embodiment of the invention (see below).

The term DNA is used throughout for simplicity. Within the scope of the current invention, the term DNA may equally well apply to all or part of the haploid, diploid or polyploid genomic DNA content of one or more germ line or somatic cell(s). The DNA may be extracted from cells taken directly from the individual(s), the DNA may be extracted from cells cultured or immortalised from the individual(s) or the DNA may be prepared from a library of clones - with inserts derived from the individual(s) and propagated in some appropriate host and cloning vector system. For the particular case wherein the term DNA refers to the expressed part of the haploid, diploid or polyploid genomic DNA content of one or more somatic cells and the DNA is prepared from a library of clones - with inserts derived from the individual(s) and propagated in some appropriate host and cloning vector system, a cDNA library (normalised or otherwise) may be used.

In the current invention, DNA is compared in fragmented form. Fragmentation can be performed after DNA extraction, prior to cloning and/or after cloning. Restriction enzyme digestion is the preferred method for such fragmentation - though other methods (e.g. shearing or sonication) will be obvious to those skilled in the art.

For the particular case wherein the DNA is prepared from a library of clones (either genomic clones or cDNA clones) - with inserts derived from the individual(s) and propagated in some appropriate host and cloning vector system and wherein restriction enzyme fragmentation is used prior to cloning, polymerase chain reaction amplification can be used to prepare the DNA for comparison in fragmented form. Priming sites within the vector sequence flanking the cloned restriction enzyme fragmented inserts may be usefully employed for one or more cycles of polymerase chain reaction amplification of the fragmented DNA of interest. The primers used for polymerase chain reaction amplification of the fragmented DNA of interest could again be used after the phenotype-determining fragment enrichment process to 'rescue' and clone the enriched fragments.

Within the scope of the current invention, the terms biotinylation and streptavidin capture are used both as an example and as the currently preferred embodiment for the invention. The streptavidin may be surface attached to inert particles (magnetic or otherwise) or to vessel walls (e.g. microtitre plate wells). The biotin may be introduced via a deoxynucleotide triphosphate analogue using a polymerase; by using a biotin-conjugated primer and polymerase chain reaction amplification; chemically or photochemically. The use of biotin and streptavidin is not a limitation for the invention. The invention could equally be used with other high affinity capture systems well known to those skilled in the art (e.g. 'his tag' introduction and metal ion affinity capture).

Within the scope of the current invention, the term 'abnormal' - used with respect to the term 'normal' - is used without limitation in order

to denote a somatic cell (or somatic cells) with a discernable phenotypic characteristic (or characteristics) arising from the acquisition of a different somatic mutation (or set of somatic mutations) from that (or those) seen in the 'normal' counterpart. Cells will most usually be considered 'abnormal' with respect to their 'normal' counterparts through the acquisition of a different somatic mutation (or set of somatic mutations) leading to one or more of the following phenotypic characteristics: altered marker gene expression, altered genomic organisation, growth under certain selective culture conditions, immortalised growth in culture, unrestrained growth *in vivo* or *in vitro*, failure of normal apoptotic control mechanisms *in vivo* or *in vitro*, induction of neovascularisation, escape of cells across epithelium, migratory cell survival or metastasis.

Within the scope of the current invention, the term mismatch recognition protein is used without limitation to denote a protein of eukaryotic, prokaryotic or archaebacterial origin capable of the selective recognition of (and binding to) a DNA duplex that is not perfectly matched along its entire length. Recognition of (and binding to) will be preferably for bases that are not engaged in correct Watson and Crick pairing and for small deletions or insertions. Many such proteins are known to those skilled in the art. Prokaryotic and eukaryotic mutS homologues, phage T4 endonuclease VII, phage T7 endonuclease I and the plant enzyme CEL-1 are just some examples.

Inter-population perfectly matched duplex depletion

In the inter-population perfectly matched duplex depletion approach, we compare (in fragmented form) the pooled DNA of 'affected' individuals with the pooled DNA of 'unaffected' individuals (both from populations as outbred and otherwise similar to each other as possible). We are only interested in those regions where differences occur between 'affected' and 'unaffected' DNA molecules. For populations as above, the only prevalent sequence differences within the 'affected' pool (compared to

the 'unaffected' pool) should be somewhere within the gene(s) (using the term gene in its widest sense to include exons, introns and all associated upstream and downstream regulatory sequences) that actually determine(s) their common phenotype. This means that we are no longer tied into working with rare (and perhaps atypical) populations where there is high genetic homogeneity.

Pooling the DNA from entire phenotypically-defined populations massively reduces the amount of labour involved.

DNA sequence variation in populations

Nickerson *et al*, Nature Genetics, Vol 19, p233 (1998), sequenced 9.7 kb of the lipoprotein lipase gene from 71 individuals (24 African-Americans, 24 Europeans and 23 European-Americans). This gene is fairly typical (90 % intron and 10 % exon - total size 30 kb with 10 exons).

88 sequence variants were found (i.e. one per 110 bp on average). Most variations were found in non-coding sequence. 90 % of these were SNPs (60 % of which were transitions and 40 % were transversions). All of the SNPs were biallelic. 10 % of the sequence variants were insertions or deletions at repeat sequences.

58 % of the sequence variants were present in all three ethnic populations. Half of these were found in heterozygous form and half in homozygous form.

Nucleotide diversity (defined as the expected number of nucleotide differences per site between a random pair of chromosomes drawn from the population) is 1/500 for DNA in general and 1/2,000 for coding sequence DNA. This means that, on average, any two DNA fragments annealed together from such a population will contain a mismatch every 500 bases.

DNA sequence variants are therefore very common. They are not, however, totally random - the variants that occur every 500 bases or so are limited; they are generally biallelic at just that single base. It is

this fact that the inter-population perfectly matched duplex depletion approach selectively exploits.

Statistical analysis of the inter-population perfectly matched duplex depletion fragmentation process

If the length of the DNA fragments is F bases and the average length between sequence differences between any two DNA molecules is 500 bases, the probability that a hybrid duplex between any two random DNA fragments will contain no mismatches is given by

$$Pr(0) = e^{-(F/500)}$$

and the probability that a hybrid duplex between any two random DNA molecules will contain any mismatches is given by

$$Pr(\geq 1) = \{1-(e^{-(F/500)})\}$$

Example values for Pr(0) and Pr(≥1) for different values of F are given below

F	Pr(0)	Pr(≥1)
10	0.98	0.02
20	0.96	0.04
50	0.90	0.10
100	0.82	0.18
200	0.67	0.33
300	0.55	0.45
400	0.45	0.55
500	0.37	0.63
600	0.30	0.70
700	0.25	0.75
800	0.20	0.80
900	0.17	0.83
1000	0.14	0.86

Example average restriction fragment sizes for DNA digestion with six 6 bp cutters and up to four 4 bp cutters

Each 6 bp cutter will cut DNA every 4,096 bp on average and each 4 bp cutter will cut DNA every 256 bp on average.

For a given set of A 6 bp cutters and B 4 bp cutters (with no duplication of restriction enzyme cutting sequence), the average fragment length F will be

$$\{(A/4,096)+(B/256)\}^{-1}$$

Example values for F as A and B are varied are given in the following table

A	B	average size (F)
6	0	683
6	1	186
6	2	108
6	3	76
6	4	59

We should note that the above situation assumes that none of the 4 bp cutter recognition sites lie within any of the 6 bp cutter recognition sites. If, for example, we have a 4 bp cutter recognition site nested within a 6 bp cutter recognition site (e.g. from the use of *MboI* and *BamHI* in the fragmentation), then we should reduce the value of A from 6 to 5 in the above.

In general, if we have a given set of A enzymes that cut DNA every a bp on average, B enzymes that cut DNA every b bp on average, ... , Z enzymes that cut DNA every z bp on average (with no duplication of restriction enzyme cutting sequence), the average fragment length F will be

$$\{(A/a)+(B/b)+ \ldots + (Z/z)\}^{-1}$$

Example sets of 6 bp cutters and 4 bp cutters for fragmentation

Example sets of 6 bp cutters and 4 bp cutters that contain panels of six 6 bp cutters that are compatible with terminal restriction site profiling array (TRSPA) analysis (see below) are given in the following:

Example enzyme set 1

For restriction in M buffer + BSA

Number	Enzyme	Site	M+BSA % activity	Optimum °C	U / µl	Supplier	Inactivate °C/min
1	*MboI*	GATC	100	37	25	NEB	65/20
2	*HaeIII*	GGCC	100	37	50	NEB	80/20
3	*MseI*	TTAA	100	37	20	NEB	65/20

Number	Enzyme	Site	M+BSA % activity	Optimum °C	U / µl	Supplier	Inactivate °C/min
1	*BamHI*	GGATCC	100	37	100	NEB	80 / 20
2	*BsrGI*	TGTACA	100	37	10	NEB	80 / 20
3	*HindIII*	AAGCTT	100	37	>40	APB	65 / 20
4	*NcoI*	CCATGG	100	37	50	NEB	65 / 20
5	*SpeI*	ACTAGT	100	37	50	NEB	65 / 20
6	*AflII*	CTTAAG	100	37	10	APB	60 / 15

Example enzyme set 2

For restriction in M buffer + BSA

Number	Enzyme	Site	M+BSA % activity	Optimum °C	U / µl	Supplier	Inactivate °C/min
1	*MboI*	GATC	100	37	25	NEB	65/20
2	*HaeIII*	GGCC	100	37	50	NEB	80/20
3	*MseI*	TTAA	100	37	20	NEB	65/20

Number	Enzyme	Site	M+BSA % activity	Optimum °C	U / µl	Supplier	Inactivate °C/min
1	*EcoRI*	GAATTC	100	37	>40	APB	65 / 20
2	*BspHI*	TCATGA	100	37	10	APB	65 / 20
3	*BglII*	AGATCT	75	37	>40	APB	No
4	*XbaI*	TCTAGA	100	37	100	NEB	65 / 20
5	*Acc65I*	GGTACC	75	37	10	NEB	65 / 20
6	*ApaLI*	GTGCAC	100	37	10	NEB	No

Aspect (I)

In one aspect the invention provides a method of providing a mixture of DNA fragments enriched in fragments that are characteristic of a phenotype of interest, by providing affected DNA in fragmented form and providing unaffected DNA in fragmented form, which method comprises:

a) mixing the fragments of the affected DNA and the fragments of the unaffected DNA under hybridising conditions;

b) recovering a mixture of hybrids that contain mismatches;

c) recovering fragments of the affected DNA from the mixture of hybrids that contain mismatches;

and optionally repeating steps a), b) and c) one or more times.

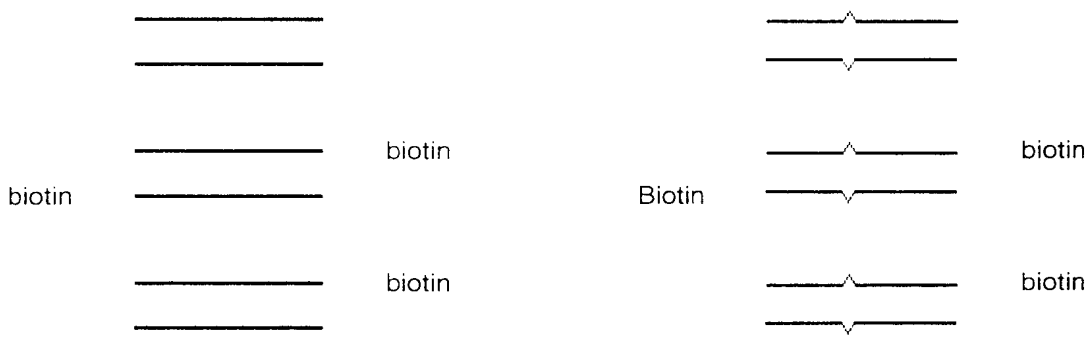
Inter-population mismatch-containing duplex selection

'Affected' versus 'unaffected' (i.e. inter-population) mismatch-containing duplex selection can be achieved by attaching a mismatch-binding protein to a solid support (or using the mismatch-binding protein in solution followed by subsequent solid-phase capture), taking fragmented and denatured 'affected' DNA and hybridising this to an excess of fragmented, denatured and biotinylated 'unaffected' DNA with ensuing capture of mismatch-containing duplex molecules. Releasing the mismatch-containing duplex molecules without denaturation, streptavidin

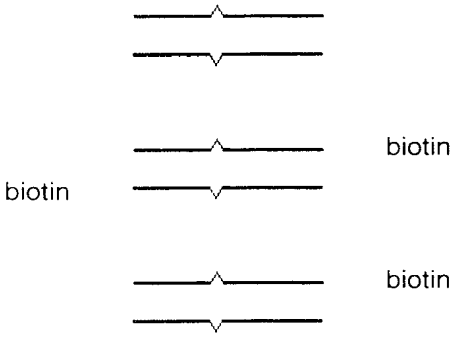
capture and then release of the non-biotinylated strands will give only the desired species as shown below.

Method

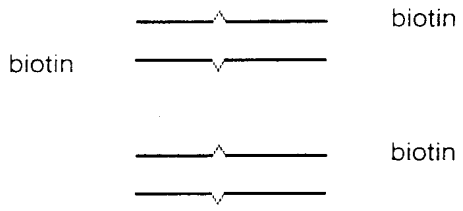
Fragment the 'affected' DNA. Fragment and derivatise the 'unaffected' DNA with biotin. Only DNA from this population will be streptavidin captured. Melt and anneal to give



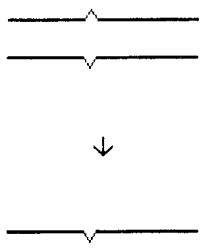
Mismatch-binding protein select. Capture only the mismatch-containing duplexes. Release without denaturation to give



Streptavidin capture to give



Release the non-biotinylated strands to give



Repeat as necessary.

Repetition of the above sequence of reactions will lead to inter-population perfectly matched duplex depletion.

What will be purified ?

Inter-population mismatch-containing duplex selection as above ensures that all of the various phenotype-determining fragments (unique to the 'affected' population) are captured for subsequent analysis - but it also causes the co-purification of very many SNP-containing ('noise') fragments.

We now need to consider the fate of the various types of fragment (i.e. those that determine the phenotype and those that do not) as we carry out inter-population perfectly matched duplex depletion cycles as above.

Recovery of 'affected' DNA molecules after streptavidin capture

For a particular fragment, if we have X molecules of 'affected' DNA and Y molecules of 'unaffected' biotinylated DNA, after complete hybridisation, there will be a ratio of {Y/(X+Y)} streptavidin-capturable molecules to {X/(X+Y)} streptavidin-non-capturable molecules. We can thus manipulate the yield of streptavidin-capturable hybrids by varying X and Y.

Example recovery and loss figures for various X and Y are shown in the following table

ratio (Y/X)	recovery = {Y/(X+Y)}	Loss = {X/(X+Y)}
1	50 %	50 %
2	67 %	33 %
3	75 %	25 %
4	80 %	20 %
9	90 %	10 %
19	95 %	5 %
99	99%	1 %
999	99.9 %	0.1 %

After n cycles, the recovery for a phenotype-determining fragment will be given by $\{Y/(X+Y)\}^n$

Loss of general SNP-containing ('noise') fragments during inter-population perfectly matched duplex depletion cycles

If we anneal fragmented DNA molecules together and capture only the mismatch-containing duplexes, then n repetitions of such a process will reduce the original number of fragments to the following fraction

$$\{1-(e^{-(F/500)})\}^n \cdot \{Y/(X+Y)\}^n$$

The enrichment for phenotype-determining fragments over SNP-containing ('noise') fragments during inter-population perfectly matched duplex depletion cycles will therefore be given by

$$\{1-(e^{-(F/500)})\}^{-n}$$

Example figures for enrichment are given below with F = 50, 100, 200. 300, 400, 500, 600, 700, 800, 900 and 1,000 and n = 1, 2, 3, 4, 5, 6 ,7, 8, 9 and 10.

	n =									
F	1	2	3	4	5	6	7	8	9	10
50	11	110	1,160	12,194	128,135	1,346,489	1.E+07	1.E+08	2.E+09	2.E+10
100	6	30	168	926	5,110	28,187	155,500	857,841	4,732,415	3.E+07
200	3	9	28	85	257	779	2,362	7,166	21,735	65,929
300	2	5	11	24	53	119	263	582	1,291	2,860
400	2	3	6	11	20	36	65	118	215	390
500	2	3	4	6	10	16	25	39	62	98
600	1	2	3	4	6	9	12	18	25	36
700	1	2	2	3	4	5	7	10	13	17
800	1	2	2	2	3	4	5	6	8	10
900	1	1	2	2	2	3	4	4	5	6
1,000	1	1	2	2	2	2	3	3	4	4

Loss of specific SNP-containing ('noise') fragments during inter-population perfectly matched duplex depletion cycles

The non-polymorphic and SNP-containing ('noise') fragments will be depleted as described above.

Not all fragments will, however, be depleted at the same rate. An individual SNP-containing ('noise') fragment will be depleted with every cycle of inter-population perfectly matched duplex depletion as outlined below.

Let us assume that there are two alleles for the polymorphism within a particular fragment. Let these be called P and Q. Let p be the

fraction of the P allele in the (outbred) 'affected' and 'unaffected' populations and let q be the fraction of the Q allele. Let (p+q)=1, so that q=(1-p).

After denaturation and annealing, the four possible events are PP, PQ, QP and QQ. PP and QQ will form perfectly matched duplexes and will therefore be lost - whereas PQ and QP will form mismatch-containing duplexes and will consequently be recovered.

After one cycle, the fraction of recovered molecules will be

$$\{2pq/(p^2+2pq+q^2)\}$$

$$= \{2pq/((p+q)^2)\}$$

$$= 2pq$$

$$= 2p(1-p)$$

Hence if we start out with M molecules of DNA from the population, there will be 2Mpq molecules remaining after the first round of inter-population mismatch-containing duplex selection. Let us denote the number of molecules entering the second round of inter-population mismatch-containing duplex selection by M', where

$$M' = 2Mpq$$

and the fraction of lost molecules will be

$$\{(p^2+q^2)/(p^2+2pq+q^2)\}$$

$$= \{(p^2+q^2)/((p+q)^2)\}$$

$$= p^2+q^2$$

$$= p^2+\{1-2p+p^2\}$$

$$= 1-2p+2p^2$$

$$= 1-2p(1-p)$$

$$=1-2pq$$

The fractional loss of P-allelic molecules will be p^2 and the fractional loss of Q-allelic molecules will be q^2 (= $(1-p)^2$).

If we start out with M molecules of DNA from the population, there will be Mp of the P-allelic molecules and Mq = M(1-p) of the Q-allelic molecules before inter-population mismatch-containing duplex selection. After inter-population mismatch-containing duplex selection, there will therefore be $Mp-Mp^2 = Mp(1-p) = Mpq$ of the P-allelic molecules and $Mq-Mq^2 = Mq(1-q) = Mqp$ of the Q-allelic molecules. In other words, after the first round of mismatch-containing duplex selection, there will be the same number of P-allelic molecules as Q-allelic molecules.

We can define new allelic frequencies p' and q' as follows

$$p' = q' = 0.5$$

If we now perform a second round of inter-population mismatch-containing duplex selection, we start out with M' molecules of DNA from the first round. There will be M'p' of the P-allelic molecules and M'q' of the Q-allelic molecules before inter-population mismatch-containing duplex selection.

After inter-population mismatch-containing duplex selection, there will therefore be $M'p'-Mpp' = M'p'(1-p) = M'p'q$ of the P-allelic

molecules and M'q'-Mqq' = M'q'(1-q) = M'q'p of the Q-allelic molecules.

The total number of molecules (M") will be

$$M'p'q + M'q'p$$

$$= M'(p'q-q'p)$$

but $p' = q' = 0.5$

hence $M" = 0.5*M'(p+q)$

but $(p+q) = 1$

so $M" = (M'/2)$

In other words, after the second round of inter-population mismatch-containing duplex selection, there will again be the same number of P-allelic molecules as Q-allelic molecules. Thus the new allelic frequencies p" and q" remain as previously p" = q" = 0.5.

A pattern now emerges. After the first cycle of inter-population mismatch-containing duplex selection, the allelic frequencies are equalled and the number of molecules is reduced to 2Mpq (Mpq of each allelic molecule). Thereafter, every cycle halves the number of molecules and keeps both alleles at the same frequency.

Consequently, after n cycles, the number of P-allelic molecules will be

$$Mpq(0.5)^{n-1}$$

and the number of Q allelic molecules will be

$$Mqp(0.5)^{n-1}$$

both are, of course, equal.

If we now take the capture yield (see above) into consideration, the SNP-containing ('noise') fragment yield will be given by

$$2Mpq(0.5)^{n-1}\cdot\{Y/(X+Y)\}^{n}$$

where both allelic variants are deemed to be captured 'noise'.

Polymorphisms that interfere with the pattern of restriction digestion

For both the loss of general and specific SNP-containing ('noise') fragments during inter-population perfectly matched duplex depletion cycles (described above) and where the SNP interferes with the pattern of restriction digestion, if the mismatch-binding protein also binds to duplex molecules with unequal lengths (e.g. from inter-population annealing around a site of restriction site polymorphism), then the above analysis still holds (with perfectly matched duplex being replaced by equal length duplex and mismatch-containing duplex being replaced by unequal partner-length duplex).

In the rare cases where a restriction site is lost due to a sequence change that actually determines the phenotype of interest, a double-length fragment will be obtained. This will give rise to a double terminal restriction site profiling array (TRSPA) signature (see below). Multiple isolates of the particular double signature will be indicative of an association between the fragment and the phenotype of interest.

Further 'kinetic' enrichment to enhance the selective removal of SNP-containing 'noise' from the pool of phenotype determining fragments

After multiple cycles of enrichment by the above procedure, the enriched DNA pool should contain many copies of all phenotype-determining fragments but also low numbers of copies of many different phenotype non-determining fragments. The total number of 'noise' fragments may exceed the number of phenotype determining fragments, despite the number of each individual 'noise' species being very small. The 'noise' fragments would therefore increase the number of probes required for TRSPA analysis before a pattern emerges. To largely eliminate this problem, a further kinetic enrichment procedure is used. Either one or both of strategies A and B below can be employed to achieve 'kinetic' enrichment.

Strategy A – Subtraction of the enriched DNA from inter-population mismatch containing duplex depletion

The enriched fragment pool from inter-population mismatch containing duplex depletion is rapidly self-hybridised - enabling the common phenotype-determining fragments to form perfectly matched duplexes with greater efficiency than the rare 'noise' fragments. Selection for perfectly matched duplexes then yields a selectively further enriched pool of fragments. Multiple cycles of subtraction could be carried out if necessary.

Strategy B – Hybridisation of the enriched DNA from inter-population mismatch containing duplex depletion against the 'affected' DNA pool

The enriched fragment pool from inter-population mismatch containing duplex depletion is then hybridised to an excess of biotinylated DNA from the 'affected' pool. This allows the common phenotype-determining fragments to form perfectly matched duplexes with greater efficiency than the rare 'noise' fragments. Selection for perfectly matched

duplexes followed by streptavidin capture and denaturation to release the non-biotinylated strands then yields a further enriched pool of fragments. Multiple such 'affected' pool back-hybridisations could be carried out if necessary.

Extension of the invention to the case of single phenotypically 'affected' individuals within populations where the distinction between 'affected' and 'unaffected' is clear

The above has described inter-population perfectly matched duplex depletion between non-biotinylated DNA fragments from an 'affected' population and biotinylated DNA fragments from an 'unaffected' population. Provided the 'unaffected' population is sufficiently complex that it contains all the non-phenotype-determining sequence variants found in a single 'affected' individual, then inter-population perfectly matched duplex depletion should be possible for single phenotypically 'affected' individuals against an 'unaffected' population where the distinction between 'affected' and 'unaffected' is clear. The latter proviso is needed in order to ensure that a small number of misdiagnosed 'affected' individuals in the 'unaffected' population do not cause the removal of phenotype-determining fragments during inter-population perfectly matched duplex molecular depletion.

Extension of the invention to the case of disease gene identification in cases where novel phenotype-determining mutations arise spontaneously within a family

Except for a small number of sequence changes, each of us contains DNA sequence derived from our parents - our individuality resulting from precisely which parental alleles we receive. If one of the above small number of sequence changes results in a change in phenotype, then we can use inter-population perfectly matched duplex depletion to enrich for fragments encoding this change in phenotype.

If we take 'unaffected total ancestral' cells (by which we mean cells derived from a complete set of 'unaffected' ancestors - e.g. both parents, or mother plus two paternal grandparents, or father plus two maternal grandparents, or two maternal grandparents and two paternal grandparents etc.) as the source of our biotinylated fragments and cells from an 'affected' descendent as the source of our non-biotinylated fragments, any fragments that have acquired phenotype-determining sequence changes between 'unaffected' ancestral generations and the 'affected' descendent generation will be unable to form perfectly matched duplexes with the biotinylated 'unaffected total ancestral' fragments. Successive cycles of such inter-population perfectly matched duplex depletion will thus lead to the enrichment of fragments carrying all such sequence - the degree of enrichment per cycle being as described below.

Statistical analysis

Let us assume that equal numbers of fragments are used from each of the 'unaffected' ancestors. Let the number of such ancestors be A.

{1/A} of the annealings in the inter-population perfectly matched duplex depletion will be self-against-ancestral-transmitted alleles - statistically equivalent to self-against-self inter-population perfectly matched duplex depletion (see below).

{(A-1)/A} of the annealings in the inter-population perfectly matched duplex depletion will be self-against-ancestral-nontransmitted alleles - statistically equivalent to inter-population perfectly matched duplex depletion between unrelated individuals.

A	{1/A}	{(A-1)/A}
2	1/2	½
3	1/3	2/3
4	1/4	¾

From the data of Nickerson *et al*, a DNA sequence variation between unrelated individuals should occur every 500 base pairs (it is this figure that we use for inter-population perfectly matched duplex depletion with self-against-ancestral-nontransmitted alleles and inter-population perfectly matched duplex depletion between unrelated individuals). In addition, a given individual should be heterozygous once every 573 base pairs (this figure is used for inter-population perfectly matched duplex depletion with self-against-ancestral-transmitted alleles and self-against-self inter-population perfectly matched duplex molecular depletion). Inter-population perfectly matched duplex depletion against the transmitted alleles and the nontransmitted alleles will now be considered separately.

Self-against-ancestral-transmitted alleles inter-population perfectly matched duplex depletion

If we anneal the two complementary strands for a fragment using DNA from an 'affected' descendent and DNA containing the transmitted alleles of 'unaffected' ancestors, then $\{1-e^{(-F/573)}\}$ of fragments will contain one or more site of heterozygosity. {1/A} of the annealings will be of this type. For such annealings, where a site of heterozygosity is present, the probability of obtaining a mismatch-containing duplex between a biotinylated fragment and a non-biotinylated fragment containing the site of heterozygosity is 0.5.

Self-against-ancestral-nontransmitted alleles inter-population perfectly matched duplex depletion

If we anneal the two complementary strands for a fragment using DNA from an 'affected' descendent and DNA containing the nontransmitted alleles of 'unaffected' ancestors, then $\{1-e^{(-F/500)}\}$ of fragments will contain one or more site of DNA sequence variation. {(A-1)/A} annealings will be of this type.

Phenotype-determining fragment enrichment

The fraction of fragments carried through the first cycle of inter-population perfectly matched duplex depletion will therefore be

$$([0.5\cdot\{1/A\}\cdot\{1-e^{(-F/573)}\}]+[\{(A-1)/A\}\cdot\{1-e^{(-F/500)}\}])\cdot\{Y/(X+Y)\}$$

where {Y/(X+Y)} represents the yield of streptavidin-capturable molecules.

Hence n repetitions of such a process will reduce the original number of fragments to the following fraction

$$([0.5\cdot\{1/A\}\cdot\{1-e^{(-F/573)}\}]+[\{(A-1)/A\}\cdot\{1-e^{(-F/500)}\}])^{n}\cdot\{Y/(X+Y)\}^{n}$$

The enrichment for phenotype-determining fragments over SNP-containing ('noise') fragments during inter-population perfectly matched duplex depletion cycles will therefore be given by

$$([0.5\cdot\{1/A\}\cdot\{1-e^{(-F/573)}\}]+[\{(A-1)/A\}\cdot\{1-e^{(-F/500)}\}])^{-n}$$

Example figures for enrichment are given for A = 2, 3 and 4 below with F = 50, 100, 200, 300, 400, 500, 600, 700, 800, 900 and 1,000 and n = 1, 2, 3, 4 , 5 ,6 ,7, 8, 9 and 10.

For A=2

	n=									
F	1	2	3	4	5	6	7	8	9	10
50	15	213	3,115	45,494	664,427	9,703,694	1.E+08	2.E+09	3.E+10	4.E+11
100	8	59	448	3,430	26,250	200,887	1,537,367	1.E+07	9.E+07	7.E+08
200	4	18	74	309	1,296	5,434	22,782	95,524	400,520	1,679,340
300	3	9	28	87	265	811	2,475	7,558	23,078	70,468
400	2	6	16	39	97	241	600	1,497	3,734	9,313
500	2	5	10	22	48	103	224	485	1,051	2,277
600	2	4	7	15	29	56	109	213	416	813
700	2	3	6	11	19	35	63	114	207	374
800	2	3	5	8	14	24	42	71	121	205
900	2	3	4	7	11	18	30	49	79	128
1,000	2	2	4	6	9	15	23	36	57	89

For A=3

	n=									
F	1	2	3	4	5	6	7	8	9	10
50	13	167	2,154	27,823	359,340	4,640,941	6.E+07	8.E+08	1.E+10	1.E+11
100	7	46	311	2,104	14,251	96,516	653,682	4,427,241	3.E+07	2.E+08
200	4	14	51	191	709	2,633	9,786	36,365	135,135	502,173
300	3	7	20	54	146	396	1,073	2,909	7,882	21,361
400	2	5	11	24	53	119	263	582	1,290	2,859
500	2	4	7	14	27	51	99	190	367	707
600	2	3	5	9	16	28	48	84	147	255
700	2	3	4	7	11	18	28	46	74	119
800	2	2	4	5	8	12	19	28	43	66
900	1	2	3	4	6	9	14	20	29	41
1,000	1	2	3	4	5	8	11	15	21	29

For A=4

	n=									
F	1	2	3	4	5	6	7	8	9	10
50	12	149	1,826	22,316	272,759	3,333,771	4.E+07	5.E+08	6.E+09	7.E+10
100	6	41	264	1,690	10,834	69,465	445,373	2.855.512	2.E+07	1.E+08
200	4	12	44	154	540	1.902	6.696	23,570	82.967	292,045
300	3	7	17	44	112	287	737	1.894	4,864	12,494
400	2	4	9	20	41	86	181	381	800	1,681
500	2	3	6	11	20	37	68	125	229	418
600	2	3	5	7	12	20	34	56	92	152
700	2	2	4	5	8	13	20	30	46	71
800	1	2	3	4	6	9	13	19	27	39
900	1	2	3	4	5	7	9	13	18	25
1,000	1	2	2	3	4	6	7	10	13	17

Extension of the invention to the case of fully comprehensive 'abnormal' cell mutational profiling within an individual

If we now take 'normal' cells as the source of our biotinylated fragments and 'abnormal' cells from the same individual as the source of our non-biotinylated fragments, any fragments that have acquired sequence changes on the way to becoming 'abnormal' will be unable to form perfectly matched duplexes with the biotinylated 'normal' fragments. Successive cycles of such inter-population perfectly matched duplex depletion will thus lead to the enrichment of fragments carrying all those sequence differences between the 'normal' cells and the 'abnormal' cells - the degree of enrichment per cycle being as described below.

From the data of Nickerson *et al*, a given individual should be heterozygous about once every 573 base pairs.

If we anneal the two complementary strands for a fragment using DNA from 'abnormal' cells and DNA from 'normal' cells, then $\{1-e^{(-F/573)}\}$ of fragments will contain one or more site of heterozygosity.

For a heterozygous site, p and q are both 0.5. The probability of obtaining a perfectly matched duplex between a biotinylated fragment

and a non-biotinylated fragment containing the site is 0.5. Similarly, the probability of obtaining a mismatch-containing duplex between a biotinylated fragment and a non-biotinylated fragment containing the site is also 0.5.

The fraction of fragments carried through the first cycle of inter-population perfectly matched duplex depletion will therefore be

$$0.5 \cdot \{1-e^{(-F/573)}\} \cdot \{Y/(X+Y)\}$$

hence n repetitions of such a process will reduce the original number of fragments to the following fraction

$$[0.5 \cdot \{1-e^{(-F/573)}\}]^{n} \cdot \{Y/(X+Y)\}^{n}$$

The enrichment for phenotype-determining fragments over SNP-containing ('noise') fragments during inter-population perfectly matched duplex depletion cycles will therefore be given by

$$[0.5 \cdot \{1-e^{(-F/573)}\}]^{-n}$$

Example figures for enrichment are given below with F = 50, 100, 200, 300, 400, 500, 600, 700, 800, 900 and 1,000 and n = 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10.

	n=									
F	1	2	3	4	5	6	7	8	9	10
50	24	573	13,711	328,171	7,854,627	2.E+08	4.E+09	1.E+11	3.E+12	6.E+13
100	12	156	1,948	24,329	303,844	3,794,731	5.E+07	6.E+08	7.E+09	9.E+10
200	7	46	313	2,123	14,412	97,830	664,076	4,507,784	3.E+07	2.E+08
300	5	24	118	580	2,845	13,958	68,490	336,072	1,649,059	8,091,708
400	4	16	63	251	999	3,977	15,831	63,012	250,815	998,346
500	3	12	41	139	479	1,644	5,650	19,411	66,688	229,114
600	3	9	29	90	278	856	2,638	8,128	25,047	77,179
700	3	8	23	65	183	520	1,475	4,183	11,862	33,639
800	3	7	19	50	133	353	937	2,491	6,621	17,599
900	3	6	16	41	103	259	654	1,652	4,171	10,532
1,000	2	6	14	34	84	202	490	1,188	2,880	6,978

This approach will enrich for fragments containing *all* sequence differences within the 'abnormal' cells - no prior knowledge of the genes (oncogenes, tumour suppressor genes etc.) that may need to be investigated is required.

Having thus isolated a fragment (or fragments) determining differences between the 'affected' and 'unaffected' populations, we can now proceed to their analysis.

Aspect (II)

In another aspect the invention provides a method of making a set of arrays of fragments of DNA of interest, which method comprises:

a) selecting, from a set of n restriction endonuclease enzymes, a subset of r restriction endonuclease enzymes;

b) digesting genomic DNA with the subset of r enzymes;

c) ligating to the resulting fragments restriction-enzyme-cutting-site-specific adapters with unique polymerase chain reaction amplifiable sequences;

d) splitting the resulting fragments into r^2 aliquots;

e) amplifying each aliquot with two-restriction enzyme-specific

primers;

f) forming an array of the r^2 aliquots of the amplimers; and

g) repeating steps a) to f) using a different subset of r restriction endonuclease enzymes. The invention also includes sets of arrays obtained or obtainable by the method.

The n restriction endonuclease enzymes may be selected from 4-cutters and 5-cutters and 6-cutters, and a set may include enzymes from one or two or three of these categories. The value of n is preferably 3 to 10, for reasons discussed below. The value of r is less than n and is preferably 2 to 4, chosen with reference to the frequency with which the chosen enzymes cut nucleic acids, and ease of fragment amplification by PCR.

Terminal restriction site profiling arrays (TRSPAs)

If we use a total of n 6 bp cutter restriction enzymes within the total set of enzymes used for fragmentation, let us use subsets of r 6 bp cutter enzymes (taken from the total set n) to make (rxr) TRSPA test matrices as follows.

Version 1 - (TRSPA-1)

For each (rxr) TRSPA test matrix, digest DNA to completion with all r restriction enzymes. Ligate on restriction enzyme cutting-site-specific adaptors with unique polymerase chain reaction amplifiable tags. Split into r^2 aliquots and for each aliquot, amplify with biotinylated restriction $enzyme_j$ and non-biotinylated restriction $enzyme_k$ tag-specific primers and array the non-biotinylated strands for all values of j and k between 1 and r.

Example

Consider the following dsDNA

--[A]--1--[B]--2--[B]--3--[C]--4--[A]--5--[B]-- +

--[A]--1--[B]--2--[B]--3--[C]--4--[A]--5--[B]-- -

where the restriction enzyme cutting sites are denoted A, B and C and the fragments after restriction digestion are denoted 1, 2, 3, 4 and 5. + and - denote the sense of the strands.

Cut to completion with A, B and C to give

A--1--B B--2--B B--3--C C--4--A A--5--B +
A--1--B B--2--B B--3--C C--4--A A--5--B -

Polymerase chain reaction amplify according to the following primer matrix and streptavidin capture to give

	A	B	C
bio-A		bio-A--1--B + A--1--B - bio-A--5--B + A--5--B -	C--4--A + C--4--A-bio -
bio-B	A--1--B + A--1--B-bio - A--5--B + A--5--B-bio -	bio-B--2--B + B--2--B - B--2--B + B--2--B-bio -	bio-B--3--C + B--3--C -
bio-C	bio-C--4--A + C--4--A -	B--3--C + B--3--C-bio -	

Keep only the non-biotinylated strands to give

	A	B	C
bio-A		A--1--B - A--5--B -	C--4--A +
bio-B	A--1--B + A--5--B +	B--2--B - B--2--B +	B--3--C -
bio-C	C--4--A -	B--3--C +	

In simplified form

	A	B	C
bio-A		1-, 5-	4+
bio-B	1+, 5+	2-, 2+	3-
bio-C	4-	3+	

Repeat for all ${}^{n}C_{r}$ combinations of restriction enzymes to generate the TRSPA.

TRSPA-1 hybridisation patterns

The two types of TRSPA-1 hybridisation pattern we should expect using a probe resulting from inter-population perfectly matched duplex depletion are

1. Hybridisation to an off-diagonal element (e.g. row x, column y - where x and y are different) and its complementary element reflected across the diagonal (i.e. row y, column x) and
2. Hybridisation to an on-diagonal element (e.g. the element at row z, column z).

TRSPA-1 analysis - a worked example for n=3 and r=2

Let us take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

where

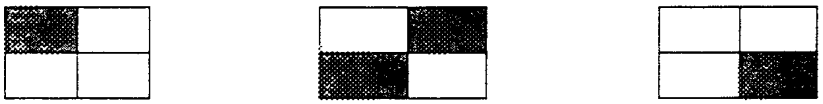
A, B and C denote restriction enzyme cutting sites. 1, 2, 3, 4, 5, 6, 7, 8 and 9 denote the restriction fragments after digestion.

The TRSPA-1 test matrices

There are ${}^{3}C_{2}=3$ TRSPA-1 test matrices - AB, BC and AC.

The test matrix hybridisation patterns

There are {(2·(2+1))/2}=3 possible hybridisation patterns for each of the test matrices



The combinatorial diversity

If there are three possible TRSPA-1 test matrix hybridisation patterns and three possible test matrices, then there will be 3^3=27 possible TRSPA-1 signatures.

Fragment 4 analysis - the AB TRSPA-1 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with A and B to give

-C-1-B
B-2-B
B-3-A
A-4-A
A-5-C-6-C-7-B
B-8-A
A-9-A-

The fragment complementary to fragment 4 will be

A-4-A

Polymerase chain reaction amplify and hybridise with just fragment 4 to give

AB matrix	A	B
bio-A		
bio-B		

Fragment 4 analysis - the BC TRSPA-1 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with B and C to give

-C-1-B

B-2-B

B-3-A-4-A-5-C

C-6-C

C-7-B

B-8-A-9-A-

The fragment complementary to fragment 4 will be

B-3-A-4-A-5-C

Polymerase chain reaction amplify and hybridise with just fragment 4 to give

BC matrix	B	C
bio-B		
bio-C		

Fragment 4 analysis - the AC TRSPA-1 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with A and C to give

-C

C-1-B-2-B-3-A

A-4-A

A-5-C

C-6-C

C-7-B-8-A

A-9-A

A-

The fragment complementary to fragment 4 will be

A-4-A

Polymerase chain reaction amplify and hybridise with just fragment 4 to give

AC matrix	A	C
bio-A		
bio-C		

Fragment 5 analysis - the AB TRSPA-1 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with A and B to give

-C-1-B

B-2-B

B-3-A

A-4-A

A-5-C-6-C-7-B

B-8-A

A-9-A-

The fragment complementary to fragment 5 will be

A-5-C-6-C-7-B

Polymerase chain reaction amplify and hybridise with just fragment 5 to give

AB matrix	A	B
bio-A		
bio-B		

Fragment 5 analysis - the BC TRSPA-1 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with B and C to give

-C-1-B

B-2-B

B-3-A-4-A-5-C

C-6-C

C-7-B

B-8-A-9-A-

The fragment complementary to fragment 5 will be

B-3-A-4-A-5-C

Polymerase chain reaction amplify and hybridise with just fragment 5 to give

BC matrix	B	C
bio-B		
bio-C		

Fragment 5 analysis - the AC TRSPA-1 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with A and C to give

-C

C-1-B-2-B-3-A

A-4-A

A-5-C

C-6-C

C-7-B-8-A

A-9-A

A-

The fragment complementary to fragment 5 will be

A-5-C

Polymerase chain reaction amplify and hybridise with just fragment 5 to give

AC matrix	A	C
bio-A		
bio-C		

Fragment 6 analysis - the AB TRSPA-1 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with A and B to give

-C-1-B

B-2-B

B-3-A

A-4-A

A-5-C-6-C-7-B

B-8-A

A-9-A-

The fragment complementary to fragment 6 will be

A-5-C-6-C-7-B

Polymerase chain reaction amplify and hybridise with just fragment 6 to give

AB matrix	A	B
bio-A		
bio-B		

Fragment 6 analysis - the BC TRSPA-1 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with B and C to give

-C-1-B

B-2-B

B-3-A-4-A-5-C

C-6-C

C-7-B

B-8-A-9-A-

The fragment complementary to fragment 6 will be

C-6-C

Polymerase chain reaction amplify and hybridise with just fragment 6 to give

BC matrix	B	C
bio-B		
bio-C		

Fragment 6 analysis - the AC TRSPA-1 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with A and C to give

-C
C-1-B-2-B-3-A
A-4-A
A-5-C
C-6-C
C-7-B-8-A
A-9-A
A-

The fragment complementary to fragment 6 will be

C-6-C

Polymerase chain reaction amplify and hybridise with just fragment 6 to give

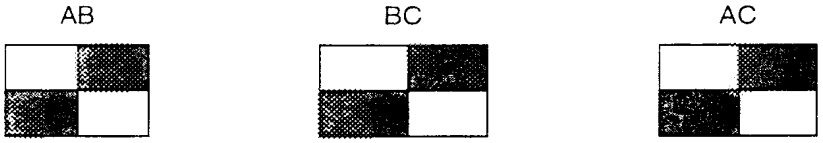
AC matrix	A	C
bio-A		
bio-C		

Overall results

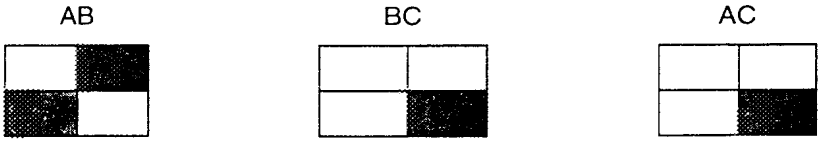
Fragment 4 TRSPA-1 analysis



Fragment 5 TRSPA-1 analysis



Fragment 6 TRSPA-1 analysis



Version 2 - (TRSPA-2)

For each (rxr) TRSPA test matrix, digest DNA to completion with all r restriction enzymes. Ligate on restriction enzyme cutting-site-specific adaptors with unique polymerase chain reaction amplifiable tags. Split into r^2 aliquots and for each aliquot, amplify with non-biotinylated restriction enzyme$_j$ and non-biotinylated restriction enzyme$_k$ tag-specific primers and array the denatured strands for all values of j and k between 1 and r.

Example

Consider the following dsDNA

--[A]--1--[B]--2--[B]--3--[C]--4--[A]--5--[B]-- +

--[A]--1--[B]--2--[B]--3--[C]--4--[A]--5--[B]-- -

where the restriction enzyme cutting sites are denoted A, B and C and the fragments after restriction digestion are denoted 1, 2, 3, 4 and 5. + and - denote the sense of the strands.

Cut to completion with A, B and C to give

A--1--B B--2--B B--3--C C--4--A A--5--B +

A--1--B B--2--B B--3--C C--4--A A--5--B -

Polymerase chain reaction amplify according to the following primer matrix to give

	A	B	C
A		A--1--B + A--1--B - B--2--B + B--2--B - A--5--B + A--5--B -	C--4--A + C--4--A -
B	A--1--B + A--1--B - B--2--B + B--2--B - A--5--B + A--5--B -	B--2--B + B--2--B -	B--2--B + B--2--B - B--3--C + B--3--C -
C	C--4--A + C--4--A -	B--2--B + B--2--B - B--3--C + B--3--C -	

In simplified form

	A	B	C
A		1, 2, 5	4
B	1, 2, 5	2	2, 3
C	4	2, 3	

Repeat for all ${}^{n}C_{r}$ combinations of restriction enzymes to generate the TRSPA.

TSPSA-2 hybridisation patterns

The two types of TRSPA-2 hybridisation pattern we should expect using a probe resulting from inter-population perfectly matched duplex depletion are

1. Hybridisation to an off-diagonal element (e.g. row x, column y - where x and y are different) and its complementary element reflected across the diagonal (i.e. row y, column x) and
2. Hybridisation to a whole row and column intersecting at an on-diagonal element (e.g. all of row z and all of column z).

TRSPA-2 analysis - a worked example for n=3 and r=2

Let us take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

where

A, B and C denote restriction enzyme cutting sites. 1, 2, 3, 4, 5, 6, 7, 8 and 9 denote the restriction fragments after digestion.

The TRSPA-2 test matrices

There are ${}^{3}C_{2}=3$ TRSPA-2 test matrices - AB, BC and AC.

The test matrix hybridisation patterns

There are $\{(2\cdot(2+1))/2\}=3$ possible hybridisation patterns for each of the test matrices





The combinatorial diversity

If there are three possible TRSPA-2 test matrix hybridisation

patterns and three possible test matrices, then there will be 3^3=27 possible TRSPA-2 signatures.

Fragment 4 analysis - the AB TRSPA-2 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with A and B to give

-C-1-B
B-2-B
B-3-A
A-4-A
A-5-C-6-C-7-B
B-8-A
A-9-A-

The fragment complementary to fragment 4 will be

A-4-A

Polymerase chain reaction amplify and hybridise with just fragment 4 to give

AB matrix	A	B
A		
B		

Fragment 4 analysis - the BC TRSPA-2 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with B and C to give

-C-1-B

B-2-B

B-3-A-4-A-5-C

C-6-C

C-7-B

B-8-A-9-A-

The fragment complementary to fragment 4 will be

B-3-A-4-A-5-C

Polymerase chain reaction amplify and hybridise with just fragment 4 to give

BC matrix	B	C
B		
C		

Fragment 4 analysis - the AC TRSPA-2 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with A and C to give

-C

C-1-B-2-B-3-A

A-4-A

A-5-C

C-6-C

C-7-B-8-A

A-9-A

A-

The fragment complementary to fragment 4 will be

A-4-A

Polymerase chain reaction amplify and hybridise with just fragment 4 to give

AC matrix	A	C
A		
C		

Fragment 5 analysis - the AB TRSPA-2 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with A and B to give

-C-1-B

B-2-B

B-3-A

A-4-A

A-5-C-6-C-7-B

B-8-A

A-9-A-

The fragment complementary to fragment 5 will be

A-5-C-6-C-7-B

Polymerase chain reaction amplify and hybridise with just fragment 5 to give

AB matrix	A	B
A		
B		

Fragment 5 analysis - the BC TRSPA-2 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with B and C to give

-C-1-B

B-2-B

B-3-A-4-A-5-C

C-6-C

C-7-B

B-8-A-9-A-

The fragment complementary to fragment 5 will be

B-3-A-4-A-5-C

Polymerase chain reaction amplify and hybridise with just fragment 5 to give

BC matrix	B	C
B		
C		

Fragment 5 analysis - the AC TRSPA-2 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with A and C to give

-C

C-1-B-2-B-3-A

A-4-A

A-5-C

C-6-C

C-7-B-8-A

A-9-A

A-

The fragment complementary to fragment 5 will be

A-5-C

Polymerase chain reaction amplify and hybridise with just fragment 5 to give

AC matrix	A	C
A		
C		

Fragment 6 analysis - the AB TRSPA-2 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with A and B to give

-C-1-B
B-2-B
B-3-A
A-4-A
A-5-C-6-C-7-B
B-8-A
A-9-A-

The fragment complementary to fragment 6 will be

A-5-C-6-C-7-B

Polymerase chain reaction amplify and hybridise with just fragment 6 to give

AB matrix	A	B
A		
B		

Fragment 6 analysis - the BC TRSPA-2 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with B and C to give

-C-1-B

B-2-B

B-3-A-4-A-5-C

C-6-C

C-7-B

B-8-A-9-A-

The fragment complementary to fragment 6 will be

C-6-C

Polymerase chain reaction amplify and hybridise with just fragment 6 to give

BC matrix	B	C
B		
C		

Fragment 6 analysis - the AC TRSPA-2 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with A and C to give

-C

C-1-B-2-B-3-A

A-4-A

A-5-C

C-6-C

C-7-B-8-A

A-9-A

A-

The fragment complementary to fragment 6 will be

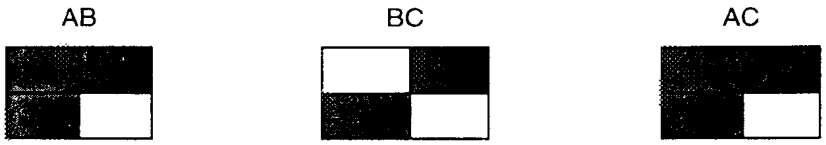
C-6-C

Polymerase chain reaction amplify and hybridise with just fragment 6 to give

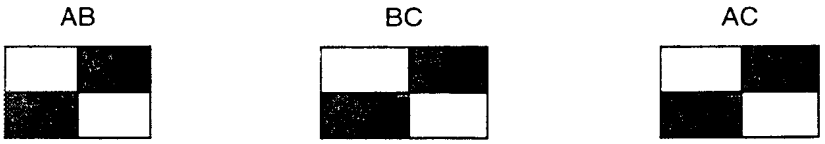
AC matrix	A	C
A		
C		

Overall results

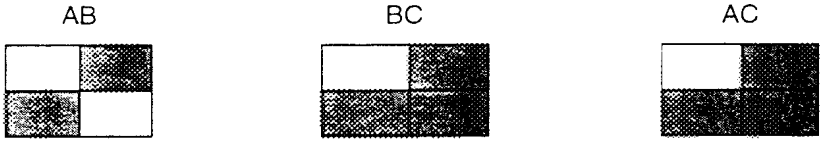
Fragment 4 TRSPA-2 analysis



Fragment 5 TRSPA-2 analysis



Fragment 6 TRSPA-2 analysis



The number of test matrices

For a total of n enzymes used for fragmentation and a panel of r enzymes per (rxr) test matrix, there will be ${}^{n}C_{r}$ possible (rxr) test matrices, where

$${}^{n}C_{r} = (n!)/[(n-r)!\cdot r!]$$

${}^{n}C_{r}$ for various n and r is given in the following table

	n=							
r	3	4	5	6	7	8	9	10
1	3	4	5	6	7	8	9	10
2	3	6	10	15	21	28	36	45
3	1	4	10	20	35	56	84	120
4	-	1	5	15	35	70	126	210
5	-	-	1	6	21	56	126	252
6	-	-	-	1	7	28	84	210
7	-	-	-	-	1	8	36	120
8	-	-	-	-	-	1	9	45
9	-	-	-	-	-	-	1	10
10	-	-	-	-	-	-	-	1

The number of TRSPA spots arrayed

The total number of TRSPA spots is given by

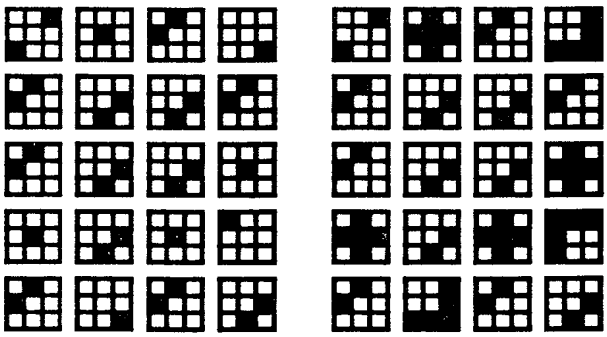
$$r^{2}\cdot\{{}^{n}C_{r}\} = r^{2}\cdot\{(n!)/[(n-r)!\cdot r!]\}$$

$r^2 \cdot \{^nC_r\}$ for various n and r is given in the following table

	n =							
R	3	4	5	6	7	8	9	10
1	3	4	5	6	7	8	9	10
2	12	24	40	60	84	112	144	180
3	9	36	90	180	315	504	756	1,080
4	-	16	80	240	560	1,120	2,016	3,360
5	-	-	25	150	525	1,400	3,150	6,300
6	-	-	-	36	252	1,008	3,024	7,560
7	-	-	-	-	49	392	1,764	5,880
8	-	-	-	-	-	64	576	2,880
9	-	-	-	-	-	-	81	810
10	-	-	-	-	-	-	-	100

Each test as above will give rise to a particular hybridisation 'signature'.

TRSPA-1 (for n=6 and r=3) TRSPA-2 (for n=6 and r=3)



There will be {1+2+...+r} = {r(r+1)/2} patterns per (rxr) test matrix. If there are ^nC_r (rxr) test matrices, the total possible number of signatures will be given by {r(r+1)/2} raised to the power ^nC_r. We require that the number of such different signatures is greatly in excess over the

number of fragments generated upon fragmentation.

If we choose n=6 and r=3, there will be 6 possible hybridisation patterns per test and 20 possible tests - giving a combinatorial diversity of $6^{20} = 3.7 \cdot 10^{15}$. The preferred scheme therefore employs six enzymes, tested in groups of three - giving 180 spots per experiment. In order to avoid restriction fragment length polymorphism problems, a duplicate analysis could be performed with a different set of enzymes sharing no cutting sites in common with the first set.

Example restriction enzyme sets for the preparation of test matrices for TRSPAs

Criteria for enzyme choice

In order to make TRSPAs, the selection of suitable enzymes is an important factor. Ideally, two sets of different enzymes are required to eliminate the small possibility that a phenotype-determining polymorphism might fall within a chosen restriction site and therefore compromise the specificity of the resulting signature. The selection of enzymes can be based upon a number of criteria

a) The enzymes should be 6 bp cutters.

b) Cleavage by any selected enzyme should leave a 4 bp overhang at the 5' end.

c) The selected enzymes in each set should all work efficiently under the same buffer conditions.

d) The selected enzymes in each set should ideally work efficiently at a single incubation temperature.

e) The chosen enzymes should be commercially available - ideally at concentrations of 10 U / µl or more.

f) The 5' overhangs left by any two enzymes in the same set should not be identical.

g) No enzyme should appear in both sets for TRSPA fabrication.

h) Enzymes should be selected to avoid or minimise the effects of mammalian methylation patterns. In particular, enzymes with CG dinucleotides in their recognition sites should be avoided unless the enzyme is known to be able to restrict ^{m5}CpG sites.

DNA methylation

In vertebrates, DNA is often methylated at the 5th position of cytosine in the sequence of CpG and this is the only chemical modification that DNA of vertebrates contains under physiological conditions. By the careful selection of enzymes which do not contain CpG sequences within the recognition site, or the selection of enzymes which freely restrict ^{m5}CpG methylated sites, it is possible to remove the potentially adverse effects of DNA methylation from the TRSPA analysis. 6 bp cutters which are known to restrict ^{m5}CpG modified DNA efficiently to leave a 5' four base overhang are *BspEI* and *XmaI*. These enzymes are therefore the only enzymes with restriction sites containing CpG dinucleotides that are potentially useful in a TRSPA analysis.

Enzyme selection method

Sixteen possible four base pair overhangs exist (excluding unusual enzymes with asymmetrical recognition sequences such as *BssSI* or *BsiI*), five of which contain CG in the sequence. A further four overhangs could potentially contain CG sequences within the restriction recognition site if preceded by a C and followed by a G. Enzymes are therefore preferentially selected from the remaining six groups.

Excluding isoschizomers, there are up to four possible enzymes which would leave a particular 5' four bp overhang. For example, enzymes leaving a CTAG overhang are

AvrII - CCTAGG
NheI - GCTAGC
SpeI - ACTAGT
XbaI - TCTAGA

Enzymes to cleave sites with all the combinations of flanking bases are not available for all overhangs - hence the enzyme choice is more limited for some overhang groups than others.

As a primary step towards enzyme selection, the enzymes supplied by Amersham Pharmacia Biotech, New England Biolabs and Promega are ordered below by overhang sequence. Supplementary details such as the percentage activity in common buffers, the reaction temperature, concentration and supplier are also recorded. For three overhang sequences, there are no available enzymes, the remaining 13 are described below. Enzymes considered unsuitable due to methylation sensitivity are shaded (darker shade). Enzymes considered unfavourable due to the presence of CpG sites even though they do restrict methylated DNA to some degree are also shaded (lighter shaded).

Candidate restriction enzymes leaving a 5' overhang of GTAC

Enzyme	Site	H/M/T buffer + BSA % activity	Optimum °C	U / µl	Supplier	Inactivate °C/min
Acc65I	GGTACC	10/50/100	37	10	NEB	65 / 20
BsiWI	CGTACG	100(50)/100(50)/25(-)	55 (37)	12	NEB	80 / 20
BsrGI	TGTACA	10/100/100	37	10	NEB	80 / 20
SplI (BsiWI)	CGTACG	100/20/60	55	10	APB	No

Candidate restriction enzymes leaving a 5' overhang of TTAA

Enzyme	Site	H/M/T buffer + BSA % activity	Optimum °C	U / µl	Supplier	Inactivate °C/min
AflII	CTTAAG	25/100/100	37	10	APB	60 / 15

Candidate restriction enzymes leaving a 5' overhang of AATT

Enzyme	Site	H/M/T buffer + BSA % activity	Optimum °C	U / µl	Supplier	Inactivate °C/min
EcoRI	GAATTC	100/100/100	37	>40	APB/NEB	65 / 20
MfeI	CAATTG	10/50/100	37	10	NEB	65 / 20

Candidate restriction enzymes leaving a 5' overhang of CATG

Enzyme	Site	H/M/T buffer + BSA % activity	Optimum °C	U / µl	Supplier	Inactivate °C/min
BspHI	TCATGA	50/100/100	37	10	NEB/APB	65 / 20
NcoI	CCATGG	100/100/100	37	50	NEB/APB	65 / 20

Candidate restriction enzymes leaving a 5' overhang of GATC

Enzyme	Site	H/M/T buffer + BSA % activity	Optimum °C	U / µl	Supplier	Inactivate °C/min
BamHI	GGATCC	50/100/75	37	>40	APB/NEB	80 / 20
BclI	TGATCA	100/100/75	50 (37)	10	NEB	No
BglII	AGATCT	100/75/10	37	>40	APB/NEB	No

Candidate restriction enzymes leaving a 5' overhang of CCGG

Enzyme	Site	H/M/T buffer + BSA % activity	Optimu m °C	U / µl	Supplier	Inactivate °C/min
BspEI	TCCGGA	100/10/0	37	10	NEB	80 / 20
NgoM IV	GCCGGC	10/50/100	37	10	NEB	80 / 20
XmaI	CCCGGG	0/100/100	37	10	NEB	65 / 20

Candidate restriction enzymes leaving a 5' overhang of GCGC

Enzyme	Site	H/M/T buffer + BSA % activity	Optimu m °C	U / µl	Supplier	Inactivate °C/min
KasI	GGCGCC	75/100/75	37	5	NEB	65 / 20

Candidate restriction enzymes leaving a 5' overhang of TCGA

Enzyme	Site	H/M/T buffer + BSA % activity	Optimu m °C	U / µl	Supplier	Inactivate °C/min
PaeR7I	CTCGAG	10/100/100	37	20	NEB	No
SalI	GTCGAC	50/0/0	37	>40	APB/NEB	60 / 15
XhoI	CTCGAG	100/100/100	37	>40	APB/NEB	65 / 20

Candidate restriction enzymes leaving a 5' overhang of AGCT

Enzyme	Site	H/M/T buffer + BSA % activity	Optimu m °C	U / µl	Supplier	Inactivate °C/min
HindIII	AAGCTT	10/100/50	37	>40	NEB/APB	65 / 20

Candidate restriction enzymes leaving a 5' overhang of CGCG

Enzyme	Site	H/M/T buffer + BSA % activity	Optimu m °C	U / µl	Supplier	Inactivate °C/min
BssHII	GCGCGC	100/100/100	50	20	NEB/APB	80 / 20
MluI	ACGCGT	100/75/50	37	10	APB/NEB	65 / 20

Candidate restriction enzymes leaving a 5' overhang of GGCC

Enzyme	Site	H/M/T buffer + BSA % activity	Optimu m °C	U / µl	Supplier	Inactivate °C/min
EagI	CGGCCG	100/25/10	37	50	NEB	65 / 20
Eco52I (EagI)	CGGCCG	20/20/20	37	10	APB	60 / 15

Candidate restriction enzymes leaving a 5' overhang of TGCA

Enzyme	Site	H/M/T buffer + BSA % activity	Optimu m °C	U / µl	Supplier	Inactivate °C/min
ApaLI	GTGCAC	10/100/100	37	10	NEB/APB	70 / 15
Alw44I	GTGCAC	-/50/-	37	10	Promega	70 / 15

Candidate restriction enzymes leaving a 5' overhang of CTAG

Enzyme	Site	H/M/T buffer + BSA % activity	Optimu m °C	U / µl	Supplier	Inactivate °C/min
AvrII	CCTAGG	50/100/100	37	4	NEB	No
NheI	GCTAGC	10/100/100	37	10	APB/NEB	65 / 20
SpeI	ACTAGT	25/100/75	37	50	NEB	65 / 20
XbaI	TCTAGA	75/100/75	37	100	NEB	65 / 20
BlnI(AvrII)	CCTAGG	40/20/20	37	10	APB	No

From the above tables, a short-list of the most useful enzymes is given below for each of the buffer conditions shown

Overhang	Buffer H + BSA		Buffer M + BSA		Buffer T + BSA	
	Enzyme	Efficiency	Enzyme	Efficiency	Enzyme	Efficiency
AATT	*EcoRI*	100	*EcoRI*	100	*MfeI*	100
			MfeI	50	-	-
CATG	*NcoI*	100	*NcoI*	100	*NcoI*	100
	BspHI	50	*BspHI*	100	*BspHI*	100
GATC	*BglII*	100	*BamHI*	100	*BamHI*	75
	BamHI	50	*BglII*	75	-	-
TATA	-	-	-	-	-	-
ACGT	-	-	-	-	-	-
CCGG	BspEI	100	*XmaI*	100	*XmaI*	100
GCGC	-	-	-	-	-	-
TCGA	-	-	-	-	-	-
AGCT	-	-	*HindIII*	100	*HindIII*	50
CGCG	-	-	-	-	-	-
GGCC	-	-	-	-	-	-
TGCA	-	-	*ApaLI*	100	*ApaLI*	100
ATAT	-	-	-	-	-	-
CTAG	*XbaI*	75	*XbaI*	100	*NheI*	100
	AvrII	50	*SpeI*	100	*AvrII*	100
	BlnI	40	*NheI*	100	*XbaI*	75
	-	-	*AvrII*	100	*SpeI*	75
GTAC	*Acc65I*	100	*BsrGI*	100	*BsrGI*	100
			Acc65I	75	-	-
TTAA	-	-	*AflII*	100	*AflII*	100

Taking into account all of the criteria explained above, an

example selection of two six-enzyme sets is described below. Reserve enzymes, which could also be used, are shown as well. These reserve enzymes can be substituted (provided this does not cause overhang duplication) if practical problems regarding enzyme availability or performance should occur.

Example enzyme set 1 - for restriction in buffer M + BSA

Number	Enzyme	Site	M+BSA % activity	Optimum °C	U / µl	Supplier	Inactivate °C/min
1	*BamHI*	GGATCC	100	37	100	NEB	80 / 20
2	*BsrGI*	TGTACA	100	37	10	NEB	80 / 20
3	*HindIII*	AAGCTT	100	37	>40	APB	65 / 20
4	*NcoI*	CCATGG	100	37	50	NEB	65 / 20
5	*SpeI*	ACTAGT	100	37	50	NEB	65 / 20
6	*AflII*	CTTAAG	100	37	10	APB	60 / 15

Example enzyme set 2 - for restriction in buffer M + BSA

Number	Enzyme	Site	M+BSA % activity	Optimum °C	U / µl	Supplier	Inactivate °C/min
1	*EcoRI*	GAATTC	100	37	>40	APB	65 / 20
2	*BspHI*	TCATGA	100	37	10	APB	65 / 20
3	*BglII*	AGATCT	75	37	>40	APB	No
4	*XbaI*	TCTAGA	100	37	100	NEB	65 / 20
5	*Acc65I*	GGTACC	75	37	10	NEB	65 / 20
6	*ApaL1*	GTGCAC	100	37	10	NEB	No

1. Example reserve enzymes

Number	Enzyme	Site	M+BSA % activity	Optimum °C	U / µl	Supplier	Inactivate °C/min
r1	*XmaI*	CCCGGG	100	37	10	NEB	65 / 20
r2	*BspEI*	TCCGGA	100	37	10	NEB	80 / 20
r3	*NheI*	GCTAGC	100	37	10	APB	65/20
r4	*AvrII*	CCTAGG	100	37	4	NEB	No

Example specific (3x3) terminal restriction site profiling array (TRSPA) test matrices for above two sets of six enzymes

Triplet combinations from *BamHI*, *BsrGI*, *HindIII*, *NcoI*, *SpeI* and *AflII*

For example set 1 - *BamHI*, *BsrGI*, *HindIII*, *NcoI*, *SpeI* and *AflII* - the 20 (= $^{6}C_{3}$) triplet combinations are

BamHI BsrGI HindIII	*BamHI HindIII SpeI*	*BsrGI HindIII NcoI*	*BsrGI SpeI AflII*
BamHI BsrGI NcoI	*BamHI HindIII AflII*	*BsrGI HindIII SpeI*	*HindIII NcoI SpeI*
BamHI BsrGI SpeI	*BamHI NcoI SpeI*	*BsrGI HindIII AflII*	*HindIII NcoI AflII*
BamHI BsrGI AflII	*BamHI NcoI AflII*	*BsrGI NcoI SpeI*	*HindIII SpeI AflII*
BamHI HindIII NcoI	*BamHI SpeI AflII*	*BsrGI NcoI AflII*	*NcoI SpeI AflII*

Triplet combinations from *EcoRI*, *BspHI*, *BglII*, *XbaI*, *Acc65I* and *ApaLI*

For example set 2 - *EcoRI*, *BspHI*, *BglII*, *XbaI*, *Acc65I* and *ApaLI* - the 20 (= $^{6}C_{3}$) triplet combinations are

EcoRI BspHI BglII	*EcoRI BglII Acc65I*	*BspHI BglII XbaI*	*BspHI Acc65I ApaLI*
EcoRI BspHI XbaI	*EcoRI BglII ApaLI*	*BspHI BglII Acc65I*	*BglII XbaI Acc65I*
EcoRI BspHI Acc65I	*EcoRI XbaI Acc65I*	*BspHI BglII ApaLI*	*BglII XbaI ApaLI*
EcoRI BspHI ApaLI	*EcoRI XbaI ApaLI*	*BspHI XbaI Acc65I*	*BglII Acc65I ApaLI*
EcoRI BglII XbaI	*EcoRI Acc65I ApaLI*	*BspHI XbaI ApaLI*	*XbaI Acc65I ApaLI*

Example TRSPA-1 test matrices for set 1 - *BamHI*, *BsrGI*, *HindIII*, *NcoI*, *SpeI* and *AflII*

matrix 1	5'-HO-*BamHI*-primer	5'-HO-*BsrGI*-primer	5'-HO-*HindIII*-primer
5'-biotin-*BamHI*-primer	1	2	3
5'-biotin-*BsrGI*-primer	4	5	6
5'-biotin-*HindIII*-primer	7	8	9

matrix 2	5'-HO-*BamHI*-primer	5'-HO-*BsrGI*-primer	5'-HO-*NcoI*-primer
5'-biotin-*BamHI*-primer	1	2	3
5'-biotin-*BsrGI*-primer	4	5	6
5'-biotin-*NcoI*-primer	7	8	9

matrix 3	5'-HO-*BamHI*-primer	5'-HO-*BsrGI*-primer	5'-HO-*SpeI*-primer
5'-biotin-*BamHI*-primer	1	2	3
5'-biotin-*BsrGI*-primer	4	5	6
5'-biotin-*SpeI*-primer	7	8	9

matrix 4	5'-HO-*BamHI*-primer	5'-HO-*BsrGI*-primer	5'-HO-*AflII*-primer
5'-biotin-*BamHI*-primer	1	2	3
5'-biotin-*BsrGI*-primer	4	5	6
5'-biotin-*AflII*-primer	7	8	9

matrix 5	5'-HO-*BamHI*-primer	5'-HO-*HindIII*-primer	5'-HO-*NcoI*-primer
5'-biotin-*BamHI*-primer	1	2	3
5'-biotin-*HindIII*-primer	4	5	6
5'-biotin-*NcoI*-primer	7	8	9

matrix 6	5'-HO-*BamHI*-primer	5'-HO-*HindIII*-primer	5'-HO-*SpeI*-primer
5'-biotin-*BamHI*-primer	1	2	3
5'-biotin-*HindIII*-primer	4	5	6
5'-biotin-*SpeI*-primer	7	8	9

matrix 7	5'-HO-*BamHI*-primer	5'-HO-*HindIII*-primer	5'-HO-*AflII*-primer
5'-biotin-*BamHI*-primer	1	2	3
5'-biotin-*HindIII*-primer	4	5	6
5'-biotin-*AflII*-primer	7	8	9

matrix 8	5'-HO-*BamHI*-primer	5'-HO-*NcoI*-primer	5'-HO-*SpeI*-primer
5'-biotin-*BamHI*-primer	1	2	3
5'-biotin-*NcoI*-primer	4	5	6
5'-biotin-*SpeI*-primer	7	8	9

matrix 9	5'-HO-*BamHI*-primer	5'-HO-*NcoI*-primer	5'-HO-*AflII*-primer
5'-biotin-*BamHI*-primer	1	2	3
5'-biotin-*NcoI*-primer	4	5	6
5'-biotin-*AflII*-primer	7	8	9

matrix 10	5'-HO-*BamHI*-primer	5'-HO-*SpeI*-primer	5'-HO-*AflII*-primer
5'-biotin-*BamHI*-primer	1	2	3
5'-biotin-*SpeI*-primer	4	5	6
5'-biotin-*AflII*-primer	7	8	9

matrix 11	5'-HO-*BsrGI*-primer	5'-HO-*HindIII*-primer	5'-HO-*NcoI*-primer
5'-biotin-*BsrGI*-primer	1	2	3
5'-biotin-*HindIII*-primer	4	5	6
5'-biotin-*NcoI*-primer	7	8	9

matrix 12	5'-HO-*BsrGI*-primer	5'-HO-*HindIII*-primer	5'-HO-*SpeI*-primer
5'-biotin-*BsrGI*-primer	1	2	3
5'-biotin-*HindIII*-primer	4	5	6
5'-biotin-*SpeI*-primer	7	8	9

matrix 13	5'-HO-*BsrGI*-primer	5'-HO-*HindIII*-primer	5'-HO-*AflII*-primer
5'-biotin-*BsrGI*-primer	1	2	3
5'-biotin-*HindIII*-primer	4	5	6
5'-biotin-*AflII*-primer	7	8	9

matrix 14	5'-HO-*BsrGI*-primer	5'-HO-*NcoI*-primer	5'-HO-*SpeI*-primer
5'-biotin-*BsrGI*-primer	1	2	3
5'-biotin-*NcoI*-primer	4	5	6
5'-biotin-*SpeI*-primer	7	8	9

matrix 15	5'-HO-*BsrGI*-primer	5'-HO-*NcoI*-primer	5'-HO-*AflII*-primer
5'-biotin-*BsrGI*-primer	1	2	3
5'-biotin-*NcoI*-primer	4	5	6
5'-biotin-*AflII*-primer	7	8	9

matrix 16	5'-HO-*BsrGI*-primer	5'-HO-*SpeI*-primer	5'-HO-*AflII*-primer
5'-biotin-*BsrGI*-primer	1	2	3
5'-biotin-*SpeI*-primer	4	5	6
5'-biotin-*AflII*-primer	7	8	9

matrix 17	5'-HO-*HindIII*-primer	5'-HO-*NcoI*-primer	5'-HO-*SpeI*-primer
5'-biotin-*HindIII*-primer	1	2	3
5'-biotin-*NcoI*-primer	4	5	6
5'-biotin-*SpeI*-primer	7	8	9

matrix 18	5'-HO-*HindIII*-primer	5'-HO-*NcoI*-primer	5'-HO-*AflII*-primer
5'-biotin-*HindIII*-primer	1	2	3
5'-biotin-*NcoI*-primer	4	5	6
5'-biotin-*AflII*-primer	7	8	9

matrix 19	5'-HO-*HindIII*-primer	5'-HO-*SpeI*-primer	5'-HO-*AflII*-primer
5'-biotin-*HindIII*-primer	1	2	3
5'-biotin-*SpeI*-primer	4	5	6
5'-biotin-*AflII*-primer	7	8	9

matrix 20	5'-HO-*NcoI*-primer	5'-HO-*SpeI*-primer	5'-HO-*AflII*-primer
5'-biotin-*NcoI*-primer	1	2	3
5'-biotin-*SpeI*-primer	4	5	6
5'-biotin-*AflII*-primer	7	8	9

Example TRSPA-1 test matrices for set 2 - *EcoRI*, *BspHI*, *BglII*, *XbaI*, *Acc65I* and *ApaLI*

matrix 21	5'-HO-*EcoRI*-primer	5'-HO-*BspHI*-primer	5'-HO-*BglII*-primer
5'-biotin-*EcoRI*-primer	1	2	3
5'-biotin-*BspHI*-primer	4	5	6
5'-biotin-*BglII*-primer	7	8	9

matrix 22	5'-HO-*EcoRI*-primer	5'-HO-*BspHI*-primer	5'-HO-*XbaI*-primer
5'-biotin-*EcoRI*-primer	1	2	3
5'-biotin-*BspHI*-primer	4	5	6
5'-biotin-*XbaI*-primer	7	8	9

matrix 23	5'-HO-*EcoRI*-primer	5'-HO-*BspHI*-primer	5'-HO-*Acc65I*-primer
5'-biotin-*EcoRI*-primer	1	2	3
5'-biotin-*BspHI*-primer	4	5	6
5'-biotin-*Acc65I*-primer	7	8	9

matrix 24	5'-HO-*EcoRI*-primer	5'-HO-*BspHI*-primer	5'-HO-*ApaLI*-primer
5'-biotin-*EcoRI*-primer	1	2	3
5'-biotin-*BspHI*-primer	4	5	6
5'-biotin-*ApaLI*-primer	7	8	9

matrix 25	5'-HO-*EcoRI*-primer	5'-HO-*BglII*-primer	5'-HO-*XbaI*-primer
5'-biotin-*EcoRI*-primer	1	2	3
5'-biotin-*BglII*-primer	4	5	6
5'-biotin-*XbaI*-primer	7	8	9

matrix 26	5'-HO-*EcoRI*-primer	5'-HO-*BglII*-primer	5'-HO-*Acc65I*-primer
5'-biotin-*EcoRI*-primer	1	2	3
5'-biotin-*BglII*-primer	4	5	6
5'-biotin-*Acc65I*-primer	7	8	9

matrix 27	5'-HO-*EcoRI*-primer	5'-HO-*BglII*-primer	5'-HO-*ApaLI*-primer
5'-biotin-*EcoRI*-primer	1	2	3
5'-biotin-*BglII*-primer	4	5	6
5'-biotin-*ApaLI*-primer	7	8	9

matrix 28	5'-HO-*EcoRI*-primer	5'-HO-*XbaI*-primer	5'-HO-*Acc65I*-primer
5'-biotin-*EcoRI*-primer	1	2	3
5'-biotin-*XbaI*-primer	4	5	6
5'-biotin-*Acc65I*-primer	7	8	9
matrix 29	5'-HO-*EcoRI*-primer	5'-HO-*XbaI*-primer	5'-HO-*ApaLI*-primer
5'-biotin-*EcoRI*-primer	1	2	3
5'-biotin-*XbaI*-primer	4	5	6
5'-biotin-*ApaLI*-primer	7	8	9

matrix 30	5'-HO-*EcoRI*-primer	5'-HO-*Acc65I*-primer	5'-HO-*ApaLI*-primer
5'-biotin-*EcoRI*-primer	1	2	3
5'-biotin-*Acc65I*-primer	4	5	6
5'-biotin-*ApaLI*-primer	7	8	9

matrix 31	5'-HO-*BspHI*-primer	5'-HO-*BglII*-primer	5'-HO-*XbaI*-primer
5'-biotin-*BspHI*-primer	1	2	3
5'-biotin-*BglII*-primer	4	5	6
5'-biotin-*XbaI*-primer	7	8	9

matrix 32	5'-HO-*BspHI*-primer	5'-HO-*BglII*-primer	5'-HO-*Acc65I*-primer
5'-biotin-*BspHI*-primer	1	2	3
5'-biotin-*BglII*-primer	4	5	6
5'-biotin-*Acc65I*-primer	7	8	9

matrix 33	5'-HO-*BspHI*-primer	5'-HO-*BglII*-primer	5'-HO-*ApaLI*-primer
5'-biotin-*BspHI*-primer	1	2	3
5'-biotin-*BglII*-primer	4	5	6
5'-biotin-*ApaLI*-primer	7	8	9

matrix 34	5'-HO-*BspHI*-primer	5'-HO-*XbaI*-primer	5'-HO-*Acc65I*-primer
5'-biotin-*BspHI*-primer	1	2	3
5'-biotin-*XbaI*-primer	4	5	6
5'-biotin-*Acc65I*-primer	7	8	9

matrix 35	5'-HO-*BspHI*-primer	5'-HO-*XbaI*-primer	5'-HO-*ApaLI*-primer
5'-biotin-*BspHI*-primer	1	2	3
5'-biotin-*XbaI*-primer	4	5	6
5'-biotin-*ApaLI*-primer	7	8	9

matrix 36	5'-HO-*BspHI*-primer	5'-HO-*Acc65I*-primer	5'-HO-*ApaLI*-primer
5'-biotin-*BspHI*-primer	1	2	3
5'-biotin-*Acc65I*-primer	4	5	6
5'-biotin-*ApaLI*-primer	7	8	9

matrix 37	5'-HO-*BglII*-primer	5'-HO-*XbaI*-primer	5'-HO-*Acc65I*-primer
5'-biotin-*BglII*-primer	1	2	3
5'-biotin-*XbaI*-primer	4	5	6
5'-biotin-*Acc65I*-primer	7	8	9

matrix 38	5'-HO-*BglII*-primer	5'-HO-*XbaI*-primer	5'-HO-*ApaLI*-primer
5'-biotin-*BglII*-primer	1	2	3
5'-biotin-*XbaI*-primer	4	5	6
5'-biotin-*ApaLI*-primer	7	8	9

matrix 39	5'-HO-*BglII*-primer	5'-HO-*Acc65I*-primer	5'-HO-*ApaLI*-primer
5'-biotin-*BglII*-primer	1	2	3
5'-biotin-*Acc65I*-primer	4	5	6
5'-biotin-*ApaLI*-primer	7	8	9

matrix 40	5'-HO-*XbaI*-primer	5'-HO-*Acc65I*-primer	5'-HO-*ApaLI*-primer
5'-biotin-*XbaI*-primer	1	2	3
5'-biotin-*Acc65I*-primer	4	5	6
5'-biotin-*ApaLI*-primer	7	8	9

Example TRSPA-2 test matrices for set 1 - *BamHI*, *BsrGI*, *HindIII*, *NcoI*, *SpeI* and *AflII*

matrix 1	5'-HO-*BamHI*-primer	5'-HO-*BsrGI*-primer	5'-HO-*HindIII*-primer
5'-HO-*BamHI*-primer	1	2	3
5'-HO-*BsrGI*-primer	4	5	6
5'-HO-*HindIII*-primer	7	8	9

matrix 2	5'-HO-*BamHI*-primer	5'-HO-*BsrGI*-primer	5'-HO-*NcoI*-primer
5'-HO-*BamHI*-primer	1	2	3
5'-HO-*BsrGI*-primer	4	5	6
5'-HO-*NcoI*-primer	7	8	9

matrix 3	5'-HO-*BamHI*-primer	5'-HO-*BsrGI*-primer	5'-HO-*SpeI*-primer
5'-HO-*BamHI*-primer	1	2	3
5'-HO-*BsrGI*-primer	4	5	6
5'-HO-*SpeI*-primer	7	8	9

matrix 4	5'-HO-*BamHI*-primer	5'-HO-*BsrGI*-primer	5'-HO-*AflII*-primer
5'-HO-*BamHI*-primer	1	2	3
5'-HO-*BsrGI*-primer	4	5	6
5'-HO-*AflII*-primer	7	8	9

matrix 5	5'-HO-*BamHI*-primer	5'-HO-*HindIII*-primer	5'-HO-*NcoI*-primer
5'-HO-*BamHI*-primer	1	2	3
5'-HO-*HindIII*-primer	4	5	6
5'-HO-*NcoI*-primer	7	8	9

matrix 6	5'-HO-*BamHI*-primer	5'-HO-*HindIII*-primer	5'-HO-*SpeI*-primer
5'-HO-*BamHI*-primer	1	2	3
5'-HO-*HindIII*-primer	4	5	6
5'-HO-*SpeI*-primer	7	8	9

matrix 7	5'-HO-*BamHI*-primer	5'-HO-*HindIII*-primer	5'-HO-*AflII*-primer
5'-HO-*BamHI*-primer	1	2	3
5'-HO-*HindIII*-primer	4	5	6
5'-HO-*AflII*-primer	7	8	9

matrix 8	5'-HO-*BamHI*-primer	5'-HO-*NcoI*-primer	5'-HO-*SpeI*-primer
5'-HO-*BamHI*-primer	1	2	3
5'-HO-*NcoI*-primer	4	5	6
5'-HO-*SpeI*-primer	7	8	9

matrix 9	5'-HO-*BamHI*-primer	5'-HO-*NcoI*-primer	5'-HO-*AflII*-primer
5'-HO-*BamHI*-primer	1	2	3
5'-HO-*NcoI*-primer	4	5	6
5'-HO-*AflII*-primer	7	8	9

matrix 10	5'-HO-*BamHI*-primer	5'-HO-*SpeI*-primer	5'-HO-*AflII*-primer
5'-HO-*BamHI*-primer	1	2	3
5'-HO-*SpeI*-primer	4	5	6
5'-HO-*AflII*-primer	7	8	9

matrix 11	5'-HO-*BsrGI*-primer	5'-HO-*HindIII*-primer	5'-HO-*NcoI*-primer
5'-HO-*BsrGI*-primer	1	2	3
5'-HO-*HindIII*-primer	4	5	6
5'-HO-*NcoI*-primer	7	8	9

matrix 12	5'-HO-*BsrGI*-primer	5'-HO-*HindIII*-primer	5'-HO-*SpeI*-primer
5'-HO-*BsrGI*-primer	1	2	3
5'-HO-*HindIII*-primer	4	5	6
5'-HO-*SpeI*-primer	7	8	9

matrix 13	5'-HO-*BsrGI*-primer	5'-HO-*HindIII*-primer	5'-HO-*AflII*-primer
5'-HO-*BsrGI*-primer	1	2	3
5'-HO-*HindIII*-primer	4	5	6
5'-HO-*AflII*-primer	7	8	9

matrix 14	5'-HO-*BsrGI*-primer	5'-HO-*NcoI*-primer	5'-HO-*SpeI*-primer
5'-HO-*BsrGI*-primer	1	2	3
5'-HO-*NcoI*-primer	4	5	6
5'-HO-*SpeI*-primer	7	8	9

matrix 15	5'-HO-*BsrGI*-primer	5'-HO-*NcoI*-primer	5'-HO-*AflII*-primer
5'-HO-*BsrGI*-primer	1	2	3
5'-HO-*NcoI*-primer	4	5	6
5'-HO-*AflII*-primer	7	8	9

matrix 16	5'-HO-*BsrGI*-primer	5'-HO-*SpeI*-primer	5'-HO-*AflII*-primer
5'-HO-*BsrGI*-primer	1	2	3
5'-HO-*SpeI*-primer	4	5	6
5'-HO-*AflII*-primer	7	8	9

matrix 17	5'-HO-*HindIII*-primer	5'-HO-*NcoI*-primer	5'-HO-*SpeI*-primer
5'-HO-*HindIII*-primer	1	2	3
5'-HO-*NcoI*-primer	4	5	6
5'-HO-*SpeI*-primer	7	8	9

matrix 18	5'-HO-*HindIII*-primer	5'-HO-*NcoI*-primer	5'-HO-*AflII*-primer
5'-HO-*HindIII*-primer	1	2	3
5'-HO-*NcoI*-primer	4	5	6
5'-HO-*AflII*-primer	7	8	9

matrix 19	5'-HO-*HindIII*-primer	5'-HO-*SpeI*-primer	5'-HO-*AflII*-primer
5'-HO-*HindIII*-primer	1	2	3
5'-HO-*SpeI*-primer	4	5	6
5'-HO-*AflII*-primer	7	8	9

matrix 20	5'-HO-*NcoI*-primer	5'-HO-*SpeI*-primer	5'-HO-*AflII*-primer
5'-HO-*NcoI*-primer	1	2	3
5'-HO-*SpeI*-primer	4	5	6
5'-HO-*AflII*-primer	7	8	9

Example TRSPA-2 test matrices for set 2 - *EcoRI*, *BspHI*, *BglII*, *XbaI*, *Acc65I* and *ApaLI*

matrix 21	5'-HO-*EcoRI*-primer	5'-HO-*BspHI*-primer	5'-HO-*BglII*-primer
5'-HO-*EcoRI*-primer	1	2	3
5'-HO-*BspHI*-primer	4	5	6
5'-HO-*BglII*-primer	7	8	9

matrix 22	5'-HO-*EcoRI*-primer	5'-HO-*BspHI*-primer	5'-HO-*XbaI*-primer
5'-HO-*EcoRI*-primer	1	2	3
5'-HO-*BspHI*-primer	4	5	6
5'-HO-*XbaI*-primer	7	8	9

matrix 23	5'-HO-*EcoRI*-primer	5'-HO-*BspHI*-primer	5'-HO-*Acc65I*-primer
5'-HO-*EcoRI*-primer	1	2	3
5'-HO-*BspHI*-primer	4	5	6
5'-HO-*Acc65I*-primer	7	8	9

matrix 24	5'-HO-*EcoRI*-primer	5'-HO-*BspHI*-primer	5'-HO-*ApaLI*-primer
5'-HO-*EcoRI*-primer	1	2	3
5'-HO-*BspHI*-primer	4	5	6
5'-HO-*ApaLI*-primer	7	8	9

matrix 25	5'-HO-*EcoRI*-primer	5'-HO-*BglII*-primer	5'-HO-*XbaI*-primer
5'-HO-*EcoRI*-primer	1	2	3
5'-HO-*BglII*-primer	4	5	6
5'-HO-*XbaI*-primer	7	8	9

matrix 26	5'-HO-*EcoRI*-primer	5'-HO-*BglII*-primer	5'-HO-*Acc65I*-primer
5'-HO-*EcoRI*-primer	1	2	3
5'-HO-*BglII*-primer	4	5	6
5'-HO-*Acc65I*-primer	7	8	9

matrix 27	5'-HO-*EcoRI*-primer	5'-HO-*BglII*-primer	5'-HO-*ApaLI*-primer
5'-HO-*EcoRI*-primer	1	2	3
5'-HO-*BglII*-primer	4	5	6
5'-HO-*ApaLI*-primer	7	8	9

matrix 28	5'-HO-*EcoRI*-primer	5'-HO-*XbaI*-primer	5'-HO-*Acc65I*-primer
5'-HO-*EcoRI*-primer	1	2	3
5'-HO-*XbaI*-primer	4	5	6
5'-HO-*Acc65I*-primer	7	8	9

matrix 29	5'-HO-*EcoRI*-primer	5'-HO-*XbaI*-primer	5'-HO-*ApaLI*-primer
5'-HO-*EcoRI*-primer	1	2	3
5'-HO-*XbaI*-primer	4	5	6
5'-HO-*ApaLI*-primer	7	8	9

matrix 30	5'-HO-*EcoRI*-primer	5'-HO-*Acc65I*-primer	5'-HO-*ApaLI*-primer
5'-HO-*EcoRI*-primer	1	2	3
5'-HO-*Acc65I*-primer	4	5	6
5'-HO-*ApaLI*-primer	7	8	9

matrix 31	5'-HO-*BspHI*-primer	5'-HO-*BglII*-primer	5'-HO-*XbaI*-primer
5'-HO-*BspHI*-primer	1	2	3
5'-HO-*BglII*-primer	4	5	6
5'-HO-*XbaI*-primer	7	8	9

matrix 32	5'-HO-*BspHI*-primer	5'-HO-*BglII*-primer	5'-HO-*Acc65I*-primer
5'-HO-*BspHI*-primer	1	2	3
5'-HO-*BglII*-primer	4	5	6
5'-HO-*Acc65I*-primer	7	8	9

matrix 33	5'-HO-*BspHI*-primer	5'-HO-*BglII*-primer	5'-HO-*ApaLI*-primer
5'-HO-*BspHI*-primer	1	2	3
5'-HO-*BglII*-primer	4	5	6
5'-HO-*ApaLI*-primer	7	8	9

matrix 34	5'-HO-*BspHI*-primer	5'-HO-*XbaI*-primer	5'-HO-*Acc65I*-primer
5'-HO-*BspHI*-primer	1	2	3
5'-HO-*XbaI*-primer	4	5	6
5'-HO-*Acc65I*-primer	7	8	9

matrix 35	5'-HO-*BspHI*-primer	5'-HO-*XbaI*-primer	5'-HO-*ApaLI*-primer
5'-HO-*BspHI*-primer	1	2	3
5'-HO-*XbaI*-primer	4	5	6
5'-HO-*ApaLI*-primer	7	8	9

matrix 36	5'-HO-*BspHI*-primer	5'-HO-*Acc65I*-primer	5'-HO-*ApaLI*-primer
5'-HO-*BspHI*-primer	1	2	3
5'-HO-*Acc65I*-primer	4	5	6
5'-HO-*ApaLI*-primer	7	8	9

matrix 37	5'-HO-*BglII*-primer	5'-HO-*XbaI*-primer	5'-HO-*Acc65I*-primer
5'-HO-*BglII*-primer	1	2	3
5'-HO-*XbaI*-primer	4	5	6
5'-HO-*Acc65I*-primer	7	8	9

matrix 38	5'-HO-*BglII*-primer	5'-HO-*XbaI*-primer	5'-HO-*ApaLI*-primer
5'-HO-*BglII*-primer	1	2	3
5'-HO-*XbaI*-primer	4	5	6
5'-HO-*ApaLI*-primer	7	8	9

matrix 39	5'-HO-*BglII*-primer	5'-HO-*Acc65I*-primer	5'-HO-*ApaLI*-primer
5'-HO-*BglII*-primer	1	2	3
5'-HO-*Acc65I*-primer	4	5	6
5'-HO-*ApaLI*-primer	7	8	9

matrix 40	5'-HO-*XbaI*-primer	5'-HO-*Acc65I*-primer	5'-HO-*ApaLI*-primer
5'-HO-*XbaI*-primer	1	2	3
5'-HO-*Acc65I*-primer	4	5	6
5'-HO-*ApaLI*-primer	7	8	9

Hybridisation patterns

There are six possible hybridisation patterns for a given probe fragment from inter-population perfectly matched duplex depletion enrichment with a (3x3) test matrix in TRSPA-1 analysis.

There are also six possible hybridisation patterns for a given probe fragment from inter-population perfectly matched duplex depletion enrichment with a (3x3) test matrix in TRSPA-2 analysis.

We can denote these patterns as follows

TRSPA-1 analysis

Pattern a (1 only)

matrix #	5'-HO-X-primer	5'-HO-Y-primer	5'-HO-Z-primer
5'-biotin-X-primer	1	2	3
5'-biotin-Y-primer	4	5	6
5'-biotin-Z-primer	7	8	9

Pattern b (2 and 4)

matrix #	5'-HO-X-primer	5'-HO-Y-primer	5'-HO-Z-primer
5'-biotin-X-primer	1	2	3
5'-biotin-Y-primer	4	5	6
5'-biotin-Z-primer	7	8	9

Pattern c (3 and 7)

matrix #	5'-HO-X-primer	5'-HO-Y-primer	5'-HO-Z-primer
5'-biotin-X-primer	1	2	3
5'-biotin-Y-primer	4	5	6
5'-biotin-Z-primer	7	8	9

Pattern d (5 only)

matrix #	5'-HO-X-primer	5'-HO-Y-primer	5'-HO-Z-primer
5'-biotin-X-primer	1	2	3
5'-biotin-Y-primer	4	5	6
5'-biotin-Z-primer	7	8	9

Pattern e (6 and 8)

matrix #	5'-HO-X-primer	5'-HO-Y-primer	5'-HO-Z-primer
5'-biotin-X-primer	1	2	3
5'-biotin-Y-primer	4	5	
5'-biotin-Z-primer	7		9

Pattern f (9 only)

matrix #	5'-HO-X-primer	5'-HO-Y-primer	5'-HO-Z-primer
5'-biotin-X-primer	1	2	3
5'-biotin-Y-primer	4	5	6
5'-biotin-Z-primer	7	8	

TRSPA-2 analysis

Pattern a (row 1 and column 1)

matrix #	5'-HO-X-primer	5'-HO-Y-primer	5'-HO-Z-primer
5'-HO-X-primer			
5'-HO-Y-primer		5	6
5'-HO-Z-primer		8	9

Pattern b (2 and 4)

matrix #	5'-HO-X-primer	5'-HO-Y-primer	5'-HO-Z-primer
5'- HO -X-primer	1		3
5'- HO -Y-primer		5	6
5'- HO -Z-primer	7	8	9

Pattern c (3 and 7)

matrix #	5'-HO-X-primer	5'-HO-Y-primer	5'-HO-Z-primer
5'- HO -X-primer	1	2	
5'- HO -Y-primer	4	5	6
5'- HO -Z-primer		8	9

Pattern d (row 2 and column 2)

matrix #	5'-HO-X-primer	5'-HO-Y-primer	5'-HO-Z-primer
5'- HO -X-primer	1		3
5'- HO -Y-primer			
5'- HO -Z-primer	7		9

Pattern e (6 and 8)

matrix #	5'-HO-X-primer	5'-HO-Y-primer	5'-HO-Z-primer
5'- HO -X-primer	1	2	3
5'- HO -Y-primer	4	5	
5'- HO -Z-primer	7		9

Pattern f (row 3 and column 3)

matrix #	5'-HO-X-primer	5'-HO-Y-primer	5'-HO-Z-primer
5'- HO -X-primer	1	2	
5'- HO -Y-primer	4	5	
5'- HO -Z-primer			

Signatures

If we have 40 such (3x3) test matrices and we denote the possible hybridisation patterns a, b, c, d, e and f, then we can write the overall TRSPA hybridisation signature as follows

Matrix	Pattern	matrix	pattern
1	a, b, c, d, e or f	21	a, b, c, d, e or f
2	a, b, c, d, e or f	22	a, b, c, d, e or f
3	a, b, c, d, e or f	23	a, b, c, d, e or f
4	a, b, c, d, e or f	24	a, b, c, d, e or f
5	a, b, c, d, e or f	25	a, b, c, d, e or f
6	a, b, c, d, e or f	26	a, b, c, d, e or f
7	a, b, c, d, e or f	27	a, b, c, d, e or f
8	a, b, c, d, e or f	28	a, b, c, d, e or f
9	a, b, c, d, e or f	29	a, b, c, d, e or f
10	a, b, c, d, e or f	30	a, b, c, d, e or f
11	a, b, c, d, e or f	31	a, b, c, d, e or f
12	a, b, c, d, e or f	32	a, b, c, d, e or f
13	a, b, c, d, e or f	33	a, b, c, d, e or f
14	a, b, c, d, e or f	34	a, b, c, d, e or f
15	a, b, c, d, e or f	35	a, b, c, d, e or f
16	a, b, c, d, e or f	36	a, b, c, d, e or f
17	a, b, c, d, e or f	37	a, b, c, d, e or f
18	a, b, c, d, e or f	38	a, b, c, d, e or f
19	a, b, c, d, e or f	39	a, b, c, d, e or f
20	a, b, c, d, e or f	40	a, b, c, d, e or f

Aspect (III)

In another aspect the invention provides a nucleic acid characterisation method which comprises presenting to the set of arrays as defined above a nucleic acid fragment of interest under hybridisation conditions, and observing a pattern of hybridisation. Preferably, a plurality of nucleic acid fragments of interest are separately presented to the set of arrays, and the resulting patterns of hybridisation are compared. Preferably, the plurality of nucleic acid fragments of interest are drawn from the mixture of DNA fragments, enriched in fragments that are characteristic of a phenotype of interest, as described under the invention (1) above.

Thus in this aspect the invention provides a method of identifying fragments of DNA that are characteristic of a phenotype of

interest, which method comprises recovering, cloning and amplifying individual DNA fragments from the mixture of DNA fragments obtained under invention (1) above, presenting the individual DNA fragments to the set of arrays as defined under hybridisation conditions, observing a pattern of hybridisation generated by each individual DNA fragment, and subjecting to further investigation any two or more individual DNA fragments whose hybridisation patterns are similar or identical.

TRSPA signatures and whole genome association studies

After a given number of cycles of inter-population perfectly matched duplex depletion, phenotype determining fragments will be enriched but will not be entirely free from 'noise' fragments. Noise may result from unequal allelic frequencies for certain SNPs between the two populations. Noise will also result from the presence of somatic mutations in the cells used to prepare DNA fragments and from the use of polymerase chain reaction in some of the embodiments of the current invention.

For the preferred embodiment, DNA is prepared from a library of clones (either genomic clones or cDNA clones) - with inserts derived from the individual(s) and propagated in some appropriate host and cloning vector system. Restriction enzyme fragmentation is used prior to cloning and polymerase chain reaction amplification is used to prepare the DNA for comparison in fragmented form. Priming sites within the vector sequence flanking the cloned restriction enzyme fragmented inserts are employed for one or more cycles of polymerase chain reaction amplification of the fragmented DNA of interest. The primers used for polymerase chain reaction amplification of the fragmented DNA of interest are again used after the phenotype-determining fragment enrichment process to 'rescue' and clone the enriched fragments. Cloned enriched fragments are colony purified, picked into appropriate storage containers, catalogued and archived, DNA probes are prepared from these single clones and are

individually hybridised against 180 spot TRSPA arrays as above. TRSPA signatures are determined for many colonies. Most noise fragments will be random in nature and will thus be randomly distributed amongst the 6^{20} = $3.7 \cdot 10^{15}$ possible types of TRSPA signature. Phenotype-determining signal fragments, however, will be those where repeat TRSPA signatures are obtained as more and more colonies are sampled. The more frequently the repetition of a given TRSPA signature occurs per unit colonies sampled, the greater the signal to noise ratio and the more successful has been the enrichment.

A great many of the steps in this process are amenable to high throughput automation - enabling very large numbers of single colony TRSPA signatures to be determined with ease and extending the power of the current invention to cases where signal to noise ratios are beyond current approaches.

Statistical correlations (associations) can initially be drawn between TRSPA signatures and phenotype. The clones giving rise to a particular TRSPA signature showing a useful association with a phenotype of interest can then be sequenced in order to determine at a DNA sequence level the association(s) with the phenotype of interest. Such associations have future predictive values for the phenotype of interest, knowing the genotype and will be of great value in medicine and pharmacogenetics.

If the genome of interest is wholly or partially sequenced, we can also *in silico* restrict the DNA with all n enzymes, calculate the expected signature for each fragment and pattern match these expected signatures with the observed signature (taking into account any loss or gain of restriction sites due to polymorphic variation compared to the reference sequence) to immediately identify the fragment of interest within a gene, genomic region, chromosome or whole genome. This latter method will be of great value in those cases where a great many phenotype-determining fragments are obtained and repeat signatures are rare or unobtained. The

clustering of phenotype-determining fragments to adjacent DNA regions thus gives an association between those genomic regions and the phenotype of interest.

Aspect (IV)

In yet another aspect the invention provides a double-stranded DNA molecule having the sequence a-A-b-B...X-y-Y-z where A, B...X and Y are unique restriction sites for n different restriction endonuclease enzymes, and a, b...y, z denotes distances in base pairs, characterised in that each fragment, obtainable by cutting the DNA molecule by means of any one or more up to n of the restriction enzymes, has a different length from every other fragment.

An example totally diagnostic internal control DNA which allows both the extent and exact nature of any example set 1 (or example set 2) 6 bp cutter partial digestion to be unambiguously determined for inter-population perfectly matched duplex depletion or TRSPA restriction

In both of the above schemes, it is important that limit digestion products are obtained. Monitoring the extent of partial digestion resulting from multi-enzymatic restriction and determining precisely which enzymes have failed to cut is a task of great importance.

If we have up to six enzymes for DNA digestion - let us label these A, B, C, D, E and F. We need to somehow determine that these have all cut to completion during the fragmentation stage for inter-population perfectly matched duplex depletion and also for the digestion step prior to adaptor ligation in TRSPA fabrication. If any of the enzymes have failed to cut to completion, we need to know which ones and to what degree in order to effectively rectify the problem.

The structure of the internal control DNA

If we construct a double stranded DNA molecule with the

following structure

end---t---A---u---B---v---C---w---D---x---E---y---F---z---end

where the A, B, C, D, E and F denote the sites for restriction enzyme cutting and t, u, v, w, x, y and z denote distances in base pairs.

This internal control DNA is either uniformly pre-labelled and added to the DNA of interest at an appropriate concentration prior to restriction or is Southern blot probed with a complementary sequence not found in the DNA of interest after restriction.

All six enzymes can cut in only one way.

One enzyme can fail to cut in ${}^{6}C_{1} = 6$ ways, these are: A, B, C, D, E or F failing to cut.

Two enzymes can fail to cut in ${}^{6}C_{2} = 15$ ways, these are: AB, AC, AD, AE, AF, BC, BD, BE, BF, CD, CE, CF, DE, DF or EF failing to cut.

Three enzymes can fail to cut in ${}^{6}C_{3} = 20$ ways, these are: ABC, ABD, ABE, ABF, ACD, ACE, ACF, ADE, ADF, AEF, BCD, BCE, BCF, BDE, BDF, BEF, CDE, CDF, CEF or DEF failing to cut.

Four enzymes can fail to cut in ${}^{6}C_{4} = 15$ ways, these are: ABCD, ABCE, ABCF, ABDE, ABDF, ABEF, ACDE, ACDF, ACEF, ADEF, BCDE, BCDF, BCEF, BDEF or CDEF failing to cut.

Five enzymes can fail to cut in ${}^{6}C_{5} = 6$ ways, these are: ABCDE, ABCDF, ABCEF, ABDEF, ACDEF or BCDEF failing to cut.

All six enzymes can fail to cut in only one way.

Each of the above possibilities will generate one or more fragments from the internal control DNA. If each possible fragment has a discernible size from any other, then we can determine exactly which enzymes have cut and which have not from the size distribution of the fragments generated. The task is therefore to design such a DNA molecule.

Example simulations

Seven simulations are given below - varying the size of inter-site fragments. Criteria for a successful outcome include the following

1. The inter-fragment spacing should be greater for larger fragments (so as to aid electrophoretic resolution).

2. All possible fragments should be unambiguously resolvable in size from each-other.

3. Size gaps between bands comprising different numbers of inter-site units should be greater than the size gaps between bands comprising the same number of inter-site units.

4. The size gaps and size spread from largest to smallest fragment should be electrophoretically compatible.

Simulation 1

Inter-site fragment sizes (in bp)

End---A	A---B	B---C	C---D	D---E	E---F	F---end
80	90	100	110	120	130	140

Possible digestion products obtained (in bp)

	Fragment length (in bp)	size gap to the next smallest fragment (in bp)
one unit fragments		
end---A	80	
A---B	90	10
B---C	100	10
C---D	110	10
D---E	120	10
E---F	130	10
F---end	140	10
two unit fragments		
end---B	170	30
A---C	190	20
B---D	210	20
C---E	230	20
D---F	250	20
E---end	270	20
three unit fragments		
end---C	270	0
A---D	300	30
B---E	330	30
C---F	360	30
D---end	390	30
four unit fragments		
end---D	380	-10
A---E	420	40
B---F	460	40
C---end	500	40
five unit fragments		
end---E	500	0
A---F	550	50
B---end	600	50
six unit fragments		
end---F	630	30
B---end	690	60
seven unit fragments		
end---end	770	80

Spread = 690 bp

Simulation 2

Inter-site fragment sizes (in bp)

end---A	A---B	B---C	C---D	D---E	E---F	F---end
90	100	110	120	130	140	150

Possible digestion products obtained (in bp)

	fragment length (in bp)	size gap to the next smallest fragment (in bp)
one unit fragments		
end---A	90	
A---B	100	10
B---C	110	10
C---D	120	10
D---E	130	10
E---F	140	10
F---end	150	10
two unit fragments		
end---B	190	40
A---C	210	20
B---D	230	20
C---E	250	20
D---F	270	20
E---end	290	20
three unit fragments		
end---C	300	10
A---D	330	30
B---E	360	30
C---F	390	30
D---end	420	30
four unit fragments		
end---D	420	0
A---E	460	40
B---F	500	40
C---end	540	40
five unit fragments		
end---E	550	10
A---F	600	50
B---end	650	50
six unit fragments		
end---F	690	40
B---end	750	60
seven unit fragments		
end---end	840	90

Spread = 750 bp

Simulation 3

Inter-site fragment sizes (in bp)

end---A	A---B	B---C	C---D	D---E	E---F	F---end
100	110	120	130	140	150	160

Possible digestion products obtained (in bp)

	fragment length (in bp)	size gap to the next smallest fragment (in bp)
one unit fragments		
end---A	100	
A---B	110	10
B---C	120	10
C---D	130	10
D---E	140	10
E---F	150	10
F---end	160	10
two unit fragments		
end---B	210	50
A---C	230	20
B---D	250	20
C---E	270	20
D---F	290	20
E---end	310	20
three unit fragments		
end---C	330	20
A---D	360	30
B---E	390	30
C---F	420	30
D---end	450	30
four unit fragments		
end---D	460	10
A---E	500	40
B---F	540	40
C---end	580	40
five unit fragments		
end---E	600	20
A---F	650	50
B---end	700	50
six unit fragments		
end---F	750	50
B---end	810	60
seven unit fragments		
end---end	910	100

Spread = 810 bp

Simulation 4

Inter-site fragment sizes (in bp)

end---A	A---B	B---C	C---D	D---E	E---F	F---end
110	120	130	140	150	160	170

Possible digestion products obtained (in bp)

	fragment length (in bp)	size gap to the next smallest fragment (in bp)
one unit fragments		
end---A	110	
A---B	120	10
B---C	130	10
C---D	140	10
D---E	150	10
E---F	160	10
F---end	170	10
two unit fragments		
end---B	230	60
A---C	250	20
B---D	270	20
C---E	290	20
D---F	310	20
E---end	330	20
three unit fragments		
end---C	360	30
A---D	390	30
B---E	420	30
C---F	450	30
D---end	480	30
four unit fragments		
end---D	500	20
A---E	540	40
B---F	580	40
C---end	620	40
five unit fragments		
end---E	650	30
A---F	700	50
B---end	750	50
six unit fragments		
end---F	810	60
B---end	870	60
seven unit fragments		
end---end	980	110

Spread = 870 bp

Simulation 5

Inter-site fragment sizes (in bp)

end---A	A---B	B---C	C---D	D---E	E---F	F---end
120	130	140	150	160	170	180

Possible digestion products obtained (in bp)

	fragment length (in bp)	size gap to the next smallest fragment (in bp)
one unit fragments		
end---A	120	
A---B	130	10
B---C	140	10
C---D	150	10
D---E	160	10
E---F	170	10
F---end	180	10
two unit fragments		
end---B	250	70
A---C	270	20
B---D	290	20
C---E	310	20
D---F	330	20
E---end	350	20
three unit fragments		
end---C	390	40
A---D	420	30
B---E	450	30
C---F	480	30
D---end	510	30
four unit fragments		
end---D	540	30
A---E	580	40
B---F	620	40
C---end	660	40
five unit fragments		
end---E	700	40
A---F	750	50
B---end	800	50
six unit fragments		
end---F	870	70
B---end	930	60
seven unit fragments		
end---end	1050	120

Spread = 930 bp

Simulation 6

Inter-site fragment sizes (in bp)

end---A	A---B	B---C	C---D	D---E	E---F	F---end
130	140	150	160	170	180	190

Possible digestion products obtained (in bp)

	fragment length (in bp)	size gap to the next smallest fragment (in bp)
one unit fragments		
end---A	130	
A---B	140	10
B---C	150	10
C---D	160	10
D---E	170	10
E---F	180	10
F---end	190	10
two unit fragments		
end---B	270	80
A---C	290	20
B---D	310	20
C---E	330	20
D---F	350	20
E---end	370	20
three unit fragments		
end---C	420	50
A---D	450	30
B---E	480	30
C---F	510	30
D---end	540	30
four unit fragments		
end---D	580	40
A---E	620	40
B---F	660	40
C---end	700	40
five unit fragments		
end---E	750	50
A---F	800	50
B---end	850	50
six unit fragments		
end---F	930	80
B---end	990	60
seven unit fragments		
end---end	1120	130

Spread = 990 bp

Simulation 7

Inter-site fragment sizes (in bp)

end---A	A---B	B---C	C---D	D---E	E---F	F---end
140	150	160	170	180	190	200

5

Possible digestion products obtained (in bp)

	fragment length (in bp)	size gap to the next smallest fragment (in bp)
one unit fragments		
end---A	140	
A---B	150	10
B---C	160	10
C---D	170	10
D---E	180	10
E---F	190	10
F---end	200	10
two unit fragments		
end---B	290	90
A---C	310	20
B---D	330	20
C---E	350	20
D---F	370	20
E---end	390	20
three unit fragments		
end---C	450	60
A---D	480	30
B---E	510	30
C---F	540	30
D---end	570	30
four unit fragments		
end---D	620	50
A---E	660	40
B---F	700	40
C---end	740	40
five unit fragments		
end---E	800	60
A---F	850	50
B---end	900	50
six unit fragments		
end---F	990	90
B---end	1050	60
seven unit fragments		
end---end	1190	140

Spread = 1050 bp

According to the above criteria for success, simulation 7 above clearly fulfils all of the requirements.

An example totally diagnostic internal control DNA which allows both the extent and exact nature of any example set 1 (or example set 2) 4 bp cutter partial digestion to be unambiguously determined for inter-population perfectly matched duplex depletion

If we have up to three enzymes for DNA digestion - let us label these A, B and C. We need to somehow determine that these have all cut to completion during the fragmentation stage for inter-population perfectly matched duplex depletion. If any of the enzymes have failed to cut to completion, we need to know which ones and to what degree in order to effectively rectify the problem.

The structure of the internal control DNA

If we construct a double stranded DNA molecule with the following structure

end---t---A---u---B---v---C---w---end

where the A, B and C denote the sites for restriction enzyme cutting and t, u, v and w denote distances in base pairs.

This internal control DNA is uniformly pre-labelled and added to the DNA of interest at an appropriate concentration prior to restriction or is Southern blot probed with a complementary sequence not found in the DNA of interest after restriction.

All three enzymes can cut in only one way.

One enzyme can fail to cut in ${}^3C_1 = 3$ ways, these are: A, B or C failing to cut.

Two enzymes can fail to cut in ${}^3C_2 = 3$ ways, these are: AB, AC or BC failing to cut.

All three enzymes can fail to cut in only one way.

Each of the above possibilities will generate one or more fragments from the internal control DNA. If each possible fragment has a discernible size from any other (and from any of the fragments in simulation 7 above for the panel of up to 6 enzymes), then we can determine exactly which enzymes have cut and which have not from the size distribution of the fragments generated. The task is therefore to design such a DNA molecule.

Example simulations

Six simulations are given below - varying the size of inter-site fragments. Criteria for a successful outcome included the following:

1. The inter-fragment spacing should be greater for larger fragments (so as to aid electrophoretic resolution).
2. All possible fragments should be unambiguously resolvable in size from each-other.
3. Size gaps between bands comprising different numbers of inter-site units should ideally be greater than the size gaps between bands comprising the same number of inter-site units.
4. The size gaps and size spread from largest to smallest fragment should be electrophoretically compatible.
5. The largest fragment obtained should ideally be smaller than the smallest fragment obtained in simulation 7 above for the panel of up to six enzymes.

Simulation 1

Inter-site fragment sizes (in bp)

end---A	A---B	B---C	C---end
10	20	30	40

Possible digestion products obtained (in bp)

	fragment length (in bp)	size gap to the next smallest fragment (in bp)
one unit fragments		
end---A	10	
A---B	20	10
B---C	30	10
C---end	40	10
two unit fragments		
end---B	30	-10
A---C	50	20
B---end	70	20
three unit fragments		
end---C	60	-10
A---end	90	30
four unit fragments		
end---end	100	10

Spread = 90 bp

Simulation 2

Inter-site fragment sizes (in bp)

end---A	A---B	B---C	C---end
15	25	35	45

Possible digestion products obtained (in bp)

	fragment length (in bp)	size gap to the next smallest fragment (in bp)
one unit fragments		
end---A	15	
A---B	25	10
B---C	35	10
C---end	45	10
two unit fragments		
end---B	40	-5
A---C	60	20
B---end	80	20
three unit fragments		
end---C	75	-5
A---end	105	30
four unit fragments		
end---end	120	15

Spread = 105 bp

Simulation 3

Inter-site fragment sizes (in bp)

end---A	A---B	B---C	C---end
20	30	40	50

Possible digestion products obtained (in bp)

	fragment length (in bp)	size gap to the next smallest fragment (in bp)
one unit fragments		
end---A	20	
A---B	30	10
B---C	40	10
C---end	50	10
two unit fragments		
end---B	50	0
A---C	70	20
B---end	90	20
three unit fragments		
end---C	90	0
A---end	120	30
four unit fragments		
end---end	140	20

Spread = 120 bp

Simulation 4

Inter-site fragment sizes (in bp)

end---A	A---B	B---C	C---end
25	35	45	55

Possible digestion products obtained (in bp)

	fragment length (in bp)	size gap to the next smallest fragment (in bp)
one unit fragments		
end---A	25	
A---B	35	10
B---C	45	10
C---end	55	10
two unit fragments		
end---B	60	5
A---C	80	20
B---end	100	20
three unit fragments		
end---C	105	5
A---end	135	30
four unit fragments		
end---end	160	25

Spread = 135 bp

Simulation 5

Inter-site fragment sizes (in bp)

end---A	A---B	B---C	C---end
20	25	30	35

Possible digestion products obtained (in bp)

	fragment length (in bp)	size gap to the next smallest fragment (in bp)
one unit fragments		
end---A	20	
A---B	25	5
B---C	30	5
C---end	35	5
two unit fragments		
end---B	45	10
A---C	55	10
B---end	65	10
three unit fragments		
end---C	75	10
A---end	90	15
four unit fragments		
end---end	110	20

Spread = 90 bp

Simulation 6

Inter-site fragment sizes (in bp)

end---A	A---B	B---C	C---end
25	30	35	40

Possible digestion products obtained (in bp)

	fragment length (in bp)	size gap to the next smallest fragment (in bp)
one unit fragments		
end---A	25	
A---B	30	5
B---C	35	5
C---end	40	5
two unit fragments		
end---B	55	15
A---C	65	10
B---end	75	10
three unit fragments		
end---C	90	15
A---end	105	15
four unit fragments		
end---end	130	25

Spread = 105 bp

According to the above criteria for success, simulation 6 clearly fulfils all of the requirements.

Example determination of the entire set of internal control DNA limit and partial digestion patterns for a panel of up to six restriction enzymes

For the example simulation 7, the entire set of internal control DNA limit and partial digestion patterns for a panel of up to six restriction enzymes can be determined as below.

Inter-site fragment sizes (in bp)

end---A	A---B	B---C	C---D	D---E	E---F	F---end
140	150	160	170	180	190	200

Possible digestion products obtained

All six enzymes can cut in only one way.

failed digests	bp	bp	bp	bp	bp	bp	bp
none	140	150	160	170	180	190	200

One enzyme can fail to cut in ${}^{6}C_{1} = 6$ ways, these are: A, B, C, D, E or F failing to cut.

failed digest	bp	bp	bp	bp	bp	bp
A	290	160	170	180	190	200
B	140	310	170	180	190	200
C	140	150	330	180	190	200
D	140	150	160	350	190	200
E	140	150	160	170	370	200
F	140	150	160	170	180	390

Two enzymes can fail to cut in ${}^{6}C_{2}$ = 15 ways, these are: AB, AC, AD, AE, AF, BC, BD, BE, BF, CD, CE, CF, DE, DF or EF failing to cut.

failed digests	bp	bp	bp	bp	bp
AB	450	170	180	190	200
AC	290	330	180	190	200
AD	290	160	350	190	200
AE	290	160	170	370	200
AF	290	160	170	180	390
BC	140	480	180	190	200
BD	140	310	350	190	200
BE	140	310	170	370	200
BF	140	310	170	180	390
CD	140	150	510	190	200
CE	140	150	330	370	200
CF	140	150	330	180	390
DE	140	150	160	540	200
DF	140	150	160	350	390
EF	140	150	160	170	570

Three enzymes can fail to cut in $^{6}C_{3}$ = 20 ways, these are: ABC, ABD, ABE, ABF, ACD, ACE, ACF, ADE, ADF, AEF, BCD, BCE, BCF, BDE, BDF, BEF, CDE, CDF, CEF or DEF failing to cut.

failed digests	bp	bp	bp	bp
ABC	620	180	190	200
ABD	450	350	190	200
ABE	450	170	370	200
ABF	450	170	180	390
ACD	290	510	190	200
ACE	290	330	370	200
ACF	290	330	180	390
ADE	290	160	540	200
ADF	290	160	350	390
AEF	290	160	170	570
BCD	140	660	190	200
BCE	140	480	370	200
BCF	140	480	180	390
BDE	140	310	540	200
BDF	140	310	350	390
BEF	140	310	170	570
CDE	140	150	700	200
CDF	140	150	510	390
CEF	140	150	330	570
DEF	140	150	160	740

5

Four enzymes can fail to cut in ${}^{6}C_{4}$ = 15 ways, these are: ABCD, ABCE, ABCF, ABDE, ABDF, ABEF, ACDE, ACDF, ACEF, ADEF, BCDE, BCDF, BCEF, BDEF or CDEF failing to cut.

failed digests	bp	bp	bp
ACCD	800	190	200
ABCE	620	370	200
ABCF	620	180	390
ABDE	450	540	200
ABDF	450	350	390
ABEF	450	170	570
ACDE	290	700	200
ACDF	290	510	390
ACEF	290	330	570
ADEF	290	160	740
BCDE	140	850	200
BCDF	140	660	390
BCEF	140	480	570
BDEF	140	310	740
CDEF	140	150	900

5

Five enzymes can fail to cut in ${}^{6}C_{5} = 6$ ways, these are: ABCDE, ABCDF, ABCEF, ABDEF, ACDEF or BCDEF failing to cut.

failed digests	bp	bp
ABCDE	990	200
ABCDF	800	390
ABCEF	620	570
ABDEF	450	740
ACDEF	290	900
BCDEF	140	1050

All six enzymes can fail to cut in only one way.

failed digests	bp
all	1190

Example determination of the entire set of internal control DNA limit and partial digestion patterns for a panel of up to three restriction enzymes

For the example simulation 6, the entire set of internal control DNA limit and partial digestion patterns for a panel of up to three restriction enzymes can be determined as below.

Inter-site fragment sizes (in bp)

end---A	A---B	B---C	C---end
25	30	35	40

Possible digestion products obtained

All three enzymes can cut in only one way.

failed digests	bp	bp	bp	bp
none	25	30	35	40

One enzyme can fail to cut in $^3C_1$ = 3 ways, these are: A, B or C failing to cut.

failed digests	bp	bp	bp
A	55	35	40
B	25	65	40
C	25	30	75

Two enzymes can fail to cut in $^3C_2$ = 3 ways, these are: AB, AC or BC failing to cut.

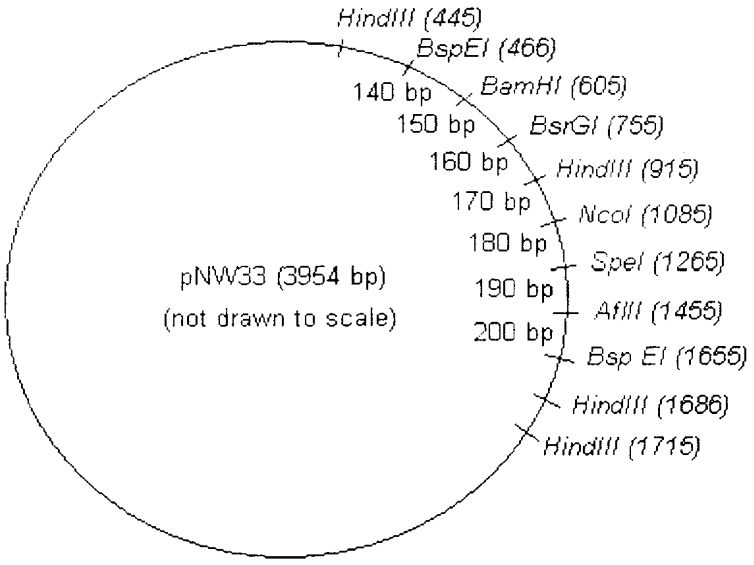
failed digests	bp	bp
AB	90	40
AC	55	75
BC	25	105

All three enzymes can fail to cut in only one way.

failed digests	bp
all	130

Example 1a – The digestion internal control plasmid for the 6 bp cutter set 1 TRSPA enzymes *BamHI, BsrGI, HindIII, NcoI, SpeI,* and *AflII*

The plasmid pNW33 (shown below) was constructed to contain an insert with all of the 6 bp cutter TRSPA enzyme sites.



BspEI sites define the outer ends of the 140 bp and the 200 bp fragments. The full sequence for pNW33 is shown below:

tcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaagcggatgcc
gggagcagacaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggctggcttaactatgcggcatcagagca
gattgtactgagagtgcaccatatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgccattc
gccattcaggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgt
gctgcaaggcgattaagttgggtaacgccagggttttcccagtcacgacgttgtaaaacgacggccagtgaattcgagctcgg
taccgggccccccctcgaggtcgacggtatcgataagcttgatcgcagctggtaatccggacgcccgcgtcgaagatgttgg
ggtgttgtaacaatatcgattccaattcagcgggggccacctgatatcctttgtatttaattaaagacttcaagcggtcaactatga
agaagtgttcgtcttcgtcccagtaaggatccgcactttgaattttgtaatcctgaagggatcgtaaaaacagctcttcttcaaatct
atacattaagacgactcgaaatccacatatcaaatatccgagtgtagtaaacattccaaaaccgtgatggaatggaacaaca
cttaaatgtacaccctggtaatccgttttagaatccatgataataattttctggattattggtaatttttttgcacgttcaaaatttttgca
acccctttttggaaacaaacactacggtaggctgcgaaatgttcatactgttgagcaattcacgttcattataagcttttcactgcat
acgacgattctgtgatttgtattcagcccatatcgtttcatagcttctgccaaccgaacggacatttcgaagtattccgcgtacgtg

atgttcacctcgatatgtgcatctgtaaaagcaattgttccaggaaccagggcgtatctcttcatagccatggaatacgccttttc
agtgttgcgatgctaatgccgttacaaatattccgagcaccaagaatggctgcgcgcttgcctggtacttgacgtcgtatttgacg
gggtccttgagaaagtatttaaactggaacacaatctgaggaatgatcaaagcaaccaacgccaacgcataataactagtg
caataccaagacctcccaataatagcacccagacttgtgtaataacctctggctctgatattgctccagatggaattggacgat
atggctcattaattgcgtcgatatctctatcataccagtcgttgattgtctgtgtatagccagtaagacaaggaccagacatcatca
tgcaaagaatcgcttaagcccttcttggcctttatgaggatctctctgatttttcttgcgtcgagttttccggtaagacctttcggtactt
cgtccacaaacacaactcctccgcgcaacttttcgcggttgttacttgactggccacgtaatccacgatctcttttttccgtcatcgt
ctttccgtgctccaaaacaacaacggcggcgggtccggattaccagctgcgatcaagcttatcgataccgtcgacctcgacct
gcaggcatgcaagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacga
gccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttc
cagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttc
cgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggtt
atccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaag
gccgcgttgctggcgtttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaa
acccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccg
gatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcg
ctccaagctgggctgtgtgcacgaacccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaaccc
ggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacag
agttcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcgga
aaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgc
agaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggatt
ttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagt
aaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgact
ccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctca
ccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctcc
atccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggc
atcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgt
gcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagc
actgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgta
tgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcatt
ggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaac
tgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataag
ggcgacacggaaatgttgaatactcatactcttccttttcaatattattgaagcatttatcagggttattgtctcatgagcggataca
tatttgaatgtatttagaaaaataaacaaatagggttccgcgcacatttccccgaaaagtgccacctgacgtctaagaaacc
attattatcatgacattaacctataaaaataggcgtatcacgaggcccttcgtc

The insert was introduced as a four fragment ligation of 140 (KpnI-BamHI), 150 / 160 (BamHI-HindIII), 170 / 180 (HindIII-SpeI) and 190 / 200 (SpeI-XhoI) into KpnI / SalI digested pUC19c DNA (Genbank X02514). The SalI and XhoI sites were lost as a result of the joining of their compatible sticky ends. The insert can be removed with BspEI or PvuII.

The insert region of 1190 bp and the short flanking regions to the vector junctions were sequenced twice in each direction in order to establish the plasmid sequence. In addition, a total of 63 analytical restriction digests and one minus-enzyme control were performed as detailed in the following table:

		Expected band sizes						
Digest	Well	bp	bp	bp	bp	bp	bp	bp
Uncut	1	1190						
BamHI	2	140	1050					
BsrGI	3	290	900					
HindIII	4	450	740					
NcoI	5	620	570					
SpeI	6	800	390					
AflII	7	990	200					
BamHI BsrGI	8	140	150	900				
BamHI HindIII	9	140	310	740				
BamHI NcoI	10	140	480	570				
BamHI SpeI	11	140	660	390				
BamHI AflII	12	140	850	200				
BsrGI HindIII	13	290	160	740				
BsrGI NcoI	14	290	330	570				
BsrGI SpeI	15	290	510	390				
BsrGI AflII	16	290	700	200				
HindIII NcoI	17	450	170	570				
HindIII SpeI	18	450	350	390				
HindIII AflII	19	450	540	200				
NcoI SpeI	20	620	180	390				
NcoI AflII	21	620	370	200				
SpeI AflII	22	800	190	200				
BamHI BsrGI HindIII	23	140	150	160	740			
BamHI BsrGI NcoI	24	140	150	330	570			
BamHI BsrGI SpeI	25	140	150	510	390			
BamHI BsrGI AflII	26	140	150	700	200			

BamHI HindIII NcoI	27	140	310	170	570			
BamHI HindIII SpeI	28	140	310	350	390			
BamHI HindIII AflII	29	140	310	540	200			
BamHI NcoI SpeI	30	140	480	180	390			
BamHI NcoI AflII	31	140	480	370	200			
BamHI SpeI AflII	32	140	660	190	200			
BsrGI HindIII NcoI	33	290	160	170	570			
BsrGI HindIII SpeI	34	290	160	350	390			
BsrGI HindIII AflII	35	290	160	540	200			
BsrGI NcoI SpeI	36	290	330	180	390			
BsrGI NcoI AflII	**37**	**290**	**330**	**370**	**200**			
BsrGI SpeI AflII	38	290	510	190	200			
HindIII NcoI SpeI	39	450	170	180	390			
HindIII NcoI AflII	40	450	170	370	200			
HindIII SpeI AflII	41	450	350	190	200			
NcoI SpeI AflII	42	620	180	190	200			
BamHI BsrGI HindIII NcoI	43	140	150	160	170	570		
BamHI BsrGI HindIII SpeI	44	140	150	160	350	390		
BamHI BsrGI HindIII AflII	45	140	150	160	540	200		
BamHI BsrGI NcoI SpeI	46	140	150	330	180	390		
BamHI BsrGI NcoI AflII	47	140	150	330	370	200		
BamHI BsrGI SpeI AflII	48	140	150	510	190	200		
BamHI HindIII NcoI SpeI	**49**	**140**	**310**	**170**	**180**	**390**		
BamHI HindIII NcoI AflII	50	140	310	170	370	200		
BamHI HindIII SpeI AflII	51	140	310	350	190	200		
BamHI NcoI SpeI AflII	52	140	480	180	190	200		
BsrGI HindIII NcoI SpeI	53	290	160	170	180	390		
BsrGI HindIII NcoI AflII	54	290	160	170	370	200		
BsrGI HindIII SpeI AflII	55	290	160	350	190	200		
BsrGI NcoI SpeI AflII	56	290	330	180	190	200		
HindIII NcoI SpeI AflII	57	450	170	180	190	200		
BamHI BsrGI HindIII NcoI SpeI	58	140	150	160	170	180	390	
BamHI BsrGI HindIII NcoI AflII	59	140	150	160	170	370	200	
BamHI BsrGI HindIII SpeI AflII	60	140	150	160	350	190	200	
BamHI BsrGI NcoI SpeI AflII	**61**	**140**	**150**	**330**	**180**	**190**	**200**	
BamHI HindIII NcoI SpeI AflII	62	140	310	170	180	190	200	
BsrGI HindIII NcoI SpeI AflII	63	290	160	170	180	190	200	
BamHI BsrGI HindIII NcoI SpeI AflII	64	140	150	160	170	180	190	200

All of these digests produced the expected fragment patterns on agarose gel electrophoresis and a number of these, shaded in the table above, are shown in figure 1. The restriction digests illustrated in figure 1

were carried out using the following conditions:

Digest 1 (Minus enzyme control)

10 µl	100 µg/ml internal control plasmid DNA in 1x NEB buffer #2 + 100 µg/ml BSA
90 µl	1x NEB buffer #2 + 100 µg/ml BSA

Digest 5

15 µl	0.33 U/µl *NcoI* in 1x NEB buffer #2 + 100 µg/ml BSA
10 µl	100 µg/ml internal control plasmid DNA in 1x NEB buffer #2 + 100 µg/ml BSA
75 µl	1x NEB buffer #2 + 100 µg/ml BSA

Digest 9

15 µl	0.33 U/µl *BamHI* in 1x NEB buffer #2 + 100 µg/ml BSA
15 µl	0.33 U/µl *HindIII* in 1x NEB buffer #2 + 100 µg/ml BSA
10 µl	100 µg/ml internal control plasmid DNA in 1x NEB buffer #2 + 100 µg/ml BSA
60 µl	1x NEB buffer #2 + 100 µg/ml BSA

Digest 37

15 µl	0.33 U/µl *BsrGI* in 1x NEB buffer #2 + 100 µg/ml BSA
15 µl	0.33 U/µl *NcoI* in 1x NEB buffer #2 + 100 µg/ml BSA
15 µl	0.33 U/µl *AflII* in 1x NEB buffer #2 + 100 µg/ml BSA
10 µl	100 µg/ml internal control plasmid DNA in 1x NEB buffer #2 + 100 µg/ml BSA
45 µl	1x NEB buffer #2 + 100 µg/ml BSA

Digest 49

15 µl	0.33 U/µl *BamHI* in 1x NEB buffer #2 + 100 µg/ml BSA
15 µl	0.33 U/µl *HindIII* in 1x NEB buffer #2 + 100 µg/ml BSA
15 µl	0.33 U/µl *NcoI* in 1x NEB buffer #2 + 100 µg/ml BSA
15 µl	0.33 U/µl *SpeI* in 1x NEB buffer #2 + 100 µg/ml BSA
10 µl	100 µg/ml internal control plasmid DNA in 1x NEB buffer #2 + 100 µg/ml BSA
30 µl	1x NEB buffer #2 + 100 µg/ml BSA

Digest 61

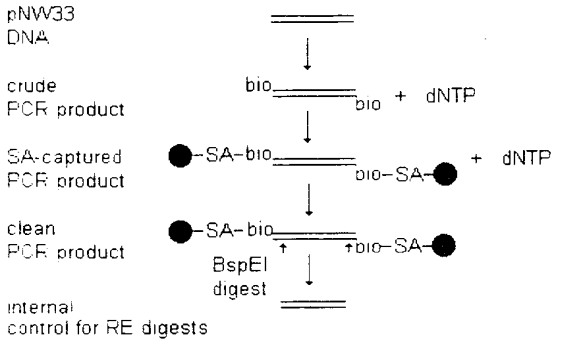
15 µl	0.33 U/µl *BamHI* in 1x NEB buffer #2 + 100 µg/ml BSA
15 µl	0.33 U/µl *BsrGI* in 1x NEB buffer #2 + 100 µg/ml BSA
15 µl	0.33 U/µl *NcoI* in 1x NEB buffer #2 + 100 µg/ml BSA
15 µl	0.33 U/µl *SpeI* in 1x NEB buffer #2 + 100 µg/ml BSA
15 µl	0.33 U/µl *AflII* in 1x NEB buffer #2 + 100 µg/ml BSA
10 µl	100 µg/ml internal control plasmid DNA in 1x NEB buffer #2 + 100 µg/ml BSA
15 µl	1x NEB buffer #2 + 100 µg/ml BSA

All 64 restriction digests were incubated at 37°C for 6 hours and samples were then electrophoresed on 2.5 % FMC MetaPhor agarose gels as illustrated in figure 1. The lanes marked M contain mixed samples from all of the digests 1-64 and these lanes were used as size markers after confirmation of the fragment sizes against Stratagene Kb ladder markers.

When spiked into a genomic DNA digest the internal control restriction fragment pattern produced is indicative of both the degree of digestion and the nature of any partial restriction at less than limit digestion. It is possible to deduce from the bands present which, if any, of the enzymes have failed to cut and therefore to take action to correct this before the DNA is used in a subsequent analysis or enrichment procedure.

Preparation of *BspEI*-released internal control DNA on a large scale

The method used for generating internal control PCR product DNA (free from contaminating dNTPs) is depicted in the following figure:



Primers and PCR

20 µM BIO140UP

5' biotin-CGCAGCTGGTAATCCGGACGCCCGCGTCGAAGATGTT 3'

20 µM BIO200DOWN

5' biotin-CGCAGCTGGTAATCCGGACCCGCCGCCGTTGTTGTT 3'

Bulk PCR amplification (192x 100 µl reactions) was carried out according to the conditions described below:

	per 20 ml	final conditions
100 mM dATP	20 µl	100 µM
100 mM dCTP	20 µl	100 µM
100 mM dGTP	20 µl	100 µM
100 mM dTTP	20 µl	100 µM
10x PCR buffer	2 ml	1x
25 mM $MgCl_2$	400 µl	1/50th volume
1 ng/µl pNW33	400 µl	2 ng per PCR
20 µM BIO140UP	200 µl	20 pmol per PCR
20 µM BIO200DOWN	200 µl	20 pmol per PCR
Taq DNA polymerase	100 µl	2.5 U per PCR

The master mix was rapidly dispensed into PCR tubes. Thermal cycling was initiated using the following parameters: 30 cycles of 97°C for 1 min, 50°C for 2 min, and 72°C for 3 min; 72°C for 5 min; and then 4°C.

After PCR amplification, samples were pooled and subjected to capture of biotinylated PCR product termini and BspEI release of internal control DNA.

Capture of biotinylated PCR product termini and *BspEI* release of internal control DNA

All separations were carried out using a Dynal MPC-1 separator (Dynal, product #12001).

20 ml of pooled PCR reaction were mixed with 20 ml of Dynabeads M-280 in 20 mM tris-HCl (pH 7.4), 2 mM EDTA (pH 8.0), 2 M NaCl. The tube was incubated at room temperature for 1 hour with rolling.

The Dynabeads were then washed four times with 20 ml of 10 mM tris-HCl (pH 7.4), 1 mM EDTA (pH 8.0), 1 M NaCl.

BspEI digestion was carried out for 1 hour at 37°C in 2 ml of 0.5 U/µl BspEI in 1x NEB buffer #3. The digest was then ethanol precipitated by the addition of 0.1 volume of 3 M sodium acetate (pH 5.2) and 2.5 volumes of ethanol, chilling to -20°C and then centrifugation. Pellets were rinsed with 70 % ethanol prior to redissolution in 500 µl of 1x TE buffer.

1 µl of the BspEI-released internal control DNA was mixed with 10 µl of 50 % glycerol AGE loading dye and electrophoresed on a 1.5 % agarose gel to confirm that the size of the purified DNA was in accordance with that expected.

High molecular weight genomic DNA digestion in the presence of internal control DNA with a dilution series of a mixture of the six set 1 6 bp cutters

High molecular weight canine genomic DNA was first mixed with a dilution series of a mixture of the six set 1 6 bp cutters – each at the same number of units. Aliquots were then removed and mixed with the BspEI-released internal control DNA described above. 20 µg of canine genomic DNA was digested with 0.25, 0.025, 0.0025, 0.00025, 0.000025, 0.0000025 and 0 U/µl BamH1 / BsrGI / HindIII / NcoI / SpeI / AflII in 200 µl. 0.4 µg of canine genomic DNA and 1 µl of BspEI-released internal control DNA were digested with 0.25, 0.025, 0.0025, 0.00025, 0.000025, 0.0000025 and 0 U/µl BamH1 / BsrGI / HindIII / NcoI / SpeI / AflII in 4 µl. Vistra Green staining was used to monitor the extent of internal control DNA digestion and high molecular weight DNA digestion. The fragment pattern from the internal control DNA is diagnostic of both the degree of digestion and the nature of any partial restriction at less than limit digestion.

A premix of restriction enzymes, buffer and BSA was prepared as detailed below:

component	µl
100 U/µl *BamHI*	1 µl
10 U/µl *BsrGI*	10 µl
40 U/µl *HindIII*	2.5 µl
50 U/µl *NcoI*	2 µl
50 U/µl *SpeI*	2 µl
20 U/µl *AflII*	5 µl
10x NEB buffer #2	4 µl
100x BSA	0.4 µl
water	13. 1 µl

Total volume = 40 µl

The premix therefore contained each restriction enzyme at

2.5 U/µl in 1x NEB buffer #2 and 1x BSA. Serial 10-fold dilutions of this premix were then prepared in 1x NEB buffer #2 and 1x BSA.

Premixes of canine genomic DNA, buffer and BSA were also prepared as detailed below:

Component	µl
1 mg/ml canine genomic DNA	20 µl
10x NEB buffer #2	8 µl
100x BSA	0.8 µl
Water	51. 2 µl

These premixes therefore contain canine genomic DNA at 0.25 mg/ml in 1x NEB buffer #2 and 1x BSA.

Canine genomic DNA and restriction enzyme mixes were then set up as follows:

tube	component #1	component #2
1	80 µl of canine genomic DNA at 0.25 mg/ml in 1x NEB buffer #2 and 1x BSA	20 µl of each restriction enzyme at 2.5 U/µl in 1x NEB buffer #2 and 1x BSA
2	80 µl of canine genomic DNA at 0.25 mg/ml in 1x NEB buffer #2 and 1x BSA	20 µl of each restriction enzyme at 0.25 U/µl in 1x NEB buffer #2 and 1x BSA
3	80 µl of canine genomic DNA at 0.25 mg/ml in 1x NEB buffer #2 and 1x BSA	20 µl of each restriction enzyme at 0.025 U/µl in 1x NEB buffer #2 and 1x BSA
4	80 µl of canine genomic DNA at 0.25 mg/ml in 1x NEB buffer #2 and 1x BSA	20 µl of each restriction enzyme at 0.0025 U/µl in 1x NEB buffer #2 and 1x BSA
5	80 µl of canine genomic DNA at 0.25 mg/ml in 1x NEB buffer #2 and 1x BSA	20 µl of each restriction enzyme at 0.00025 U/µl in 1x NEB buffer #2 and 1x BSA
6	80 µl of canine genomic DNA at 0.25 mg/ml in 1x NEB buffer #2 and 1x BSA	20 µl of each restriction enzyme at 0.000025 U/µl in 1x NEB buffer #2 and 1x BSA
7	80 µl of canine genomic DNA at 0.25 mg/ml in 1x NEB buffer #2 and 1x BSA	20 µl of 1x NEB buffer #2 and 1x BSA

2 µl aliquots were then removed from tubes 1-7 and added to 1 µl of BspEI-released internal control DNA and 1 µl of 2x NEB buffer #2 and 2x BSA to give samples 1-7ic. Samples 1-7ic were then overlaid with 50 µl of mineral oil in order to prevent evaporation.

The remaining volume from tubes 1-7 was then added to 100 µl of 1x NEB buffer #2 and 1x BSA.

All samples were finally incubated at 37°C overnight.

After digestion, 20 µl of digests 1-7 were mixed with 10 µl of 50 % glycerol AGE loading dye and 4 µl of digests 1-7ic were mixed with 2 µl of 50 % glycerol AGE loading dye. Digests in loading dye were then electrophoresed on a 2.5 % MetaPhor agarose gel in 1x TBE. The gel was stained for 60 min in 500 ml of 1x TBE containing 50 µl of Vistra Green. The stained gel was finally imaged on a Fluorimager with the following settings: a 488 nm laser; a 570 DF 30 filter; a PMT setting of 700 V; 200 µm resolution; and low sensitivity.

Restriction digests - final conditions

Restriction digests 1-7 were therefore performed under the following conditions:

tube	µg canine genomic DNA	U *BamH1*	U *BsrGI*	U *HindIII*	U *NcoI*	U *SpeI*	U *AflII*	NEB buffer #2	µg/ml BSA
1	20	50	50	50	50	50	50	1x	100
2	20	5	5	5	5	5	5	1x	100
3	20	0.5	0.5	0.5	0.5	0.5	0.5	1x	100
4	20	0.05	0.05	0.05	0.05	0.05	0.05	1x	100
5	20	0.005	0.005	0.005	0.005	0.005	0.005	1x	100
6	20	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	1x	100
7	20	0	0	0	0	0	0	1x	100

Total volume = 200 µl

Likewise, restriction digests 1-7ic were performed under the following conditions:

tube	µg canine genomic DNA	µl IC DNA	U *BamH1*	U *BsrGI*	U *HindIII*	U *NcoI*	U *SpeI*	U *AflII*	NEB buffer #2	µg/ml BSA
1ic	0.4	1	1	1	1	1	1	1	1x	100
2ic	0.4	1	0.1	0.1	0.1	0.1	0.1	0.1	1x	100
3ic	0.4	1	0.01	0.01	0.01	0.01	0.01	0.01	1x	100
4ic	0.4	1	0.001	0.001	0.001	0.001	0.001	0.001	1x	100
5ic	0.4	1	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	1x	100
6ic	0.4	1	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	1x	100
7ic	0.4	1	0	0	0	0	0	0	1x	100

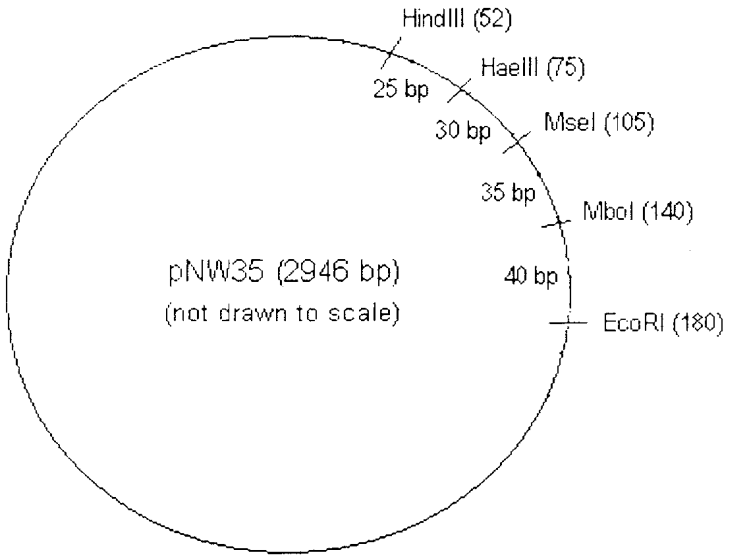
IC = BspEI-released internal control DNA

Total volume = 4 µl

The results are shown in figure 2.

Example 1b – The digestion internal control plasmid for the 4 bp cutter set 1 TRSPA enzymes *HaeIII, MboI,* and *MseI*

The plasmid pNW35 (shown below) was constructed to contain an insert with all of the 4 bp cutter TRSPA enzyme sites.



HindIII and EcoRI sites define the outer ends of the 25 bp and the 40 bp fragments. The sequence of pNW35 is shown below with the inserted region shown in bold type:

atgaccatgattacgccaagctctaatacgactcactatagggaa**agcttccggacgtctcaggctaatgttggcccacc**
gacgttccacgatggggcgctcttaagggcttagaccctcgtcgggagtatttctgtgatctggcgacactcacgcg
agaagtcattaccggcgatatgaattcactggccgtcgttttacaacgtcgtgactgggaaaaccctggcgttacccaactta
atcgccttgcagcacatccccctttcgccagctggcgtaatagcgaagaggcccgcaccgatcgcccttcccaacagttgcg
cagcctgaatggcgaatgggaaattgtaaacgttaatattttgttaaaattcgcgttaaatttttgttaaatcagctcatttttaacca
ataggccgaaatcggcaaaatcccttataaatcaaaagaatagaccgagatagggttgagtgttgttccagtttggaacaag
agtccactattaaagaacgtggactccaacgtcaaagggcgaaaaaccgtctatcagggcgatggcccactacgtgaacc
atcaccctaatcaagtttttggggtcgaggtgccgtaaagcactaaatcggaaccctaaagggagcccccgatttagagcttg
acggggaaagccggcgaacgtggcgagaaaggaagggaagaaagcgaaaggagcgggcgctagggcgctggcaa
gtgtagcggtcacgctgcgcgtaaccaccacacccgccgcgcttaatgcgccgctacagggcgcgtcaggtggcacttttcg
gggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaat
gcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattccctttttgcggcattttgccttcctgttt
ttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctc
aacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggt

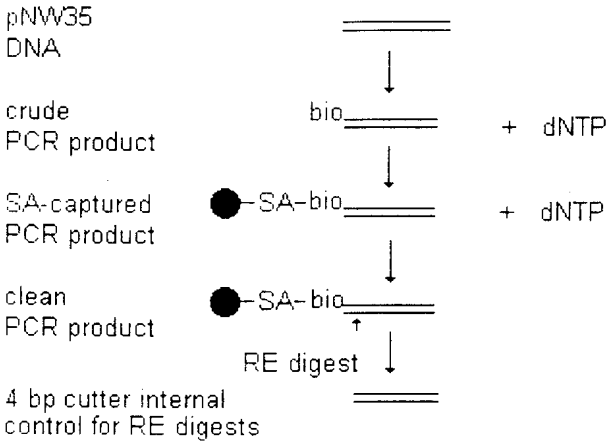
attatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtc
acagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggcc
aacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgat
cgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgtt
gcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcag
gaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcatt
gcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacga
aatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgat
ttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttttgataatctcatgaccaaaatcccttaacgtgagttttcgttcc
actgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaa
aaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactcttttccgaaggtaactggcttcagcagagcg
cagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctc
tgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataa
ggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatac
ctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcg
gaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgactt
gagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggcc
ttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgc
tcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctc
tccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaat
taatgtgagttagctcactcattaggcaccccaggctttacactttatgcttccggctcgtatgttgtgtggaattgtgagcggataa
caatttcacacaggaaacagct

The 4 bp internal control plasmid for restriction enzymes HaeIII, MboI, and MseII was prepared by the insertion of a synthetic 130 bp fragment into HindIII / EcoRI-digested pMOSblue (Amersham Pharmacia Biotech).

The insert region of 130 bp was sequenced twice in each direction in order to establish the plasmid sequence. The presence of the restriction sites and the mobility of the fragments released were also checked by restriction digestion and polyacrylamide gel electrophoresis.

Preparation of the internal control spike DNA from the plasmid

The following illustration describes the strategy used for the preparation of EcoRI-released internal control DNA:



Primers and PCR

U-19 mer bio primer	5' bio-GTTTTCCCAGTCACGACGT 3'
ICPCR(F) primer	5' TCCGGACGTCTCAGGCTAATGTT 3'

Bulk PCR amplification (96x 100 µl reactions, repeated to give 192 reactions in total) was carried out according to the conditions described below (all volumes are in µl):

	x1 (µl)	x100 (µl)	Final conditions
Water	84.4	8440	-
10x PCR buffer	10	1000	1x
25 mM dNTP mix	0.8	80	200 µM
20 µM T7 promoter-bio primer	1	100	20 pmol per PCR
20 µM U19 mer-bio primer	1	100	20 pmol per PCR
1 ng/µl pNW35	2	200	2 ng per PCR
5 U/µl *Taq* polymerase	0.8	80	4U per PCR
Total	**100**	**10000**	-

The PCR master mix was rapidly dispensed into 96 PCR tubes. Thermal cycling was initiated using the following parameters: 94°C for 2 min; 50 °C for 2 min; 29 cycles of 72°C for 2 min, 94°C for 45 sec, and 50°C for 1 min; 72°C for 8 min; and then 4°C.

Capture of biotinylated PCR product termini and *EcoRI*- release of internal control DNA

All separations were carried out using a Dynal MPC-1 separator (Dynal, product #12001).

20 ml of pooled PCR reaction were mixed with 20 ml of Dynabeads M-280 in 20 mM tris-HCl (pH 7.4), 2 mM EDTA (pH 8.0), 2 M NaCl. The tube was incubated at room temperature for 1 hour with mixing on a Denley Spiramix 5.

The Dynabeads were then washed four times with 20 ml of 10 mM tris-HCl (pH 7.4), 1 mM EDTA (pH 8.0), 1 M NaCl. A fifth wash was performed in 20 ml of 1x buffer M.

EcoRI-digestion was carried out for 1 hour at 37°C in 5 ml of 0.25 U/µl EcoRI in 1x buffer M. The digest was then divided into ten 500 µl aliquot parts. Each aliquot was ethanol precipitated by the addition of 1 µl of See DNA, 0.1 volume of 3 M sodium acetate (pH 5.2), and 2.5 volumes of ethanol. The precipitations were mixed and chilled to 0°C on ice for 30 minutes and then centrifugation at 20,000 maxRCF for 10 minutes. The pellets were rinsed with 70 % ethanol before dissolving in a total of 500 µl of 1x TE buffer.

A 1 µl aliquot of the EcoRI-released internal control DNA was electrophoresed on an 8 % polyacrylamide gel to confirm that the size of the purified DNA was in accordance with that expected.

High molecular weight genomic DNA digestion in the presence of the 4 bp cutter internal control DNA with a dilution series of the enzymes *HaeIII, MseI,* and *MboI*

High molecular weight human placental DNA was mixed with a dilution series of a mixture of the three set 1 4 bp cutter restriction endonucleases – each at the same number of units. Aliquots were then removed and mixed with the EcoRI-released internal control DNA. 3.6 µg of Human placental DNA (Sigma) was digested with 0.5 U/µl, 0.1 U/µl, 0.02 U/µl, 0.004 U/µl, and 0 U/µl each of HaeIII, MboI, and MseI in a total volume of 36 µl. 0.4 µg of placental DNA and 1 µl of EcoRI-released internal control DNA were digested with 0.5 U/µl, 0.1 U/µl, 0.02 U/µl, 0.004 U/µl, and 0 U/µl each of HaeIII, MboI, and MseI in a total volume of 4 µl. Vistra Green (Amersham Pharmacia Biotech) staining was used to monitor the extent of internal control DNA digestion and high molecular weight DNA digestion.

A premix of restriction enzymes (5 U/µl each enzyme), buffer, and BSA was prepared as described below:

Component	µl
50 U/µl *HaeIII*	1
20 U/µl *MseI*	2.5
25 U/µl *MboI*	2
10x NEB buffer #2	1
10x BSA	1
Water	2.5
Total	**10**

Serial 5-fold dilutions of the premix were prepared in 1x NEB buffer #2 and 1x BSA.

A 6x premix of human placental DNA, buffer, and BSA were also prepared as described below:

Component	6x mix (µl)	per reaction (µl)
1mg/ml human placental DNA	24	4
10x NEB buffer #2	9.6	1.6
10x BSA	9.6	1.6
Water	52.8	8.8

This premix contained placental DNA at 0.25 mg/ml in 1x NEB buffer #2 and 1x BSA.

Human placental DNA and restriction mixes were then set up as follows:

Tube	Component #1	Component #2
1	16 µl of placental DNA at 0.25 mg/ml in 1x NEB buffer #2 and 1x BSA	4 µl of each restriction enzyme at 5 U/µl in 1x NEB buffer #2 and 1x BSA
2	16 µl of placental DNA at 0.25 mg/ml in 1x NEB buffer #2 and 1x BSA	4 µl of each restriction enzyme at 1 U/µl in 1x NEB buffer #2 and 1x BSA
3	16 µl of placental DNA at 0.25 mg/ml in 1x NEB buffer #2 and 1x BSA	4 µl of each restriction enzyme at 0.2 U/µl in 1x NEB buffer #2 and 1x BSA
4	16 µl of placental DNA at 0.25 mg/ml in 1x NEB buffer #2 and 1x BSA	4 µl of each restriction enzyme at 0.04 U/µl in 1x NEB buffer #2 and 1x BSA
5	16 µl of placental DNA at 0.25 mg/ml in 1x NEB buffer #2 and 1x BSA	4 µl of 1x NEB buffer #2 and 1x BSA

2 µl aliquots were then removed from tubes 1-5 and added to 1 µl of EcoRI-released internal control DNA and 1 µl of 2x NEB buffer #2 and 2x BSA to give samples 1-5ic. Samples 1-5ic were then overlaid with 50 µl of mineral oil in order to prevent evaporation.

The remaining volume from tubes 1-5 was then added to 18 µl of 1x NEB buffer #2 and 1x BSA.

All samples were finally incubated at 37°C overnight.

Restriction digests - final conditions

Restriction digests 1-5 were therefore performed under the following conditions:

tube	µg human placental DNA	U *HaeIII*	U *MseI*	U *MboI*	NEB buffer #2	µg/ml BSA
1	3.6	20	20	20	1x	100
2	3.6	4	4	4	1x	100
3	3.6	0.8	0.8	0.8	1x	100
4	3.6	0.16	0.16	0.16	1x	100
5	3.6	0	0	0	1x	100

Total volume = 36 µl

Likewise, restriction digests 1-5ic were performed under the following conditions:

tube	µg human placental DNA	µl IC DNA	U *HaeIII*	U *MseI*	U *MboI*	NEB buffer #2	µg/ml BSA
1	0.4	1	2	2	2	1x	100
2	0.4	1	0.4	0.4	0.4	1x	100
3	0.4	1	0.08	0.08	0.08	1x	100
4	0.4	1	0.016	0.016	0.016	1x	100
5	0.4	1	0	0	0	1x	100

IC = *EcoRI*-released internal control DNA

Total volume = 4 µl

10 µl of digests 1-5 were each mixed with 3 µl of loading dye and 4 µl of digests 1-5ic were mixed with 1 µl of loading dye. To sample number 5ic, 180 ng of PCR molecular weight markers were added to serve as size standards. The band sizes for these markers are 50 bp, 100 bp, 200 bp, 300 bp, 400 bp, 500 bp, 525 bp, 700 bp, and 1000 bp. The digests

were then electrophoresed on an 8 % polyacrylamide gel in 1x TBE as described below:

10 ml of 10x TBE
19.5 ml of 40 % acrylamide
19 ml of 2 % methylene bisacrylamide
50.4 ml of water
1 ml of freshly prepared 10 % (w/v) ammonium persulphate
100 µl of TEMED

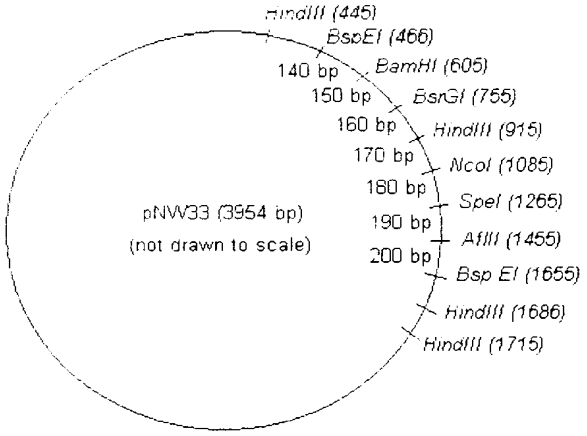
A Cambridge Electrophoresis Ltd. vertical protein electrophoresis unit was used with 1 mm plate spacing. The samples were electrophoresed at 30 mA for 2 hours in 1x TBE. The gel was then stained for 30 minutes in 500 ml of 1x TBE containing 50 µl of Vistra Green. The stained gel was imaged on a Fluorimager with the following settings: a 488 nm laser; a 570 DF 30 filter; a PMT setting of 700 V; 200 µm resolution; and low sensitivity.

The results are shown in figure 3.

Example 2

TRSPA-2 analysis with the pNW33 *BamHI*, *HindIII*, and *AflII* matrix and the pNW33 *HindIII*, *NcoI*, and *SpeI* matrix probed with the 140 bp *BspEI* to *BamHI* fragment from pNW33

As an example of the case where the probe hybridizes to an arrayed PCR fragment with a different restriction site at each end, the pNW33 BamHI, HindIII, and AflII matrix (matrix #7) was probed with a PCR product from the 140 bp BspEI (466) to BamHI (605) restriction fragment within pNW33. In this example, the probe binds to a 204 bp HindIII to BamHI PCR product derived from the 160 bp HindIII (445) to BamHI (605) restriction fragment within pNW33 (see below).



As an example of the case where the probe hybridizes to an arrayed PCR fragment with the same restriction site at each end, the pNW33 HindIII, NcoI, and SpeI matrix (matrix #17) was probed with a PCR product from the 140 bp BspEI (466) to BamHI (605) restriction fragment within pNW33. In this example, the probe binds to a 514 bp HindIII to HindIII PCR product derived from the 470 bp HindIII (445) to HindIII (915) restriction fragment within pNW33 (see above).

Oligonucleotides

BamHI short PCR primer	5' TGTAACGACACATTGCTGGATACC 3'
HindIII short PCR primer	5' ATATAACTCTCGCTCCTTGATAAC 3'
NcoI short PCR primer	5' AGGCGTCTGAGGCTGCGGCTATGG 3'
SpeI short PCR primer	5' AACCCGTCGCGACGAGAGTCTAAG 3'
AflII short PCR primer	5' GATATACGTGATATATTTTGATTG 3'
BamHI adaptor	5' pGATCGGTATCCAGCAATGTGTCGTTACA 3'
HindIII adaptor	5' pAGCTGTTATCAAGGAGCGAGAGTTATAT 3'
NcoI adaptor	5' pCATGCCATAGCCGCAGCCTCAGACGCCT 3'
SpeI adaptor	5' pCTAGCTTAGACTCTCGTCGCGACGGGTT 3'

AflII adaptor 5' pTTAACAATCAAAATATATCACGTATATC 3'

BamHI long PCR primer 5' TGTAACGACACATTGCTGGATACCGATCC 3'
HindIII long PCR primer 5' ATATAACTCTCGCTCCTTGATAACAGCTT 3'
NcoI long PCR primer 5' AGGCGTCTGAGGCTGCGGCTATGGCATGG 3'
SpeI long PCR primer 5' AACCCGTCGCGACGAGAGTCTAAGCTAGT 3'
AflII long PCR primer 5' GATATACGTGATATATTTTGATTGTTAAG 3'

Luc140down primer
5' GCGCTAGGGATCCTTACTGGGACGAAGACGAA 3'

Luc140up-bio primer
5' biotin-CGCAGCTGGTAATCCGGACGCCCGCGTCGAAGATGTT3'

Restriction digestion of pNW33

Restriction digests of pNW33 were set up by combining the following:

Matrix #7

450 µl	22 µg/ml pNW33 in 1x NEB buffer #2 + 100 µg/ml BSA
50 µl	0.6 U/µl *BamHI* in 1x NEB buffer #2 + 100 µg/ml BSA
50 µl	0.6 U/µl *HindIII* in 1x NEB buffer #2 + 100 µg/ml BSA
50 µl	0.6 U/µl *AflII* in 1x NEB buffer #2 + 100 µg/ml BSA

Matrix #17

450 µl	22 µg/ml pNW33 in 1x NEB buffer #2 + 100 µg/ml BSA
50 µl	0.6 U/µl *HindIII* in 1x NEB buffer #2 + 100 µg/ml BSA
50 µl	0.6 U/µl *NcoI* in 1x NEB buffer #2 + 100 µg/ml BSA
50 µl	0.6 U/µl *SpeI* in 1x NEB buffer #2 + 100 µg/ml BSA

The samples were then vortex mixed, and incubated at 37°C overnight.

Calf intestinal alkaline phosphatase (CIAP) digestion of pNW33 restriction digests

400 µl fractions of the 20 restriction digests (each containing 6.6 µg of digested pNW33) were ethanol precipitated by the addition of 1 µl of See DNA, 0.1 volume of 3 M sodium acetate (pH 5.2), and 2.5 volumes of ethanol, chilling to -20°C, and then centrifugation. Pellets were rinsed with 70 % ethanol prior to re-dissolution in 20 µl of 1x CIAP buffer containing 40 U of CIAP. The CIAP digests were carried out for 5 hours at 37°C and were then made up to 400 µl with 1x TE buffer.

Phenol extraction of pNW33 CIAP digests

The diluted digests were extracted with 400 µl of phenol and then ethanol precipitated as described above but with a 100 % ethanol wash after the 70 % ethanol wash. Samples were finally re-dissolved in 20 µl of TE buffer.

Annealing of short PCR primers to cognate adaptors

Short PCR primers and their cognate adaptors were annealed by adding 1 µl of 200 µM short PCR primer to 1 µl of 200 µM cognate adaptor in 20 µl of 50 mM NaCl, 1x TE buffer. The mixed oligonucleotides were overlaid with 30 µl of light mineral oil and were then heated to 90°C for 5 minutes followed by slow cooling to room temperature. The annealed short PCR primer / cognate adaptor complexes were then diluted with 1 ml of 1x TE buffer and stored frozen at -20°C.

Ligation to annealed short PCR primers and cognate adaptors

1 µl of phenol extracted pNW33 digest was used per ligation. Ligations were performed in 100 µl of 1x ligase buffer containing 1 mM ATP

and 10 µl (100 U) of T4 DNA ligase. 20 µl aliquots from the 1 ml of annealed short PCR primer / cognate adaptor complexes were added according to the following table.

#7	*BamHI / HindIII / AflII*
#17	*HindIII / NcoI / SpeI*

Ligation reactions were carried out for 24 hours at 16°C. Samples were then diluted to 1 ml with TE buffer and stored at -20°C.

The diluted ligation reactions were then further diluted 1 in 10 and 10 µl was used as PCR template per 100 µl reaction.

Restriction digestion of human placental DNA

1 U (~50 µg) of human placental DNA was digested overnight at 37°C with 100 U each of *BamHI*, *BsrGI*, *HindIII*, *NcoI*, *SpeI*, and *AflII* in 400 µl of 1x NEB buffer #2 containing 100 µg/ml BSA.

CIAP digestion of human placental DNA restriction digest

The digest was ethanol precipitated by the addition of 1 µl of See DNA, 0.1 volume of 3 M sodium acetate (pH 5.2), and 2.5 volumes of ethanol, chilling to -20°C, and then centrifugation. The pellet was rinsed with 70 % ethanol prior to re-dissolution in 50 µl of 1x CIAP buffer containing 100 U of CIAP. The CIAP digest was carried out for 5 hours at 37°C and was then made up to 400 µl with 1x TE buffer.

Phenol extraction of human placental DNA CIAP digest

The diluted CIAP digest was extracted with 400 µl of phenol and then ethanol precipitated as described above – again with a 100 % ethanol wash after the 70 % ethanol wash. The sample was finally re-dissolved in 10 µl of TE buffer.

Ligation to annealed short PCR primers and cognate adaptors

Short PCR primers were annealed to their cognate adaptors as described above.

The ligation to annealed short PCR primers and cognate adaptors was carried out in 100 µl of 1x ligase buffer with 1 mM ATP, 100 U of T4 DNA ligase and 10 µl of each of the six short PCR primer / cognate adaptor complexes as above.

The ligation reaction was carried out for 24 hours at 16°C. The sample was then diluted to 1 ml with 1x TE buffer and stored at -20°C.

0.2 µl was used as PCR template per 100 µl reaction.

PCR amplification conditions

An initial touch-down reaction was carried out in 50 µl of 1x PCR buffer with all four dNTPs at 200 µM and *Taq* DNA polymerase at 0.05 U/µl. Long PCR primers were used at 400 nM. 10 µl of pNW33 PCR template was used per reaction and 0.2 µl of human placental DNA PCR template was used per reaction. The samples were overlaid with 40 µl of light mineral oil and were touch-down thermocycled as described below:

98°C for 1 min, 72°C for 2 min, 72°C for 5 min
98°C for 1 min, 69°C for 2 min, 72°C for 5 min
98°C for 1 min, 66°C for 2 min, 72°C for 5 min
98°C for 1 min, 63°C for 2 min, 72°C for 5 min
98°C for 1 min, 60°C for 2 min, 72°C for 5 min

The main thermal cycling was then carried out in 100 µl of 1x PCR buffer with all four dNTPs at 200 µM and *Taq* DNA polymerase at 0.05 U/µl. Long PCR primers were used at 400 nM. The samples were subjected to 20 cycles of 98°C for 1 min, 60°C for 2 min, 72°C for 5 min and then 72°C for 10 min followed by chilling at 10°C and recovery from under oil.

Arraying onto nylon membranes

1 µl aliquots from the PCR amplified samples were spotted onto Hybond N+ nylon membranes. The membranes were then transferred to a stack of three sheets of 3MM paper, saturated with 0.4 M NaOH, and incubated for 10 minutes. The NaOH was rinsed away in 2x SSC and the membranes were used directly for the pre-hybridization.

Probe synthesis

A PCR master mix was prepared as follows (all volumes are in µl):

Master Mix Preparation	**140**
Water	183.6
10x PCR buffer	30
2 mM dATP	15
2 mM dGTP	15
2 mM dTTP	15
2 mM dCTP	3
140 probe template	1
Luc140up-bio	2.5
Luc140down	2.5
Taq DNA polymerase (5 U/µl)	2.4
Total	**270**

Five PCR reactions were performed. For each reaction, 5 µl of 33P-labelled dCTP was added to the PCR tube. The reactions were then made up to 50 µl by the addition of 45 µl of the master mix to each tube. Each reaction was gently mixed.

PCR cycling parameters

94°C	2 min	
50 °C	2 min	
72°C	2 min	} 25 cycles
94°C	45 sec	
50 °C	1 min	
72°C	8 min	
4°C	hold	

After thermal cycling, the five PCR amplifications were pooled. The 250 µl of labelled DNA was then used for the following magnetic bead purification procedure.

Capture of biotinylated PCR product and release of non-biotinylated strands

All separations were carried out using a Dynal MPC-4 separator (Dynal, product #12004).

250 µl of the pooled PCR was mixed with an equal volume of 10 mg/ml streptavidin-coated colloidal Fe3O4 particles in 20 mM tris-HCl (pH 7.4), 2 mM EDTA (pH 8.0), 2 M NaCl. The tube was incubated at room temperature for 1 hour with mixing on a Denley Orbital Mixer.

The streptavidin-coated colloidal Fe3O4 particles were then washed with 1 ml of 10 mM tris-HCl (pH 7.4), 1 mM EDTA (pH 8.0), 1 M NaCl. Three more identical washes were performed.

The washed streptavidin-coated colloidal Fe3O4 particles were incubated in 500 µl of 0.1 M NaOH for 10 minutes at room temperature. The supernatant was removed and added to 500 µl of 0.5 M

HEPES. The samples were then ethanol precipitated by the addition of 0.1 volume of 3 M sodium acetate (pH 5.2) and 2.5 volumes of ethanol, chilling to 0°C (on ice for 30 minutes), and then centrifugation at 20,000 maxRCF for 10 minutes. The pellets were rinsed with 70 % ethanol before dissolving in 100 µl of TE buffer.

Hybridization

Each membrane was placed in a 55 mm x 35 mm x 21 mm plastic box and 1.25 ml of pre-hybridization solution (5x SSC; Denhardt's solution; 1 % SDS; 10 % dextran sulphate [Mw 500,000]; 0.3 % tetrasodium pyrophosphate; 100 µg/ml denatured, sonicated DNA – pre-warmed to 65°C) was added. Each box was closed and incubated at 65°C for 50 minutes on a rocking platform. The pre-hybridization solution was removed and replaced with hybridization solution (5x SSC; Denhardt's solution; 1 % SDS; 10 % dextran sulphate [Mw 500,000]; 0.3 % tetrasodium pyrophosphate; 100 µg/ml denatured, sonicated DNA – containing 5 µl of the appropriate 33P-labelled probe) and the box was incubated at 65°C for 3 hours on a rocking platform.

Washing

The membranes were drained and transferred to 200 ml of 2x SSC, 0.1 % SDS at 68°C for 30 minutes. A further wash was carried out in 0.2x SSC, 0.1 % SDS at 71°C for 30 minutes. The membranes were rinsed in 2x SSC at room temperature and laid out on blotting paper to remove excess liquid. Once dry, the membrane was covered in Saran Wrap and exposed to a Kodak Phosphor Screen for 1 hour. The phosphor screen was subsequently imaged using a Molecular Dynamics Storm 860 Phosphorimager.

Results are shown in figures 4 and 5.

Example #3a

A single cycle of inter-population perfectly matched duplex depletion wherein E.*coli* MutS protein is used to capture an A/A mismatch-containing duplex

'Affected' versus 'unaffected' (i.e. inter-population) mismatch-containing duplex selection can be achieved by: attaching a mismatch-binding protein to a solid support (or using the mismatch-binding protein in solution followed by subsequent solid-phase capture); taking denatured 'affected' DNA fragments and hybridizing these to denatured and biotinylated 'unaffected' DNA fragments; and capture of mismatch-containing duplex molecules with the mismatch-binding protein. Releasing the mismatch-containing duplex molecules (without strand denaturation), streptavidin capture and then release of the non-biotinylated strands will give only the desired species.

In this example, PCR fragments are prepared and used to demonstrate each of the individual steps for a single cycle of inter-population perfectly matched duplex depletion using E.*coli* MutS protein.

Clone insert design

The clone inserts were constructed using standard cloning methodology well known to those skilled in the art and were inserted between the *AvaI* site and *EcoRI* of pMOS*Blue* (Amersham Pharmacia Biotech).

The clone inserts contain a common 9 base pair internal core sequence in which a single nucleotide change or an insertion can be located. The internal core sequence is derived from codons 272-274 of human p53. These codons (GTG CGT GGT) correspond to a mutational hotspot found in lung and other types of cancer (R273L). For the design of the clone inserts containing a mismatch (only one from a complete series of which is shown below), the nucleotide in position 5 of this core sequence is

modified (GTGCXTGGT).

The internal core sequence is flanked by a random sequence – allowing the independent detection of the clone #1 and the clone #7 insert sequence in a mixed population of clone inserts.

Clone #1

Mutant sequence (#1M)

```
5'CCCGGGGGATCCTCGTTTTATTGGGCCGAGTTTTGGTCCGTAGTGCTTGGTTAGATATGCTTAT
3'GGGCCCCCTAGGAGCAAAATAACCCGGCTCAAAACCAGGCATCACGAACCAATCTATACGAATA

GTTCACAAAATCATCCTTGTACAGAATTC3'
CAAGTGTTTTAGTAGGAACATGTCTTAAG5'
```

Control sequence (#1C)

```
5'CCCGGGGGATCCTCGTTTTATTGGGCCGAGTTTTGGTCCGTAGTGCATGGTTAGATATGCTTAT
3'GGGCCCCCTAGGAGCAAAATAACCCGGCTCAAAACCAGGCATCACGTACCAATCTATACGAATA

GTTCACAAAATCATCCTTGTACAGAATTC3'
CAAGTGTTTTAGTAGGAACATGTCTTAAG5'
```

Clone #7

Control sequence (#7C)

```
5'CCCGGGTGTACACAAAAGTTTACCTGAAGAACGTGGGGGGTCGTGCCTGGTCTTGCGTCACCTG
3'GGGCCCACATGTGTTTTCAAATGGACTTCTTGCACCCCCCAGCACGGACCAGAACGCAGTGGAC

GTCTCAGGAGAGGGTCCCCATGGGAATTC3'
CAGAGTCCTCTCCCAGGGGTACCCTTAAG5'
```

Preparation of upper strand 5'-biotinylated and lower strand 5'-biotinylated #1C double-stranded PCR product, and upper strand 5'-biotinylated and lower strand 5'-biotinylated #7C double-stranded PCR product

Oligonucleotides

BIOUPST2	5' bio-CTACTGATCGGATCCCCG 3'
BIODOWN3	5' bio-AAACGACGGCCAGTGAAT 3'

PCR reaction set-up (#1C)

#1C in pMOS*Blue* at 2.50 µg/ml	100 µl
Water	8400 µl
10x PCR buffer	1000 µl
50 µM BIOUPST2	100 µl
50 µM BIODOWN3	100 µl
10 mM dNTPs	200µl
Taq DNA polymerase (5 U/µl)	100 µl

PCR reaction set-up (#7C)

#7C in pMOS*Blue* at 3.32 µg/ml	100 µl
Water	8400 µl
10x PCR buffer	1000 µl
50 µM BIOUPST2	100 µl
50 µM BIODOWN3	100 µl
10 mM dNTPs	200µl
Taq DNA polymerase (5 U/µl)	100 µl

96x 200 µl reactions were carried out for template #1C and 96x 200 µl reactions were carried out for template #7C on a 96-well Perkin Elmer Cetus GeneAmp PCR System 9600 machine as described below:

95 °C, 5 minutes	1 cycle
95 °C, 1 minute	30 cycles
50 °C, 1 minute	
72 °C, 1 minute	
72 °C, 5 minutes	1 cycle
4°C, hold	

The PCR products were pooled together and precipitated by adding 0.1 volumes of 3 M sodium acetate and 1 volume of isopropanol followed by centrifugation at 16,000 rpm for 30 minutes at 4°C (in a Centrikon T-2070 ultracentrifuge; swinging bucket Kontron rotor TST 41.14). Pellets were washed with 14 ml of ethanol and centrifuged at 20,000 rpm for 30 minutes. Finally, the pellets were air-dried and resuspended in a total volume of 0.6 ml of TE buffer.

The 0.6 ml PCR sample was purified in twelve Microspin S-300 HR columns (50 µl per column) following the manufacturer's protocol. Briefly, the resin in the columns was resuspended by vortexing. Columns were centrifuged at 735 x g (3000 rpm in a microfuge) for 1 minute. The sample was then applied to the centre of the resin, being careful not to disturb the bed. The columns were centrifuged at 735 x g for 2 minutes and the flow-through containing the PCR product was collected. The twelve eluted 50 µl volumes were pooled together (pool 1). Columns were washed with 50 µl of TE buffer and the eluted fractions were loaded onto a fresh S-300 HR column. Product yield and removal efficiency of the PCR primers were analysed on a 1.5 % agarose gel.

Preparation of non-biotinylated #1M single-stranded DNA and non-biotinylated #7C single-stranded DNA

Oligonucleotides

BIOUPST2	5' bio-CTACTGATCGGATCCCCG 3'
DOWN3	5' AAACGACGGCCAGTGAAT 3'

PCR reaction set-up (#1M)

#1M in pMOS*Blue* at 3.56 µg/ml	100 µl
Water	8400 µl
10x PCR buffer	1000 µl
50 µM BIOUPST2	100 µl
50 µM DOWN3	100 µl
10 mM dNTPs	200µl
Taq DNA polymerase (5 U/µl)	100 µl

PCR reaction set-up (#7C)

#7C in pMOS*Blue* at 3.32 µg/ml	100 µl
Water	8400 µl
10x PCR buffer	1000 µl
50 µM BIOUPST2	100 µl
50 µM DOWN3	100 µl
10 mM dNTPs	200µl
Taq DNA polymerase (5 U/µl)	100 µl

48x 200 µl reactions were carried out for template #1M and 48x 200 µl reactions were carried out for template #7C on a 96-well Perkin

Elmer Cetus GeneAmp PCR System 9600 machine as described below:

95 °C, 5 minutes	1 cycle
95 °C, 1 minute	30 cycles
50 °C, 1 minute	
72 °C, 1 minute	
72 °C, 5 minutes	1 cycle
4°C, hold	

Capture of biotinylated PCR product strands and release of non-biotinylated strands for the preparation of non-biotinylated #1M single-stranded DNA and non-biotinylated #7C single-stranded DNA

10 ml of pooled #1M and #7C PCR amplifications were each mixed with an equal volume of 4 mg/ml streptavidin-coated colloidal Fe_3O_4 particles taken up in 20 mM tris-HCl (pH 7.4), 2 mM EDTA (pH 8.0), 2 M NaCl. The tubes were incubated at room temperature for 60 minutes with mixing.

The streptavidin-coated colloidal Fe_3O_4 particles were then washed with 20 ml of 10 mM tris-HCl (pH 7.4), 1 mM EDTA (pH 8.0), 1 M NaCl. Two more identical washes were performed.

The washed streptavidin-coated colloidal Fe_3O_4 particles were finally incubated in 800 µl of 0.1 M NaOH for 10 minutes at room temperature. The supernatants were removed and added to 200 µl of 2 M HEPES (free acid). Samples were quantified by absorbance at 260 nm.

Denaturation and annealing

Denaturation and annealing reactions were prepared by mixing:

A = 400 ng of upper strand 5'-biotinylated and lower strand 5'-biotinylated #1C double-stranded PCR product
B = 400 ng of upper strand 5'-biotinylated and lower strand 5'-biotinylated #7C double-stranded PCR product
C = 200 ng of non-biotinylated #1M single-stranded DNA (the lower of the two #1M strands shown above)
D = 200 ng of non-biotinylated #7C single-stranded DNA (the lower of the two #7C strands shown above)

Reannealing between the upper strand of A and the single-stranded C will therefore give rise to an A/A mismatch-containing duplex, 5'-biotinylated on the upper strand, for clone insert #1.

Reannealing between the upper strand of B and the single-stranded D will therefore give rise to a perfectly matched duplex, 5'-biotinylated on the upper strand, for clone insert #7.

0.1 volumes of 1 M NaOH were then added, followed by incubation at room temperature for 10 minutes. 0.25 volumes of 2 M HEPES (free acid) were finally added followed by incubation at 42°C for 1 hour.

Samples were adjusted to 50 µl and were made 1x in PBS and 1 mg/ml in BSA ready for reaction with MutS protein-coated magnetic beads.

One 50 µl sample (the pre-enrichment control, sample 6) was used directly for capture of biotinylated PCR product strands and release of non-biotinylated strands.

Mismatch capture

20 µl of M_2B_2 MutS protein-coated magnetic particles (Genecheck, lot #20) were added to the annealed DNA above. Samples were incubated for 1 hour at room temperature with shaking.

Samples were then washed twice with 200 µl of ice-cold PBS.

Samples were finally eluted from the magnetic beads for 10 minutes at room temperature in 50 µl of the following:

Sample 1	1 M NaCl
Sample 2	PBS
Sample 3	1 M urea
Sample 4	1% (w/v) SDS
Sample 5	10 mM NaOH

Capture of biotinylated PCR product strands and release of non-biotinylated strands

All separations were carried out using an Amersham magnetic separator (RPN1682, batch #1).

50 µl of the eluates from the MutS protein-coated magnetic beads and the pre-enrichment control (sample 6) were each mixed with an equal volume of 4 mg/ml streptavidin-coated colloidal Fe_3O_4 particles in 20 mM tris-HCl (pH 7.4), 2 mM EDTA (pH 8.0), 2 M NaCl. The tubes were incubated at room temperature for 30 minutes with regular mixing.

The streptavidin-coated colloidal Fe_3O_4 particles were then washed twice with 500 µl of 10 mM tris-HCl (pH 7.4), 1 mM EDTA (pH 8.0), 1 M NaCl at room temperature.

The washed streptavidin-coated colloidal Fe_3O_4 particles were incubated in 10 µl of 0.1 M NaOH for 10 minutes at room temperature. The supernatant was removed and added to 2.5 µl of 2 M HEPES (free acid).

5 µl fractions were finally spotted onto Hybond N+ nylon membranes along with 200 ng, 20 ng, 2 ng, and 200 pg amounts of the following:

C = non-biotinylated #1M single-stranded DNA (the lower of the two #1M strands shown previously)
D = non-biotinylated #7C single-stranded DNA (the lower of the two #7C strands shown previously)

Probe labelling

Oligonucleotides

#1 probe oligo 5' GGCCGAGTTTTGGTCCGTAG 3'

#7 probe oligo 5' GTCTTGCGTCACCTGGTCTCAG 3'

Preparation of probes

#1 probe oligo and #7 probe oligo were radioactively 5' end-labelled using T4 polynucleotide kinase as described below (all volumes are in µl):

	#1 probe	#7 probe
#1 probe oligo (25 pmol)	1.25	-
#7 probe oligo (25 pmol)	-	1.25
10x PNK buffer	2.5	2.5
[γ-^{33}P] ATP >92 TBq/mmol, >2500 Ci/mmol, 370 mBq/ml, 10 mCi/ml, (Amersham Pharmacia Biotech., AH9968, lot #B0006)	12.5	12.5
T4 PNK (10 U/µl) (Amersham International, E70031Y, lot #201226)	2.5	2.5
Water	6.25	6.25

The reactions were incubated at 37°C for 30 minutes and

then heated to 70°C for 5 minutes to denature the enzyme.

MicroSpin G-25 column purification

Two G-25 columns (APB 27-5325-01, lot #9015325011) were resuspended by vortexing, and the bottom closures were snapped off as described in the manufacturer's instructions. A pre-spin of the columns was carried out for 1 minute at 730 maxRCF – 2670 rpm in a Hettich Zeutrifugen EBA 12 benchtop centrifuge – and the eluates were discarded. The 25 µl reactions were applied to each column, and the columns were centrifuged for 2 minutes at 730 maxRCF. The eluates from the second spin were stored and used as probes.

Hybridization

Each membrane was placed in a 55 mm x 35 mm x 21 mm plastic box and 2.5 ml of pre-hybridization solution (5x SSC; Denhardt's solution; 1 % SDS; 10 % dextran sulphate [MW 500,000]; 0.3 % tetrasodium pyrophosphate; 100 µg/ml denatured, sonicated DNA – pre-warmed to 42°C) was added. Each box was closed and incubated at 42°C for 1 hour on a rocking platform. The pre-hybridization solution was removed and replaced with hybridization solution (5x SSC; Denhardt's solution; 1 % SDS; 10 % dextran sulphate [MW 500,000]; 0.3 % tetrasodium pyrophosphate; 100 µg/ml denatured, sonicated DNA – containing 2.5 µl of the appropriate ^{33}P-labelled probe) and the box was incubated at 42°C overnight on a rocking platform.

Washing

The membranes were drained and transferred to 200 ml of 2x SSC, 0.1 % SDS at 42°C for 10 minutes. A further wash was carried out in 0.2x SSC, 0.1 % SDS at 42°C for 10 minutes. The membranes were rinsed in 2x SSC at room temperature and laid out on blotting paper to remove excess liquid. Once dry, the membrane was covered in Saran

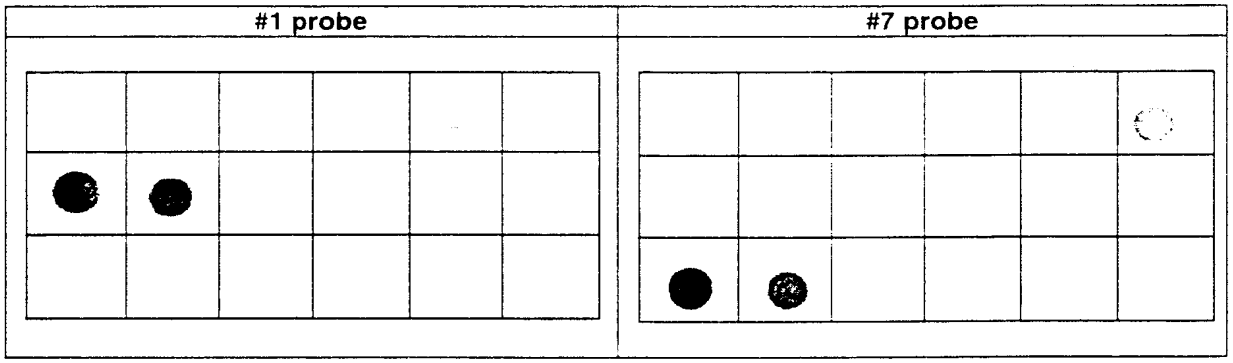
Wrap and exposed to a Kodak Phosphor Screen for 1 hour. The phosphor screen was subsequently imaged using a Molecular Dynamics Storm 860 Phosphorimager.

Dot blot layout

Sample 1 5 µl fraction	Sample 2 5 µl fraction	Sample 3 5 µl fraction	Sample 4 5 µl fraction	Sample 5 5 µl fraction	Sample 6 5 µl fraction
200 ng #1M ssDNA	20 ng #1M ssDNA	2 ng #1M ssDNA	200 pg #1M ssDNA	-	-
200 ng #7C ssDNA	20 ng #7C ssDNA	2 ng #7C ssDNA	200 pg #7C ssDNA	-	-

ssDNA = single-stranded DNA

Probe hybridisation results



Signal intensities from the spots above were reported using ImageQuant 5.0 software (Molecular Dynamics), with the SumAboveBG figures being used after drawing a 6x3 grid over the array of spots.

#1 probe SumAboveBG signals

Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
142257	24356	180359	394296	465242	658323
200 ng #1M ssDNA	20 ng #1M ssDNA	2 ng #1M ssDNA	200 pg #1M ssDNA	-	-
15961011	2842937	252392	25316	2281	2475
200 ng #7C ssDNA	20 ng #7C ssDNA	2 ng #7C ssDNA	200 pg #7C ssDNA	-	-
66536	37955	10621	2736	1955	1934

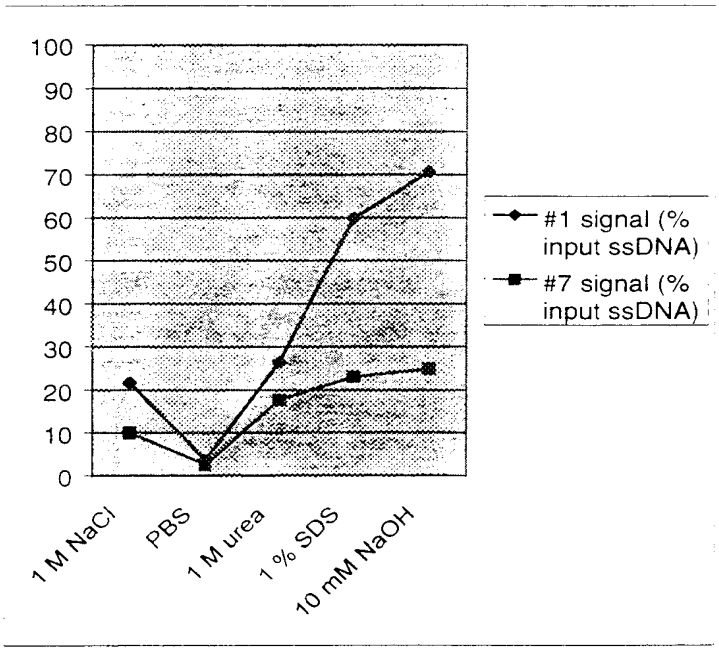
#7 probe SumAboveBG signals

Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
66392	17071	118109	153372	166211	666853
200 ng #1M ssDNA	20 ng #1M ssDNA	2 ng #1M ssDNA	200 pg #1M ssDNA	-	-
80873	23256	6371	2399	1831	1857
200 ng #7C ssDNA	20 ng #7C ssDNA	2 ng #7C ssDNA	200 pg #7C ssDNA	-	-
10619299	2052027	207827	15470	2055	1844

Recovery

Elution conditions	#1 signal (% input ssDNA)	#7 signal (% input ssDNA)	#1/#7 signal ratio
1 M NaCl	21.6 %	10.0 %	2.2
PBS	3.7 %	2.6 %	1.4
1 M urea	26.4 %	17.7 %	1.5
1 % SDS	59.9 %	23.0 %	2.6
10 mM NaOH	70 .7 %	24.9 %	2.8

Recovery figures are plotted below for the various different elution conditions:



Example #3b

A single cycle of inter-population perfectly matched duplex depletion wherein bacteriophage T4 endonuclease VIII protein containing a cleavage-inactivating N62D point mutation is used to capture an A/A mismatch-containing duplex

In this example, PCR fragments are again prepared and used to demonstrate each of the individual steps for a single cycle of inter-population perfectly matched duplex depletion using bacteriophage T4 endonuclease VIII protein containing a cleavage-inactivating N62D point mutation.

Clone design and DNA preparation

Clone insert design was exactly as described in example #3a (see clone #1: mutant sequence (#1M), control sequence (#1C) and clone #7: control sequence (#7C)).

Preparation of upper strand 5'-biotinylated and lower strand 5'-biotinylated #1C double-stranded PCR product, and upper strand 5'-biotinylated and lower strand 5'-biotinylated #7C double-stranded PCR product were as described in example #3a.

Preparation of non-biotinylated #1M single-stranded DNA and non-biotinylated #7C single-stranded DNA were again as described in example #3a.

Denaturation and annealing

Denaturation and annealing reactions were prepared as described in example #3a.

Reannealing will therefore again give rise to an A/A mismatch-containing duplex, 5'-biotinylated on the upper strand, for clone insert #1 and a perfectly matched duplex, 5'-biotinylated on the upper strand, for clone insert #7.

One 50 µl sample (the pre-enrichment control) was adjusted to 100 µl with TE buffer and was used directly for capture of biotinylated PCR product strands and release of non-biotinylated strands.

Mismatch capture

50 µl samples of the annealing reaction were mixed with an equal volume of 200 mM sodium phosphate (pH 6.5), 100 mM KCl. 10 µg of GST-tagged T4 endonuclease VII N62D mutant (obtained from Prof. Börries Kemper, Univ. Cologne) were added to the annealed DNA and the samples were incubated for 15 minutes at 16°C.

Samples were then mixed with 200 µl of a 50 % slurry of Glutathione Sepharose 4B (Amersham Pharmacia Biotech, lot #279991) in 100 mM sodium phosphate (pH 6.5), 50 mM KCl. The tubes were incubated at 16°C for 30 minutes with regular mixing. The mixture was then transferred to a spin column to separate solid from liquid phases.

Samples were finally eluted from the Glutathione Sepharose

4B matrix for 10 minutes at room temperature in 100 µl of 10 mM reduced glutathione in 50 mM tris-HCl (pH 8.0).

Capture of biotinylated PCR product strands and release of non-biotinylated strands

All separations were carried out using an Amersham magnetic separator (RPN1682, batch #1).

100 µl of the eluate from the N62D T4 endonuclease VII mismatch-capture reaction, and the pre-enrichment control were each mixed with an equal volume of 4 mg/ml streptavidin-coated colloidal Fe_3O_4 particles in 20 mM tris-HCl (pH 7.4), 2 mM EDTA (pH 8.0), 2 M NaCl. The tubes were incubated at room temperature for 30 minutes with regular mixing.

The streptavidin-coated colloidal Fe_3O_4 particles were then washed twice with 500 µl of 10 mM tris-HCl (pH 7.4), 1 mM EDTA (pH 8.0), 1 M NaCl at room temperature.

The washed streptavidin-coated colloidal Fe_3O_4 particles were incubated in 10 µl of 0.1 M NaOH for 10 minutes at room temperature. The supernatant was removed and added to 2.5 µl of 2 M HEPES (free acid).

5 µl fractions were finally spotted onto Hybond N+ nylon membranes along with 200 ng, 20 ng, 2 ng, and 200 pg amounts of the following:

C = non-biotinylated #1M single-stranded DNA (the lower of the two #1M strands shown previously)
D = non-biotinylated #7C single-stranded DNA (the lower of the two #7C strands shown previously)

Probe labelling

Probe labelling was exactly as described in example #3a.

Hybridization

Each membrane was placed in a 55 mm x 35 mm x 21 mm plastic box and 2.5 ml of pre-hybridization solution (5x SSC; Denhardt's solution; 1 % SDS; 10 % dextran sulphate [MW 500,000]; 0.3 % tetrasodium pyrophosphate; 100 µg/ml denatured, sonicated DNA – pre-warmed to 42°C) was added. Each box was closed and incubated at 42°C for 10 minutes on a rocking platform. The pre-hybridization solution was removed and replaced with hybridization solution (5x SSC; Denhardt's solution; 1 % SDS; 10 % dextran sulphate [MW 500,000]; 0.3 % tetrasodium pyrophosphate; 100 µg/ml denatured, sonicated DNA – containing 2.5 µl of the appropriate ^{33}P-labelled probe) and the box was incubated at 42°C overnight on a rocking platform.

Washing

The membranes were drained and transferred to 200 ml of 2x SSC, 0.1 % SDS at 42°C for 10 minutes. A further wash was carried out in 0.2x SSC, 0.1 % SDS at 42°C for 10 minutes. The membranes were rinsed in 2x SSC at room temperature and laid out on blotting paper to remove excess liquid. Once dry, the membrane was covered in Saran Wrap and exposed to a Kodak Phosphor Screen for 1 hour. The phosphor screen was subsequently imaged using a Molecular Dynamics Storm 860 Phosphorimager.

Signal intensities from the spots above were again reported using ImageQuant 5.0 software (Molecular Dynamics), with the SumAboveBG figures being used after drawing a grid over the array of spots.

#1 probe SumAboveBG signals

200 ng #1M ssDNA 22418970	20 ng #1M ssDNA 3317389	2 ng #1M ssDNA 270664	200 pg #1M ssDNA 25763	GSH eluted sample 144398
200 ng #7C ssDNA 229951	20 ng #7C ssDNA 105343	2 ng #7C ssDNA 16354	200 pg #7C ssDNA 3632	Pre-enrichment control 997077

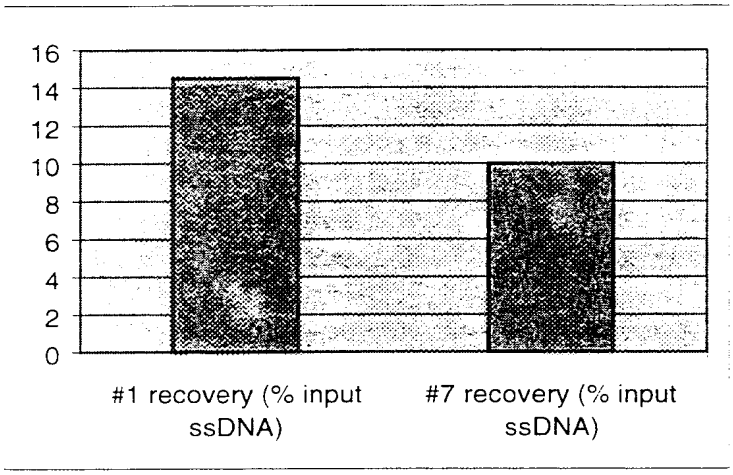
#7 probe SumAboveBG signals

200 ng #1M ssDNA 117267	20 ng #1M ssDNA 42226	2 ng #1M ssDNA 9421	200 pg #1M ssDNA 2812	GSH eluted sample 113608
200 ng #7C ssDNA 3665967	20 ng #7C ssDNA 22248096	2 ng #7C ssDNA 203869	200 pg #7C ssDNA 19406	Pre-enrichment control 1135786

Recovery

Elution conditions	#1 signal (% input ssDNA)	#7 signal (% input ssDNA)	#1/#7 signal ratio
10 mM reduced glutathione in 50 mM tris-HCl (pH 8.0)	14.5 %	10.0 %	1.45

The enrichment results are presented graphically below:



Dedicated to the memory of Chris Griffin and Richard Beer

CLAIMS

1. A method of providing a mixture of DNA fragments enriched in fragments that are characteristic of a phenotype of interest, by providing affected DNA in fragmented form and providing unaffected DNA in fragmented form, which method comprises:

a) mixing the fragments of the affected DNA and the fragments of the unaffected DNA under hybridising conditions;

b) recovering a mixture of hybrids that contain mismatches;

c) recovering fragments of the affected DNA from the mixture of hybrids that contain mismatches;

and optionally repeating steps a), b) and c) one or more times.

2. The method of claim 1 wherein the affected DNA is pooled DNA of individuals who show the phenotype of interest, and the unaffected DNA is pooled DNA of individuals who do not show the phenotype of interest.

3. The method of claim 1, wherein the affected DNA is DNA of one individual who shows the phenotype of interest, and the unaffected DNA is pooled DNA of individuals who do not show the phenotype of interest.

4. The method of claim 1, wherein the affected DNA is DNA of one individual who shows the phenotype of interest, and the unaffected DNA is pooled DNA of a complete set of ancestors who do not show the phenotype of interest.

5. The method of claim 1, wherein the affected DNA is DNA from cells of an individual that show the phenotype of interest, and the unaffected DNA is DNA from cells of the individual that do not show the phenotype of interest.

6. The method of any one of claims 1 to 5, wherein step b) is performed by use of a mismatch-binding protein.

7. The method of any one of claims 1 to 6, wherein either the fragments of the affected DNA or the fragments of the unaffected DNA are tagged by one member of a specific binding pair, and step c) is performed by using the other member of the specific binding pair.

8. The method of claim 7, wherein the fragments of the unaffected DNA are tagged with biotin, and step c) is performed by use of immobilised streptavidin.

9. The method of any one of claims 1 to 8, wherein the mixture of DNA fragments enriched in fragments that are characteristic of the phenotype of interest, is subjected to self-hybridisation followed by recovery of perfectly matched duplexes.

10. The method of any one of claims 1 to 9, wherein the mixture of DNA fragments enriched in fragments that are characteristic of the phenotype of interest, is mixed with an excess of the fragments of the affected DNA under hybridisation conditions, followed by recovery of perfectly matched duplexes.

11. The method of any one of claims 1 to 10, wherein each of the affected DNA and the unaffected DNA is provided in fragmented form by digestion with from 4 to 7 six-cutter restriction endonuclease enzymes together with from 0 to 5 four-cutter restriction endonuclease enzymes.

12. A mixture of DNA fragments enriched in fragments that are characteristic of the phenotype of interest, provided by the method of any one of claims 1 to 11.

13. A method of making a set of arrays of fragments of DNA of interest, which method comprises:

a) selecting, from a set of n restriction endonuclease enzymes, a subset of r restriction endonuclease enzymes;

b) digesting genomic DNA with the subset of r enzymes;

c) ligating to the resulting fragments restriction-enzyme-cutting-site-specific adapters with unique polymerase chain reaction amplifiable sequences;

d) splitting the resulting fragments into r^2 aliquots;

e) amplifying each aliquot with two restriction-enzyme-specific primers of which one is tagged;

f) forming an array of the r^2 aliquots of non-tagged amplimer strands; and

g) repeating steps a) to f) using one or more different subsets of r restriction endonuclease enzymes.

14. A method of making a set of arrays of fragments of DNA of interest, which method comprises:

a) selecting, from a set of n restriction endonuclease enzymes, a subset of r restriction endonuclease enzymes;

b) digesting genomic DNA with the subset of r enzymes;

c) ligating to the resulting fragments restriction-enzyme-cutting-site-specific adapters with unique polymerase chain reaction amplifiable sequences;

d) splitting the resulting fragments into r^2 aliquots;

e) amplifying each aliquot with two restriction-enzyme-specific primers;

f) forming an array of the r^2 aliquots of the amplimer strands; and

g) repeating steps a) to f) using one or more different subsets of r restriction endonuclease enzymes.

15. The method of claim 13 or claim 14, wherein steps a to f) are repeated using each different subset of r restriction endonuclease enzymes to give (n!)/[(n-r)!r!] different arrays.

16. The method of any one of claims 13 to 15, wherein the n restriction endonuclease enzymes are selected from 4-cutters and 5-cutters and 6-cutters.

17. The method of any one of claims 13 to 16, wherein n is 3 to 10 and r is 2 to 4.

18. The method of claim 17, where n = 6 and r = 3.

19. A set of arrays of fragments of DNA of interest, which set results from performance of the method of any one of claims 13 to 18.

20. The set of arrays of claim 19, which set results from performance of the method of claim 13 and claim 14 and claim 15.

21. The set of arrays of claim 19 or claim 20, derived from a set of n = 6 six-cutter restriction endonuclease enzymes which are *BamHI; Bsr GI; Hind III; NcoI; SpeI;* and *AflII.*

22. The set of arrays of claim 19 or claim 20, derived from the set of n = 6 six-cutter restriction endonuclease enzymes which are *EcoRI; BspHI; BglII; XbaI; Acc65I;* and *ApaLI.*

23. A nucleic acid characterisation method which comprises presenting to the set of arrays of any one of claims 19 to 22 a nucleic acid fragment of interest under hybridisation conditions, and observing a pattern of hybridisation.

24. The method of claim 23, wherein a plurality of nucleic acid fragments of interest are separately presented to the set of arrays, and the resulting patterns of hybridisation are compared.

25. The method of claim 24, wherein the plurality of nucleic acid fragments of interest are drawn from the mixture of DNA fragments, enriched in fragments that are characteristic of a phenotype of interest, of claim 13.

26. A method of identifying fragments of DNA that are characteristic of a phenotype of interest, which method comprises recovering, cloning and amplifying individual DNA fragments from the mixture of DNA fragments of claim 12, presenting the individual DNA fragments to the set of arrays of any one of claims 19 to 22 under

hybridisation conditions, observing a pattern of hybridisation generated by each individual DNA fragment, and subjecting to further investigation any two or more individual DNA fragments whose hybridisation patterns are similar or identical, or near to each other in a genome of interest.

27. A double-stranded DNA molecule having the sequence a-A-b-B...X-y-Y-z where A, B...X and Y are unique restriction sites for n different restriction endonuclease enzymes, and a, b...y, z denotes distances in base pairs, characterised in that each fragment, obtainable by cutting the DNA molecule by means of any one or more up to n of the restriction enzymes, has a different length from every other fragment.

28. The double-stranded DNA molecule of claim 27, wherein the following criteria are satisfied:

a) inter-fragment length differences are greater for larger fragments;

b) all possible fragments are unambiguousiy resolvable by electrophoresis from one another;

c) size gaps between bands comprising different numbers of inter-restriction-site units are larger than size gaps between bands comprising the same number of inter-restriction-site units;

d) the size gaps and size spread from the largest to the smallest fragment are electrophorectically compatible.

Figure 1

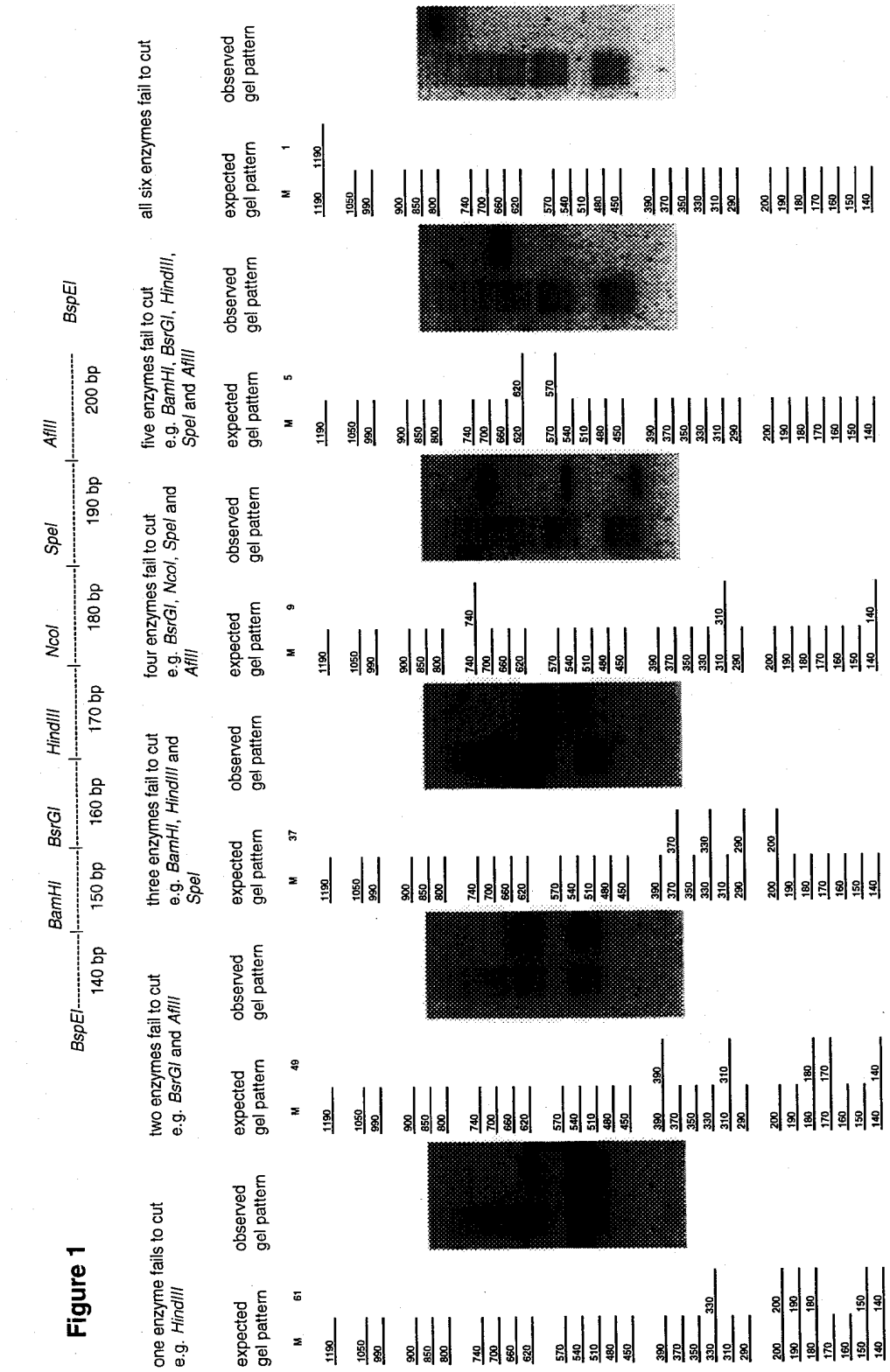
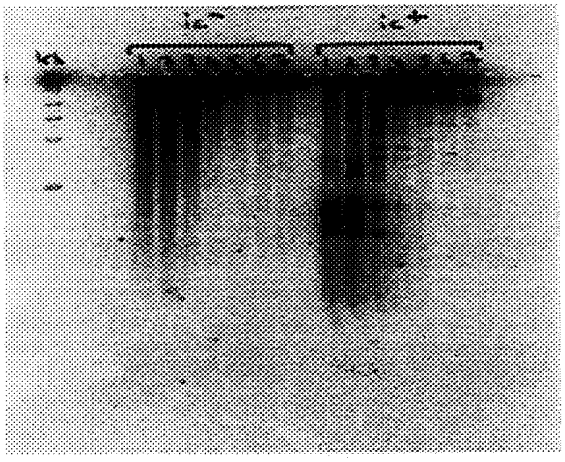

BspEI
BamHI
BsrGI
HindIII
NcoI
SpeI
AflII
BspEI
140 bp
150 bp
160 bp
170 bp
180 bp
190 bp
200 bp
one enzyme fails to cut
e.g. HindIII
two enzymes fail to cut
e.g. BsrGI and AflII
three enzymes fail to cut
e.g. BamHI, HindIII and SpeI
four enzymes fail to cut
e.g. BsrGI, NcoI, SpeI and AflII
five enzymes fail to cut
e.g. BamHI, BsrGI, HindIII, SpeI and AflII
all six enzymes fail to cut
expected gel pattern
observed gel pattern
M
61
49
37
9
5
1
1190
1050
990
900
850
800
740
700
660
620
570
540
510
480
450
390
370
350
330
310
290
200
190
180
170
160
150
140

Figure 2

Vistra Green fluorescence

1 µg of Stratagene Kb DNA Ladder (band sizes: 250 bp, 500 bp, 750 bp, 1 kb, 1.5 kb, 2 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb, 10 kb and 12 kb), 3x gaps, 2 µg of canine genomic DNA digested with 0.25, 0.025, 0.0025, 0.00025, 0.000025, 0.0000025 and 0 U/µl *BamH1 / BsrGI / HindIII / NcoI / SpeI / AflII*, gap, 0.4 µg of canine genomic DNA and 1 µl of *BspEI*-released internal control DNA digested with 0.25, 0.025, 0.0025, 0.00025, 0.000025, 0.0000025 and 0 U/µl *BamH1 / BsrGI / HindIII / NcoI / SpeI / AflII*.

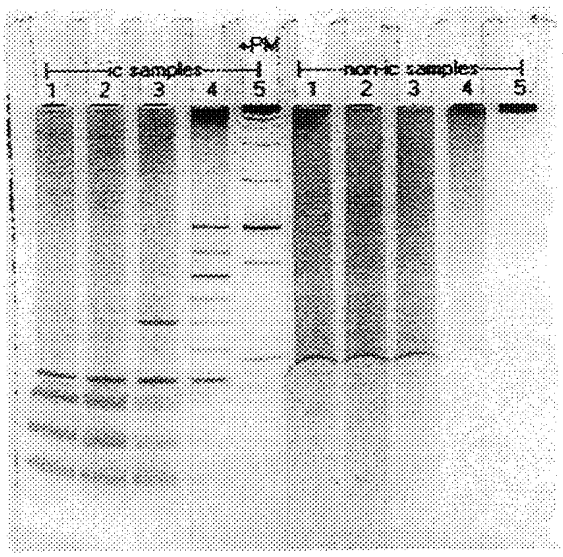


From the results above, the fragment pattern from the internal control DNA is diagnostic of both the degree of digestion and the nature of any partial restriction at less than limit digestion.

Very little digestion of canine genomic DNA and *BspEI*-released internal control DNA is observed for 0.00025 U/µl *BamH1 / BsrGI / HindIII / NcoI / SpeI / AflII* and fewer units. With 0.0025 U/µl, *BamHI* and *AflII* are both failing to cut more than the other enzymes (the 140 bp and 200 bp bands in 3ic are diagnostic of this). With 0.025 U/µl, most enzymes have cut – though not quite to completion. 0.25 U/µl thus appears to be necessary in order to achieve limit digestion of 20 µg of canine genomic DNA in a volume of 200 µl.

Figure 3

Figure 3 shows the products of the restriction digests after electrophoresis on an 8 % polyacrylamide gel. The numbers 1–5 indicate the dilution of enzyme mix used. 1=0.5 U/µl, 2=0.1 U/µl, 3=0.02 U/µl, 4=0.004 U/µl, and 5=0 U/µl. 'ic' indicates the samples that were spiked with 4 bp cutter internal control DNA. PM indicates that 180 ng of PCR molecular weight markers were added as a size standard.



All five of the partial digestion products are visible in lane ic4, between the top 130 bp band and the bottom 40 bp band, indicating that all three enzymes have failed to cut to completion at 0.004 U/µl. In lane ic3, the four complete digestion products are all visible but the 65 bp band is also present. *HaeIII* and *MboI* have digested to completion but the *MseI* digestion is still partial at 0.2 U/µl. At 0.1 U/µl and at 0.5 U/µl all three enzymes have produced limit digests.

Using the internal control spike, it is a great deal easier to determine the point at which limit digestion has occurred. The undigested internal control DNA is 130 bp. Partial digestion products generate bands of 105 bp, 90 bp, 75 bp, 65 bp, and 55 bp. The products of complete digestion generate bands of 40 bp, 35 bp, 30 bp, and 25 bp. The fragment pattern from the internal control DNA is therefore diagnostic of both the degree of digestion and the nature of any partial restriction at less than limit digestion.

Figure 4

Expected TRSPA-2 pattern for the pNW33 *BamHI*, *HindIII*, and *AflII* matrix probed with the 140 bp *BspEI* to *BamHI* fragment from pNW33

The TRSPA-2 pattern expected for matrix #7 (*BamHI*, *HindIII*, and *AflII*) hybridized with a ^{33}P-labelled probe derived from the 140 bp *BspEI – BamHI* fragment of pNW33 is shown below:

#7	*BamHI*	*HindIII*	*AflII*
BamHI	-	HYB	-
HindIII	HYB	-	-
AflII	-	-	-

z

Observed TRSPA-2 pattern for the pNW33 *BamHI*, *HindIII*, and *AflII* matrix probed with the 140 bp *BspEI* to *BamHI* fragment from pNW33

The TRSPA-2 pattern observed for matrix #7 (*BamHI*, *HindIII*, and *AflII*) hybridized with a ^{33}P-labelled probe derived from the 140 bp *BspEI – BamHI* fragment of pNW33 is shown below:

#7	*BamHI*	*HindIII*	*AflII*
BamHI			
HindIII			
AflII			

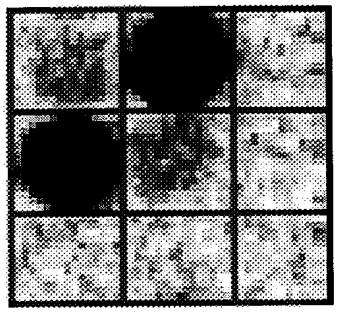


Figure 5

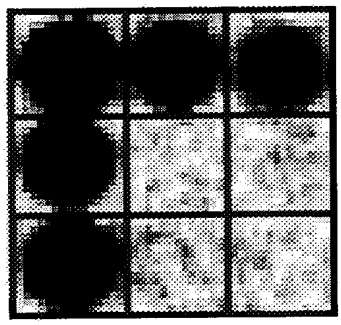
Expected TRSPA-2 pattern for the pNW33 *HindIII*, *NcoI*, and *SpeI* matrix probed with the 140 bp *BspEI* to *BamHI* fragment from pNW33

The TRSPA-2 pattern expected for matrix #17 (*HindIII*, *NcoI*, and *SpeI*) hybridized with a ^{33}P-labelled probe derived from the 140 bp *BspEI* – *BamHI* fragment of pNW33 is shown below:

#17	*HindIII*	*NcoI*	*SpeI*
HindIII	HYB	HYB	HYB
NcoI	HYB	-	-
SpeI	HYB	-	-

Observed TRSPA-2 pattern for the pNW33 *HindIII*, *NcoI*, and *SpeI* matrix probed with the 140 bp *BspEI* to *BamHI* fragment from pNW33

The TRSPA-2 pattern expected for matrix #17 (*HindIII*, *NcoI*, and *SpeI*) hybridized with a ^{33}P-labelled probe derived from the 140 bp *BspEI* – *BamHI* fragment of pNW33 is shown below:

#17	*HindIII*	*NcoI*	*SpeI*
HindIII			
NcoI			
SpeI			

Espacenet

Bibliographic data: WO02072772 (A2) — 2002-09-19

METHODS AND COMPOSITIONS FOR AMPLIFICATION OF RNA SEQUENCES

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

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

Classification: - international: *C07H21/02; C12N15/09; C12P19/34; C12Q1/68;* (IPC1-7): C12N
- cooperative: **C12Q1/6853** more

Application number: WO2002US07306 20020311 Global Dossier

Priority number (s): US20010274550P 20010309

Also published as: WO02072772 (A3) AT361996 (T) AU2002252279 (B2) BR0205268 (A) CA2416963 (A1) more

Abstract of WO02072772 (A2)

The invention provides methods for isothermal amplification of RNA. The methods are particularly suitable for amplifying a plurality of RNA species in a sample. The methods employ a composite primer, a second primer and strand displacement to generate multiple copies of DNA products comprising sequences complementary to an RNA sequence of interest. In another aspect, the methods employ a single primer (which is a composite primer) and strand displacement to generate multiple copies of DNA products comprising sequences complementary to an RNA sequence of interest. In some embodiments, a transcription step is included to generate multiple copies of sense RNA of an RNA sequence of interest. The methods are useful for preparation of nucleic acid libraries and substrates for analysis of gene expression of cells in biological samples. The invention also provides compositions and kits for practicing the amplification methods, as well as methods which use the amplification products.

IDT Exhibit 1020.9245

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 September 2002 (19.09.2002)

PCT

(10) International Publication Number
WO 02/072772 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/07306

(22) International Filing Date: 11 March 2002 (11.03.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/274,550 9 March 2001 (09.03.2001) US

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

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

(74) Agents: **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, 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, 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, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/072772 A2

(54) Title: METHODS AND COMPOSITIONS FOR AMPLIFICATION OF RNA SEQUENCES

(57) Abstract: The invention provides methods for isothermal amplification of RNA. The methods are particularly suitable for amplifying a plurality of RNA species in a sample. The methods employ a composite primer, a second primer and strand displacement to generate multiple copies of DNA products comprising sequences complementary to an RNA sequence of interest. In another aspect, the methods employ a single primer (which is a composite primer) and strand displacement to generate multiple copies of DNA products comprising sequences complementary to an RNA sequence of interest. In some embodiments, a transcription step is included to generate multiple copies of sense RNA of an RNA sequence of interest. The methods are useful for preparation of nucleic acid libraries and substrates for analysis of gene expression of cells in biological samples. The invention also provides compositions and kits for practicing the amplification methods, as well as methods which use the amplification products.

IDT Exhibit 1020.9246

METHODS AND COMPOSITIONS FOR AMPLIFICATION OF RNA SEQUENCES

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the priority benefit of the provisional patent application U.S. Serial No. 60/274,550, filed March 9, 2001, which is incorporated by reference in its entirety.

TECHNICAL FIELD

The invention relates to the field of polynucleotide amplification. More particularly, the invention provides methods, compositions and kits for amplifying (i.e., making multiple copies) RNA sequences of interest which employ an RNA/DNA composite primer, and, optionally transcription.

BACKGROUND ART

The ability to amplify ribonucleic acid (RNA) is an important aspect of efforts to elucidate biological processes. To date, RNA (generally, mRNA) amplification is most commonly performed using the reverse transcriptase-polymerase chain reaction (RT-PCR) method and variations thereof. These methods are based on replication of RNA by reverse transcriptase to form single stranded DNA complementary to the RNA (cDNA), which is followed by polymerase chain reaction (PCR) amplification to produce multiple copies of double stranded DNA. Although these methods are most commonly used, they have some significant drawbacks: a) the reactions require thermocycling; b) the products are double stranded, thus rendering them less accessible to binding to probes; c) the reactions are prone to contamination with products of prior amplification, thus requiring strict containment of reaction mixtures; and d) the exponential nature of amplification of these methods renders them prone to generate pools of products which do not truly reflect the representation of the various RNA sequences in the input total RNA sample, due to unequal efficiency of amplification of different sequences, and the nature of exponential

IDT Exhibit 1020.9247

amplification which is based on replication of amplification products rather than on continued replication of the input target RNAs.

Total cellular mRNA represents gene expression activity at a defined time. Gene expression is affected by cell cycle progression, developmental regulation, response to internal and external stimuli and the like. The profile of expressed genes for any cell type in an organism reflects normal or disease states, response to various stimuli, developmental stages, cell differentiation, and the like.

Various methods for the analysis of gene expression have been developed in recent years. See, for example, U.S. Pat. Nos. 5,744,308; 6,143,495; 5,824,517; 5,829,547; 5,888,779; 5,545,522; 5,716,785; 5,409,818; EP 0971039A2; EP0878553A2. These include quantification of specific mRNAs, and the simultaneous quantification of a large number of mRNAs, as well as the detection and quantification of patterns of expression of known and unknown genes. The analysis of gene expression profiles is currently one of the most powerful tools in the study of cellular differentiation and cellular development, and in the investigation of normal and disease states of various organisms, in particular in human. This analysis is crucial for gene discovery, molecular medicine and drug discovery processes.

Amplification of the total cellular mRNAs prepared from any cell or tissue is generally critical for gene expression profiling. Although analysis of non-amplified mRNA is feasible, a significant amount of starting mRNA would be required. However, the total amount of sample mRNA that is available is frequently limited by the amount of biological sample from which it is derived. Biological samples are often limited in amount and precious. Moreover, the amount of the various mRNA species is not equal; some species are more abundant than others, and these are more likely and easier, to analyze. The ability to amplify mRNA sequences enables the analysis of less abundant, rare mRNA species. The ability to analyze small samples, by means of nucleic acid amplification, is also advantageous for design parameters of large scale screening of effector molecule libraries, for which reduction in sample volume is a major concern both for the ability to perform very large scale screening or ultra high throughput screening, and in view of the limiting amounts of library components.

Therefore, there is a need for improved RNA amplification methods that overcome drawbacks in existing methods. The invention provided herein fulfills this need and provides additional benefits.

IDT Exhibit 1020.9248

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 RNA amplification, as well as applications of the amplification methods.

Accordingly, in one aspect, the invention provides methods of generating multiple copies of a polynucleotide sequence complementary to an RNA sequence of interest, said method comprising the steps of: (a) extending a first primer hybridized to a target RNA with an RNA-dependent DNA polymerase, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, whereby a complex comprising a first primer extension product and the target RNA is produced; (b) cleaving RNA in the complex of step (b) with an enzyme that cleaves RNA from an RNA/DNA hybrid; (c) extending a second primer hybridized to the first primer extension product with a DNA-dependent DNA polymerase, whereby a second primer extension product is produced to form a complex of first and second primer extension products; (d) cleaving RNA from the composite primer in the complex of first and second primer extension products with an enzyme that cleaves RNA from an RNA/DNA hybrid such that a composite primer hybridizes to the second primer extension product, wherein the composite primer comprises an RNA portion and a 3' DNA portion; (e) extending the composite primer hybridized to the second primer extension product with a DNA-dependent DNA polymerase; whereby said first primer extension product is displaced, and whereby multiple copies of a polynucleotide sequence complementary to the RNA sequence of interest are generated.

In another aspect, the invention provides methods of generating multiple copies of a polynucleotide sequence complementary to an RNA sequence of interest, said method comprising the steps of: (a) extending a second primer hybridized to a first primer extension product with a DNA-dependent DNA polymerase, wherein the first primer extension product comprises an RNA portion at the 5' end, said first primer extension product comprising a sequence complementary to an RNA sequence, whereby a second primer extension product is produced to form a complex of first and second primer

IDT Exhibit 1020.9249

extension products; (b) cleaving RNA in the complex of first and second primer extension products with an enzyme that cleaves RNA from an RNA/DNA hybrid such that a composite primer hybridizes to the second primer extension product, wherein the composite primer comprises an RNA portion and a 3' DNA portion; (c) extending the composite primer hybridized to the second primer extension product with a DNA-dependent DNA polymerase; whereby said first primer extension product is displaced, and whereby multiple copies of a polynucleotide sequence complementary to the RNA sequence of interest are generated. In some embodiments of the invention, the first primer extension product is produced by extension of a first primer hybridized to a target RNA with a RNA-dependent DNA polymerase, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion.

In another aspect, the invention provides methods of generating multiple copies of a polynucleotide sequence complementary to an RNA sequence of interest, said method comprising the steps of: (a) cleaving RNA from a complex of first and second primer extension products with an enzyme that cleaves RNA from an RNA/DNA hybrid such that a composite primer hybridizes to the second primer extension product, wherein the composite primer comprises an RNA portion and a 3' DNA portion, wherein the first primer extension product is produced by extension of a first primer hybridized to a target RNA with a RNA-dependent DNA polymerase, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion; (b) hybridizing a composite primer to the second primer extension product and extending the composite primer with a DNA-dependent DNA polymerase; whereby said first primer extension product is displaced, and whereby multiple copies of a polynucleotide sequence complementary to the RNA sequence of interest are generated.

In another aspect, the invention provides methods of generating multiple copies of a polynucleotide sequence complementary to an RNA sequence of interest, said method comprising the step of: extending a composite primer in a complex comprising: (i) a complex of a first and second primer extension products, wherein the first primer extension product is produced by extension of a first primer hybridized to a target RNA with a RNA-dependent DNA polymerase, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, wherein the second primer extension product is generated by extension of a second primer hybridized to the first primer extension product, and wherein RNA from the complex of first and second primer extension products is

IDT Exhibit 1020.9250

cleaved with an enzyme that cleaves RNA from an RNA/DNA hybrid; and (ii) a composite primer, said composite primer comprising an RNA portion and a 3' DNA portion, wherein the composite primer is hybridized to the second primer extension product, and wherein the composite primer may be the same or different from the first primer; whereby said first primer extension product is displaced, and whereby multiple copies of a polynucleotide sequence complementary to the RNA sequence of interest are generated.

In another aspect, the invention provides methods of generating multiple copies of a polynucleotide sequence complementary to an RNA sequence of interest, said method comprising the steps of: (a) extending a composite primer hybridized to a second primer extension product, wherein said primer extension product (i) comprises a complement or a sequence of a first primer extension product generated by extension of a first primer hybridized to template RNA, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion; and (ii) is hybridized to a second primer extension product generated by extension of a second primer hybridized to the first primer extension product, and cleavage of RNA from the first primer with an enzyme that cleaves RNA from an RNA/DNA hybrid; whereby said first primer extension product is displaced, and whereby multiple copies of a polynucleotide sequence complementary to the RNA sequence of interest are generated.

In another aspect, the invention provides methods of generating multiple copies of an RNA sequence of interest, said method comprising the steps of:

hybridizing the displaced first primer extension product from any of the methods of generating multiple copies of a polynucleotide sequence complementary to an RNA sequence of interest described herein, with a polynucleotide comprising a propromoter and a region which is hybridizable to the displaced first primer extension product under conditions which allow transcription to occur by RNA polymerase, such that RNA transcripts are produced comprising sequences complementary to the displaced first primer extension product, whereby multiple copies of the RNA sequence of interest are generated.

In another aspect, the invention provides methods of generating multiple copies of an RNA sequence of interest, said method comprising the steps of: hybridizing a first primer extension product from any of the methods of generating multiple copies of a polynucleotide sequence complementary to an RNA sequence of interest described herein, with a polynucleotide comprising a propromoter and a region which is hybridizable to the displaced first primer extension product under conditions which allow transcription to

IDT Exhibit 1020.9251

occur by RNA polymerase, such that RNA transcripts are produced comprising sequences complementary to the displaced first primer extension product, wherein the primer extension product is a displaced primer extension product generated by: (a) extending a first primer hybridized to a target RNA with an RNA-dependent DNA polymerase, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, whereby a complex comprising a first primer extension product and the target RNA is produced; (b) cleaving RNA in the complex of step (b) with an enzyme that cleaves RNA from an RNA/DNA hybrid; (c) extending a second primer hybridized to the first primer extension product with a DNA-dependent DNA polymerase, whereby a second primer extension product is produced to form a complex of first and second primer extension products; (d) cleaving RNA from the composite primer in the complex of first and second primer extension products with an enzyme that cleaves RNA from an RNA/DNA hybrid such that a composite primer hybridizes to the second primer extension product, wherein the composite primer comprises an RNA portion and a 3' DNA portion; (e) extending said composite primer hybridized to the second primer extension product with a DNA-dependent DNA polymerase, whereby said first primer extension product is displaced; whereby multiple copies of the RNA sequence of interest are generated.

In another aspect, the invention provides methods of generating multiple copies of (amplifying) a polynucleotide sequence complementary to an RNA sequence of interest, said method comprising the steps of: (a) hybridizing a first primer to a target RNA, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion; (b) extending the first primer with an RNA-dependent DNA polymerase, whereby a complex comprising a first primer extension product and the target RNA is produced; (c) cleaving RNA in the complex of step (b) with an agent (such as an enzyme) that cleaves RNA from an RNA/DNA hybrid; (d) hybridizing a second primer to the first primer extension product; (e) extending the second primer with a DNA-dependent DNA polymerase, whereby a second primer extension product is produced to form a complex of first and second primer extension products; (f) cleaving RNA from the composite primer in the complex of first and second primer extension products with an agent (such as an enzyme) that cleaves RNA from an RNA/DNA hybrid such that a composite primer (which may or may not be the same as the first primer) hybridizes to the second primer extension product, wherein said composite primer comprises an RNA portion and a 3' DNA portion; (g) extending said composite primer of step (f) with a DNA-dependent DNA polymerase,

IDT Exhibit 1020.9252

whereby said first primer extension product is displaced, and whereby multiple copies of a polynucleotide sequence complementary to the RNA sequence of interest are generated.

In some aspects, the invention provides methods of generating multiple copies of a polynucleotide sequence complementary to an RNA sequence of interest, said method comprising the steps of: combining and reacting: the complex of first and second primer extension products of step (e) described above; a composite primer (which may or may not be the same as the first primer) that is hybridizable to the second primer extension product, wherein said composite primer comprises an RNA portion and a 3' DNA portion; a DNA-dependent DNA polymerase; and an agent (such as an enzyme) that cleaves RNA from an RNA/DNA hybrid; wherein the mixture is incubated under conditions (which include necessary substrates and buffer conditions) that permit primer hybridization, RNA cleavage, and displacement of the first primer extension product from the complex of step (e) described above when its RNA is cleaved and a composite primer binds to the second primer extension product in the complex.

As is clear to one skilled in the art, reference to production of copies of an RNA sequence of interest or copies of a polynucleotide sequence complementary to an RNA sequence of interest refers to products that may contain, comprise or consist of such sequences. As is evident to one skilled in the art, aspects that refer to combining and incubating the resultant mixture also encompasses method embodiments which comprise incubating the various mixtures (in various combinations and/or subcombinations) so that the desired products are formed.

In another aspect, the invention provides methods of generating multiple copies of a polynucleotide sequence complementary to an RNA sequence of interest comprising incubating a reaction mixture, said reaction mixture comprising: (a) a target RNA; (b) a first primer that is hybridizable to the target RNA, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion; (c) a second primer that is hybridizable to an extension product of the first primer; (d) an RNA-dependent DNA polymerase; (e) a DNA-dependent DNA polymerase; (f) an enzyme that cleaves RNA from an RNA/DNA hybrid; wherein the incubation is under conditions that permit primer hybridization, extension from the first primer and the second primer to form a complex comprising first and second primer extension products, RNA cleavage, and displacement of the first primer extension product from the complex when its RNA is cleaved and a composite primer binds to the second primer extension product in the complex and is

IDT Exhibit 1020.9253

extended, whereby multiple copies of a polynucleotide sequence complementary to the RNA sequence of interest are generated.

In another aspect, the invention provides methods of generating multiple copies of (amplifying) an RNA sequence of interest, said method comprising the steps of: (a) hybridizing a first primer to a target RNA, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion; (b) extending the first primer with an RNA-dependent DNA polymerase, whereby a complex comprising a first primer extension product and the target RNA is produced; (c) cleaving RNA in the complex of step (b) with an agent (such as an enzyme) that cleaves RNA from an RNA/DNA hybrid; (d) hybridizing a second primer to the first primer extension product; (e) extending the second primer with a DNA-dependent DNA polymerase, whereby a second primer extension product is produced to form a complex of first and second primer extension products; (f) cleaving RNA from the composite primer in the complex of first and second primer extension products with an agent (such as an enzyme) that cleaves RNA from an RNA/DNA hybrid such that a composite primer (which may or may not be the same as the first primer) hybridizes to the second primer extension product, wherein said composite primer comprises an RNA portion and a 3' DNA portion; (g) extending said composite primer of step (f) with a DNA-dependent DNA polymerase whereby said first primer extension product is displaced; (h) hybridizing the displaced first primer extension product with a polynucleotide comprising a propromoter and a region which is hybridizable to the displaced first primer extension product under conditions which allow transcription to occur by RNA polymerase, such that RNA transcripts are produced comprising sequences complementary to the displaced primer extension products, whereby multiple copies of the RNA sequence of interest are generated. In some embodiments, the invention provides methods of generating multiple copies of an RNA sequence of interest, said method comprising the steps of: (a) combining: a first complex, wherein the first complex is the complex of first and second primer extension products of step (e) described above in this paragraph; a composite primer (which may or may not be the same as the first primer) that is hybridizable to the second primer extension product, wherein said composite primer comprises an RNA portion and a 3' DNA portion; a DNA-dependent DNA polymerase; an RNA polymerase; a propromoter polynucleotide comprising a propromoter and a region which hybridizes to a second primer extension product; and an agent (such as an enzyme) that cleaves RNA from an RNA/DNA hybrid; and (b) incubating the mixture of step (a)

IDT Exhibit 1020.9254

under conditions (which include necessary substrates and buffer conditions) that permit primer hybridization, RNA cleavage, displacement of the first primer extension product from the first complex when its RNA is cleaved and a composite primer binds to the second primer extension product in the first complex, hybridization of the propromoter polynucleotide to the displaced first primer extension product to form a second complex comprising the displaced primer extension product and the propromoter polynucleotide, and RNA transcription from said second complex.

In another aspect, the invention provides methods of generating multiple copies of an RNA sequence of interest comprising incubating a reaction mixture, said reaction mixture comprising: (a) a target RNA; (b) a first primer that is hybridizable to the target RNA, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion; (c) a second primer that is hybridizable to an extension product of the first primer; (d) an RNA-dependent DNA polymerase; (e) a DNA-dependent DNA polymerase; (f) an RNA polymerase; (g) a propromoter polynucleotide comprising a propromoter and a region which hybridizes to an extension product of the second primer; and (h) an enzyme that cleaves RNA from an RNA/DNA hybrid; wherein the incubation is under conditions that permit primer hybridization, extension from the first primer and the second primer to form a first complex comprising first and second primer extension products, RNA cleavage, displacement of the first primer extension product from the first complex when its RNA is cleaved and a composite primer binds to the second primer extension product in the first complex, hybridization of the propromoter polynucleotide to the displaced first primer extension product to form a second complex comprising the displaced primer extension product and the propromoter polynucleotide, and RNA transcription from said second complex, whereby multiple copies of the RNA sequence of interest are generated.

In still another aspect, the invention provides methods of generating multiple copies of (amplifying) an RNA sequence of interest comprising incubating a reaction mixture, said reaction mixture comprising: (a) a target RNA; (b) a first primer that is hybridizable to the target RNA, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion; (c) a second primer that is hybridizable to an extension product of the first primer; (d) an RNA-dependent DNA polymerase; (e) a DNA-dependent DNA polymerase; and (f) an enzyme that cleaves RNA from an RNA/DNA hybrid; wherein the incubation is under conditions (which include necessary substrates and buffer conditions) that permit primer hybridization, extension from the first primer and the second primer to

IDT Exhibit 1020.9255

form a complex comprising first and second primer extension products, RNA cleavage, and displacement of the first primer extension product from the complex when its RNA is cleaved and another first primer binds to the second primer extension product in the complex, whereby multiple copies of a polynucleotide sequence complementary to the RNA sequence of interest are generated.

In another aspect, the invention provides methods of generating multiple copies of (amplifying) an RNA sequence of interest comprising: (a) incubating a reaction mixture, said reaction mixture comprising: (i) a target RNA; and (ii) a first primer that is hybridizable to a target RNA, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion; and (iii) a RNA-dependent DNA polymerase; wherein the incubation is under conditions that permit primer hybridization and formation of a complex comprising a first primer extension product and the target RNA; (b) incubating a reaction mixture, said reaction mixture comprising: (i) the complex comprising a first primer extension product and the target RNA; and (ii) an enzyme capable of cleaving RNA from an RNA/DNA hybrid; wherein the incubation is under conditions that permit cleavage of RNA in the complex comprising a first primer extension product and the target RNA; (c) incubating a reaction mixture, said reaction mixture comprising: (i) the first primer extension product; (ii) a second primer that is hybridizable to the first primer extension product; and (iii) DNA-dependent DNA polymerase; wherein the incubation is under conditions that permit primer hybridization and formation of a complex comprising the first primer extension product and a second primer extension product; (d) incubating a reaction mixture, said reaction mixture comprising: (i) the complex comprising the first primer extension product and a second primer extension product; (ii) an enzyme capable of cleaving RNA from an RNA/DNA hybrid; wherein the incubation is under conditions that permit cleavage of RNA in the complex comprising the first primer extension product and a second primer extension product; (e) incubating a reaction mixture, said reaction mixture comprising: (i) a composite primer, wherein the composite primer comprises a RNA portion and a 3' DNA portion; (ii) the cleaved complex comprising the first primer extension product and a second primer extension product; and (iii) DNA-dependent DNA polymerase; wherein the incubation is under conditions that permit composite primer hybridization, and displacement of the first primer extension product from the complex comprising the first primer extension product and a second primer extension product;

IDT Exhibit 1020.9256

whereby multiple copies of a polynucleotide sequence complementary to the RNA sequence of interest are generated.

The reaction mixtures may be combined (thus reducing the number of incubations) in any way, with one or more reaction mixtures above combined. Accordingly, in some embodiments, the reaction mixtures of steps (a) and (b) are the same reaction mixture (four incubations or incubation steps). In other embodiments, the reaction mixtures of steps (d) and (e) are the same reaction mixture (four incubations or incubation steps). In still another embodiment, the reaction mixtures of steps (a) and (b) are the same reaction mixture and the reaction mixtures of steps (d) and (e) are the same reaction mixture (three incubations or incubation steps). In still other embodiments, the reaction mixtures of steps (a)-(c) are the same reaction mixture (three incubations or incubation steps). In still another embodiment, the reaction mixtures of steps (a)-(c) are the same reaction mixture and the reaction mixtures of steps (d) and (e) are the same reaction mixture (two incubations or incubation steps). In other embodiments, the reaction mixtures of steps (a)-(d) are the same reaction mixture (two incubations or incubation steps). In other embodiments, the reaction mixtures of (b) and (c) are the same reaction mixture (four incubations or incubation steps). In other embodiments, the reaction mixtures of (c) and (d) are the same reaction mixture (four incubations or incubation steps). In yet another embodiment, the reaction mixtures of (a)-(e) are the same reaction mixture (one incubation). It is understood that any combination of these incubation steps, and any single incubation step, to the extent that the incubation is performed as part of any of the methods described herein, fall within the scope of the invention.

In another aspect of the invention, the methods of generating multiple copies of (amplifying) an RNA sequence of interest further comprise:(a) incubating a reaction mixture, said reaction mixture comprising: (i) a copy of a polynucleotide sequence complementary to the RNA sequence of interest; (ii) a propromoter polynucleotide comprising a propromoter and a region which hybridizes to the copy of a polynucleotide sequence complementary to the RNA sequence of interest; and RNA polymerase; and wherein the incubation is under conditions that permit hybridization of the propromoter polynucleotide to the copy of a polynucleotide sequence complementary to the RNA sequence of interest to form a second complex comprising the copy of a polynucleotide sequence complementary to the RNA sequence of interest and the propromoter polynucleotide, and RNA transcription from said second complex.

IDT Exhibit 1020.9257

In yet another aspect, the invention provides methods of generating multiple copies of (amplifying) an RNA sequence of interest, said method comprising the steps of: (a) combining: a target RNA; a first primer that is hybridizable to the target RNA, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion; a second primer that is hybridizable to an extension product of the first primer; an RNA-dependent DNA polymerase; a DNA-dependent DNA polymerase; an RNA polymerase; a propromoter polynucleotide comprising a propromoter and a region which hybridizes to an extension product of the second primer; and an agent (such as an enzyme) that cleaves RNA from an RNA/DNA hybrid; and (b) incubating the mixture of step (a) under conditions (which includes necessary substrates and buffer conditions) that permit primer hybridization, extension from the first primer and the second primer to form a first complex comprising first and second primer extension products, RNA cleavage, displacement of the first primer extension product from the first complex when its RNA is cleaved and another first primer binds to the second primer extension product in the first complex, hybridization of the propromoter polynucleotide to the displaced first primer extension product to form a second complex comprising the displaced primer extension product and the propromoter polynucleotide, and RNA transcription from said second complex, whereby multiple copies of the RNA sequence of interest are generated.

In yet another aspect, the invention provides methods of generating multiple copies of (amplifying) a polynucleotide sequence complementary to an RNA sequence of interest, said method comprising the steps of: (a) hybridizing a primer to a target RNA, wherein the primer is a composite primer comprising an RNA portion and a 3' DNA portion; (b) extending the primer with an RNA-dependent DNA polymerase, whereby a complex comprising a primer extension product and the target RNA is produced; (c) cleaving RNA in the complex of step (b) with an agent (such as an enzyme) that cleaves RNA from an RNA/DNA hybrid, such that at least one fragment of the target RNA remains hybridized to the primer extension product; (d) extending the at least one fragment of the target RNA with a DNA-dependent DNA polymerase, whereby a fragment extension product comprising the sequence of interest is produced to form a complex of primer and fragment extension products; (e) cleaving RNA from the composite primer in the complex of primer and fragment extension products with an agent (such as an enzyme) that cleaves RNA from an RNA/DNA hybrid such that a composite primer (which may or may not be the same as the first primer) hybridizes to the fragment extension product and repeats primer extension

IDT Exhibit 1020.9258

by strand displacement, wherein said composite primer comprises an RNA portion and a 3' DNA portion; whereby multiple copies of a polynucleotide sequence complementary to the RNA sequence of interest are generated. In some embodiments, the invention provides methods of generating multiple copies of a polynucleotide sequence complementary to an RNA sequence of interest, said method comprising the steps of: (a) combining: the complex of primer and fragment extension product of step (d) described above in this paragraph; a composite primer (which may or may not be the same as the first primer) that is hybridizable to the fragment extension product, wherein said composite primer comprises an RNA portion and a 3' DNA portion; a DNA-dependent DNA polymerase; and an agent (such as an enzyme) that cleaves RNA from an RNA/DNA hybrid; and (b) incubating the mixture of step (a) under conditions (which include necessary substrates and buffer conditions) that permit primer hybridization, RNA cleavage, and displacement of the primer extension product from the complex of step (d) described above in this paragraph when its RNA is cleaved and a composite primer binds to the fragment extension product in the complex.

In yet another aspect, the invention provides methods of generating multiple copies of (amplifying) an RNA sequence of interest, said method comprising the steps of: (a) hybridizing a first primer to a target RNA, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion; (b) extending the first primer with an RNA-dependent DNA polymerase, whereby a complex comprising a first primer extension product and the target RNA is produced; (c) cleaving the target RNA in the complex of step (b) with an agent (such as an enzyme) that cleaves RNA from an RNA/DNA hybrid, such that fragments of the target RNA remains hybridized to the first primer extension product; (d) extending the at least one fragment of the target RNA with a DNA-dependent DNA polymerase, whereby a fragment extension product comprising the sequence of interest is produced to form a complex of primer and fragment extension products; (e) cleaving RNA from the composite primer in the complex of primer and fragment extension products with an agent (such as an enzyme) that cleaves RNA from an RNA/DNA hybrid such that a composite primer (which may or may not be the same as the first primer) hybridizes to the fragment extension product and repeats primer extension by strand displacement, wherein said composite primer comprises an RNA portion and a 3' DNA portion; (f) hybridizing a displaced primer extension product with a polynucleotide comprising a propromoter and a region which is hybridizable to the displaced primer

IDT Exhibit 1020.9259

extension product under conditions which allow transcription to occur by RNA polymerase, such that RNA transcripts are produced comprising sequences complementary to the displaced primer extension product, whereby multiple copies of the RNA sequence of interest are generated. In some embodiments, the invention provides methods of generating multiple copies of an RNA sequence of interest, said method comprising the steps of: (a) combining: a first complex, wherein the first complex is the complex of primer and fragment extension product of step (d) described above in this paragraph; a composite primer (which may or may not be the same as the first primer) that is hybridizable to the fragment extension product, wherein said composite primer comprises an RNA portion and a 3' DNA portion; a DNA-dependent DNA polymerase; an RNA polymerase; a propromoter polynucleotide comprising a propromoter and a region which hybridizes to a second primer extension product; and an agent (such as an enzyme) that cleaves RNA from an RNA/DNA hybrid; and (b) incubating the mixture of step (a) under conditions (which include necessary substrates and buffer conditions) that permit primer hybridization, RNA cleavage, displacement of the primer extension product from the first complex when its RNA is cleaved and a composite primer binds to the fragment extension product in the first complex, hybridization of the propromoter polynucleotide to the displaced primer extension product to form a second complex comprising the displaced primer extension product and the propromoter polynucleotide, and RNA transcription from said second complex.

In yet another aspect, the invention provides methods of generating multiple copies of (amplifying) a polynucleotide sequence complementary to an RNA sequence of interest, said method comprising the steps of: (a) combining: a target RNA; a primer that is hybridizable to the target RNA, wherein the primer is a composite primer comprising an RNA portion and a 3' DNA portion; an RNA-dependent DNA polymerase; a DNA-dependent DNA polymerase; and an agent (such as an enzyme) that cleaves RNA from an RNA/DNA hybrid; and (b) incubating the mixture of step (a) under conditions (which includes necessary substrates and buffer conditions) that permit primer hybridization, RNA cleavage, wherein RNA cleavage of a first complex comprising a target RNA and a primer extension product is such that a fragment of the target RNA remains hybridized to the primer extension product, primer extension from the primer and the fragment of the target RNA to form a second complex comprising a primer extension product and a fragment extension product that comprises the sequence of interest, and displacement of the primer

IDT Exhibit 1020.9260

extension product from the second complex when its RNA is cleaved and another primer binds to the fragment extension product in the second complex, whereby multiple copies of a polynucleotide sequence complementary to the RNA sequence of interest are generated.

In one other aspect, the invention provides methods of generating multiple copies of (amplifying) an RNA sequence of interest, said method comprising the steps of: (a) combining: a target RNA; a primer that is hybridizable to the target RNA, wherein the primer is a composite primer comprising an RNA portion and a 3' DNA portion; an RNA-dependent DNA polymerase; a DNA-dependent DNA polymerase; an RNA polymerase; a propromoter polynucleotide; and an agent (such as an enzyme) that cleaves RNA from an RNA/DNA hybrid; and (b) incubating the mixture of step (a) under conditions (which includes necessary substrates and buffer conditions) that permit primer hybridization, RNA cleavage, wherein RNA cleavage of a first complex comprising a target RNA and a primer extension product occurs such that a fragment of the target RNA remains hybridized to the primer extension product, primer extension from the primer and the fragment of the target RNA to form a second complex comprising a primer extension product and a fragment extension product that comprises the sequence of interest, displacement of the primer extension product from the second complex when its RNA is cleaved and another primer binds to the second complex, hybridization of the propromoter polynucleotide to a displaced primer extension product to form a third complex comprising the displaced primer extension product and the propromoter polynucleotide, and transcription from said third complex, whereby multiple copies of the RNA sequence of interest are generated.

In another aspect, the invention provides methods of generating multiple copies of an RNA sequence of interest present on a target RNA, said method comprising: formation of a complex of first and second primer extension products comprising a 3' single stranded portion according to any of the methods described herein; (c) hybridizing a propromoter oligonucleotide to the 3' single stranded portion described in step (b); and (d) transcription using DNA-dependent RNA polymerase, whereby multiple copies of sense RNA products are generated.

In another aspect, the invention provides methods of generating multiple copies of a sequence complementary to an RNA sequence of interest present on a target RNA and multiple copies of an RNA sequence of interest present on a target RNA, said methods comprising the steps of: (a) extending a first primer hybridized to a target RNA with an RNA-dependent DNA polymerase, wherein the first primer is a composite primer

IDT Exhibit 1020.9261

comprising an RNA portion and a 3' DNA portion, whereby a complex comprising a first primer extension product and the target RNA is produced; (b) cleaving RNA in the complex of step (b) with an enzyme that cleaves RNA from an RNA/DNA hybrid; (c) extending a second primer hybridized to the first primer extension product with a DNA-dependent DNA polymerase, wherein the second primer is a composite primer comprising an RNA portion and a 3' DNA portion, whereby a second primer extension product is produced to form a complex of first and second primer extension products; (d) cleaving RNA from the first and second composite primers in the complex of first and second primer extension products with an enzyme that cleaves RNA from an RNA/DNA hybrid such that another composite primer hybridizes to the second primer extension product and another composite primer hybridizes to the first primer extension product; (e) extending said composite primers of step (d) with a DNA-dependent DNA polymerase; whereby said first primer extension product is displaced, and whereby multiple copies of a polynucleotide sequence complementary to the RNA sequence of interest are generated; and whereby said second primer extension product is displaced, and whereby multiple copies of the RNA sequence of interest are generated.

In another aspect, the methods of generating multiple copies of a polynucleotide sequence complementary to an RNA sequence of interest comprise generating multiple copies of a polynucleotide sequence complementary to of two or more different sequences of interest.

In another aspect, the methods of generating multiple copies of an RNA sequence of interest comprise generating multiple copies of two or more different sequences of interest.

Various embodiments of the composite primer and second primer used in the methods of the invention are described herein. For example, in some embodiments, the RNA portion of a 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. In some embodiments, a composite primer comprises a 5' portion (for example, a 5' RNA portion) that is not hybridizable to a target RNA under conditions which the composite primer hybridizes to target RNA. In yet other embodiments, a composite primer comprises a 3' portion (for example, a 3' DNA portion) that comprises a random sequence. In some embodiments wherein a target RNA is mRNA, a composite primer may comprise a poly-dT sequence. In other embodiments, a composite primer is a random primer. In still other

IDT Exhibit 1020.9262

embodiments, a plurality of composite primers are used for hybridizing to the target RNA. In some embodiments, the composite primer that hybridizes to target RNA and the composite primer that hybridizes to second primer extension product are the same. In some embodiments, the composite primer that hybridizes to target RNA and the composite primer that hybridizes to second primer extension product are different.

In still other embodiments, the second primer is a primer comprising DNA (in some embodiments, consisting of DNA). In other embodiments, the second primer comprises one or more fragments of target RNA hybridized to the first primer extension product, said one or more fragments generated by cleaving RNA in the complex of target RNA and first primer extension product with an enzyme that cleaves RNA from an RNA/DNA hybrid. In other embodiments, the second primer comprises a portion (for example, a 3' portion) that comprises a random sequence. In yet another embodiment, the second primer is a random primer. In some embodiments, the second primer comprises a 5' portion that is not hybridizable to a first primer extension product. For the methods described herein, one or more composite primers or second primers can be used.

The enzymes which may be used in the methods and compositions are described herein. For example, the agent (such as an enzyme) that cleaves RNA may be an RNaseH, and the RNA-dependent DNA polymerase may be reverse transcriptase. The RNA-dependent DNA polymerase may comprise an RNase H enzyme activity, or separate enzymes may be used. Similarly, a DNA polymerase may comprise both RNA-dependent and DNA-dependent DNA polymerase enzyme activities, or separate enzymes may be used. A DNA-dependent DNA polymerase and an enzyme that cleaves RNA may also be the same enzyme. A DNA-dependent DNA polymerase, an RNA-dependent DNA polymerase, and the enzyme that cleaves RNA can also be the same enzyme.

In some embodiments, methods of the invention are used to generate labeled polynucleotide products (generally DNA or RNA products). In some embodiments of methods for generating labeled DNA products, at least one type of dNTP used is a labeled dNTP. In some embodiments of methods for generating labeled RNA products, at least one type of rNTP used is a labeled rNTP. In other embodiments of methods for generating labeled DNA products, a labeled composite primer is used.

In some aspects, a propromoter polynucleotide (for example, a PTO) comprises a region at the 3' end which hybridizes to the displaced primer extension product, whereby

IDT Exhibit 1020.9263

DNA polymerase extension of displaced extension product produces a double stranded promoter from which transcription occurs.

The methods are applicable to amplifying any RNA target, including, for example, mRNA and ribosomal RNA. One or more steps may be combined and/or performed sequentially (often in any order, as long as the requisite product(s) are able to be formed), and, as is evident, the invention includes various combinations of the steps described herein. It is also evident, and is described herein, that the invention encompasses methods in which the initial, or first, step is any of the steps described herein. For example, the methods of the invention do not require that the first step be production of the first primer extension product from the RNA template. Methods of the invention encompass embodiments in which later, "downstream" steps are an initial step.

Further, in various embodiments of the invention, it is understood that the first primer extension product comprises (a) a sequence complementary to an RNA sequence and (b) a 5' portion, preferably a 5' end, that is RNA. As described, this product generally arises from primer extension of a composite primer with an RNA portion along an RNA template. As such, reference to a first primer extension product refers to a product comprising (a) and (b) above.

The invention also provides methods which employ (usually, analyze) the products of the amplification methods of the invention, such as sequencing, detection of sequence alteration(s) (e.g., genotyping or nucleic acid mutation detection); determining presence or absence of a sequence of interest; gene expression profiling; subtractive hybridization; preparation of a subtractive hybridization probe; differential amplification; preparation of libraries (including cDNA and differential expression libraries); preparation of an immobilized nucleic acid (which can be a nucleic acid immobilized on a microarray), and characterizing (including detecting and/or quantifying) amplified nucleic acid products generated by the methods of the invention.

In some aspects, the invention provides methods of sequencing an RNA sequence of interest, said methods comprising amplifying a target RNA containing the sequence of interest by the amplification methods of the invention in the presence of a mixture of dNTPs and dNTP analogs (which may be labeled or unlabeled), such that primer extension is terminated upon incorporation of a dNTP analog which may be labeled or unlabeled, and analyzing the amplification products to determine sequence. In embodiments wherein amplified products are RNA transcripts, the methods may comprise (a) amplifying a target

IDT Exhibit 1020.9264

RNA containing the sequence of interest by the amplification methods of the invention in the presence of a mixture of rNTPs and rNTP analogs (which may be labeled or unlabeled), such that RNA transcription is terminated upon incorporation of an rNTP analog which may be labeled or unlabeled; and (b) analyzing the amplification products to determine sequence.

In some aspects, the invention provides methods of detecting a mutation (or, in some aspects, characterizing a sequence) in a target RNA, comprising (a) amplifying the target RNA by a method described herein; and (b) analyzing the amplification products of the method for single stranded conformation, wherein a difference in conformation as compared to a reference single stranded RNA indicates a mutation in the target RNA. In other embodiments, the invention provides methods of detecting a mutation (or, in some aspects, characterizing a sequence) in a target RNA comprising analyzing amplification products of any of the methods described herein for single stranded conformation, wherein a difference in conformation as compared to a reference single stranded RNA indicates a mutation in the target RNA (or, in some aspects, characterizes the target sequence).

In another aspect, the invention provides methods of determining presence or absence of a sequence of interest, said methods comprising (i) amplifying a target RNA containing the sequence of interest, said amplification comprising extending a composite primer hybridized to cleaved complex of first and second primer extension product prepared by any of the methods described here, 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 hybridizable to the RNA portion of the composite primer indicates that the second primer extension product does not comprise a sequence hybridizable to the RNA portion of the composite primer and is a sequence variant with respect to the sequence hybridizable 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 hybridizable to the RNA portion of the composite primer indicates that the second primer extension product comprises a sequence hybridizable to the RNA portion of the composite primer and is not a

IDT Exhibit 1020.9265

sequence variant with respect to the sequence hybridizable to the RNA portion of the composite primer.

In another aspect, the invention provides methods of producing a nucleic acid immobilized to a substrate (which includes methods of producing a microarray), comprising (a) amplifying a target RNA by any of the methods described herein; and (b) immobilizing the amplification products on a substrate. The amplification products can be labeled or unlabeled. In other aspects, the invention provides methods of producing a microarray, comprising (a) amplifying a target RNA by an amplification method described herein; and (b) immobilizing the amplification products on a substrate (which can be solid or semi-solid). In some embodiments, microarrays are produced by immobilizing amplification products onto a substrate to make a microarray of amplification products. The microarray can comprise at least one amplification product immobilized on a solid or semi-solid substrate fabricated from a material selected from the group consisting of paper, glass, ceramic, plastic, polystyrene, polypropylene, nylon, polyacrylamide, nitrocellulose, silicon and other metals, and optical fiber. An amplification product can be immobilized on the solid or semi-solid substrate in a two-dimensional configuration or a three-dimensional configuration comprising pins, rods, fibers, tapes, threads, beads, particles, microtiter wells, capillaries, and cylinders.

Any of the methods of the invention can be used to generate polynucleotide (generally, DNA or RNA) products that are suitable for characterization of an RNA sequence of interest in a sample. In one embodiment, the invention provides methods for characterizing (for example, detecting (presence or absence) and/or quantifying) an RNA sequence of interest comprising: (a) amplifying a target RNA by any of the methods described herein; and (b) analyzing the amplification products. Step (b) of analyzing the amplification products can be performed by any method known in the art or described herein, for example by detecting and/or quantifying amplification products that are hybridized to a probe. These amplification products may or may not be labeled. Any of the methods of the invention can be used to generate DNA or RNA products that are labeled by incorporating labeled nucleotides and/or labeled composite primers into appropriate step(s) of the methods. These labeled products are particularly suitable for quantification and/or identification by methods known in the art, which include the use of arrays such as cDNA microarrays and oligonucleotide arrays. In one aspect, the invention provides a method of characterizing an RNA sequence of interest, comprising (a)

IDT Exhibit 1020.9266

amplifying a target RNA by a method described herein to generate labeled DNA products; and (b) analyzing the labeled DNA products. In some embodiments, the step of analyzing DNA products comprises determining amount of said products, whereby the amount of the RNA sequence of interest present in a sample is quantified.

In another aspect, the invention provides a method of characterizing an RNA sequence of interest, comprising (a) amplifying a target RNA by a method described herein to generate labeled RNA products; and (b) analyzing the labeled RNA products. In some embodiments, the step of analyzing RNA products comprises determining amount of said products, whereby the amount of the RNA sequence of interest present in a sample is quantified. The DNA or RNA products can be analyzed by, for example, contacting them with at least one probe. In some embodiments, the at least one probe is provided as a microarray. The microarray can comprise at least one probe immobilized on a solid or semi-solid substrate fabricated from a material selected from the group consisting of paper, glass, ceramics, plastic, polypropylene, polystyrene, nylon, polyacrylamide, nitrocellulose, silicon, other metals, and optical fiber. A probe can be immobilized on the solid or semi-solid substrate in a two-dimensional configuration or a three-dimensional configuration comprising pins, rods, fibers, tapes, threads, beads, particles, microtiter wells, capillaries, and cylinders.

In another aspect, the invention provides methods of determining gene expression profile in a sample, the methods comprising: (a) amplifying at least one RNA sequence of interest in the sample using any of the methods described herein; and (b) determining amount of amplification products of each RNA sequence of interest, wherein each said amount is indicative of amount of each RNA sequence of interest in the sample, whereby the gene expression profile of the sample is determined.

In another aspect, the invention provides methods of preparing a library (including cDNA and subtractive hybridization libraries), said methods comprising: amplifying at least one RNA sequences of interest using any of the methods described herein to generate a single or double stranded nucleic acid product.

In another aspect, the invention provides methods of preparing a subtractive hybridization probe, said methods comprising generating multiple single stranded polynucleotide, preferably DNA, copies of the complement of at least one RNA sequences of interest from a first RNA population using any of the methods described herein.

IDT Exhibit 1020.9267

In another aspect, the invention provides methods of performing subtractive hybridization, said methods comprising: (a) generating multiple copies of the complement of at least one RNA sequences of interest from a first RNA population using any of the methods described herein; and (b) hybridizing the multiple copies to a second mRNA population, whereby a subpopulation of the second mRNA population forms a complex with a DNA copy. In some embodiments, the methods further comprise: (c) cleaving RNA in the complex of step (b) with an enzyme that cleaves RNA from an RNA/DNA hybrid; and (d) amplifying an unhybridized subpopulation of the second mRNA population (using any method, including the methods described herein), whereby multiple copies of single stranded DNA complementary to the unhybridized subpopulation of the second mRNA population are generated.

In another aspect, the invention provides methods for differential amplification, the methods comprising: (a) generating multiple nucleic acid, generally DNA, copies of the complement of at least one RNA sequence of interest from a first RNA population using any of the methods described herein; (b) hybridizing the multiple copies to a second mRNA population, whereby a subpopulation of the second mRNA population forms a complex with a DNA copy; (c) cleaving RNA in the complex of step (b) with an enzyme that cleaves RNA from an RNA/DNA hybrid; and (d) amplifying an unhybridized subpopulation of the second mRNA population using any method, including those described herein, whereby multiple copies of single stranded DNA complementary to the unhybridized subpopulation of the second mRNA population are generated.

In another aspect, the invention provides methods for making a library, said method comprising preparing a subtractive hybridization probe as described herein, or differential amplification as described herein.

Any of these applications can use any of the amplification methods (including various components and various embodiments of any of the components) as described herein. For example, the composite primer used may have a 5' RNA portion, which may be adjacent to the 3' DNA portion.

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. In one aspect, for example, the invention provides compositions comprising a composite primer, wherein the composite primer comprises an RNA portion and a 3' DNA portion, and a second primer, wherein the

IDT Exhibit 1020.9268

second primer is a random primer. In some embodiments, these compositions further comprise an RNA-dependent DNA polymerase (which can be a reverse transcriptase). In another embodiment, the invention provides a composition comprising a composite primer and a second primer that comprises a sequence that is not hybridizable (under conditions in which a region of primer is hybridizable) to a composite primer extension product (generally, but not necessarily, the use of this primer results in generating primer extension products to which a propromoter polynucleotide can hybridize). In some embodiments, the 5' RNA portion of a composite primer is adjacent to its 3' DNA portion. In still other embodiments, the 5' RNA portion of a composite primer is about 5 to about 20 nucleotides and its 3' DNA portion is about 5 to about 15 nucleotides. In some embodiments, the propromoter polynucleotide is a propromoter template oligonucleotide (PTO). In still other embodiments, the invention provides a composition comprising: (a) a composite primer; (b) a second primer (which can be a random primer); and (c) a propromoter polynucleotide (which in some embodiments is a PTO).

In another aspect, the invention provides compositions comprising any of the complexes (which are generally considered as intermediates with respect to the final amplification products) described herein (see also the figures for exemplary schematic depictions of these various complexes). For example, the invention provides compositions comprising a complex of: (a) a first primer extension strand, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion; and (b) a target RNA strand. In yet another aspect, the invention provides compositions comprising a complex of: (a) a first primer extension product, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion; and (b) a second primer (or fragment) extension product. In still another aspect, the invention provides a complex of (a) a cleaved primer extension product, wherein the primer is a composite primer comprising an RNA portion and a 3' DNA portion; (b) a second primer (or fragment) extension product; and (c) a composite primer.

In another aspect, the invention includes any one or more products (including intermediates) and compositions comprising the products (including intermediates) produced by any aspect of the methods of the invention. The products include libraries and any other population produced, which are generally based on the nature of the primer(s) used in the methods described herein.

IDT Exhibit 1020.9269

In another aspect, the invention provides reaction mixtures (or compositions comprising reaction mixtures) which contain various combinations of components described herein. For example, the invention provides reaction mixtures comprising (a) a target RNA; (b) a composite primer comprising a 3' DNA portion and an RNA portion; (c) a second primer; and (d) a DNA polymerase. As described herein, any of the composite primers may be in the reaction mixture (or a plurality of composite primers), including a composite primer that comprises a 5' RNA portion which is adjacent to the 3' DNA portion. The reaction mixture could also further comprise an enzyme which cleaves RNA from an RNA/DNA hybrid, such as RNase H. A reaction mixture of the invention can also comprise a propromoter polynucleotide (which in some embodiments is a PTO), and/or an RNA polymerase. A reaction mixture of the invention can also comprise (a) a displaced primer extension product (which contains a 5' end sequence complementary to the 3' DNA portion of a composite primer, but not sequences complementary to the RNA portion of a composite primer); (b) a propromoter polynucleotide (such as a PTO); and (c) an RNA polymerase.

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 comprising a 3' DNA portion and an RNA portion (which may be 5' and may further be adjacent to the 3' DNA portion), and a second primer (which may or may not be separately packaged). In some embodiments, these kits further comprise instructions for using the primers to amplify RNA. The composite primer in the kits can be any described herein. 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. Any of these kits can further comprise instructions for using the components to amplify RNA.

In another aspect, the invention provides systems for effecting the amplification methods described herein. For example, the invention provides systems for amplifying a target RNA, comprising (a) a composite primer comprising a 3' DNA portion and an RNA portion; (b) a second primer; (c) an RNA-dependent DNA polymerase; (d) a DNA-dependent DNA polymerase; and (e) an enzyme which cleaves RNA from an RNA/DNA

IDT Exhibit 1020.9270

hybrid (such as RNaseH). The composite primer may be any (one or more) described herein, including a composite primer which comprises a 5' RNA portion which is adjacent to the 3' DNA portion. The systems can further comprise a propromoter polynucleotide (such as a PTO) and/or an RNA polymerase. As described herein, systems of the invention generally comprise one or more apparatuses appropriate for carrying out methods of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGURES 1A-1B show a diagrammatic representation of a linear isothermal RNA amplification process using a composite primer, a second primer, and strand displacement to generate multiple copies of single stranded DNA product comprising sequences complementary to the target RNA.

FIGURES 2A-C show a diagrammatic representation of an enhanced linear isothermal RNA amplification process using a composite primer, a second primer, strand displacement and transcription to generate multiple copies of the target RNA.

FIGURES 3A-B show a diagrammatic representation of a linear isothermal RNA amplification process using a single composite primer and strand displacement to generate multiple copies of single stranded DNA product comprising sequences complementary to the target RNA.

FIGURES 4A-B show a diagrammatic representation of an enhanced linear isothermal RNA amplification process using a single composite primer, strand displacement and transcription to generate multiple copies of the target RNA.

FIGURE 5 shows a diagrammatic representation of a linear isothermal RNA amplification process using a single composite primer containing a poly-dT sequence and strand displacement to generate multiple copies of single stranded DNA product comprising sequences complementary to the target RNA.

FIGURE 6 shows a diagrammatic representation of a linear isothermal RNA amplification process using a single composite primer containing a random sequence and strand displacement to generate multiple copies of single stranded DNA product comprising sequences complementary to the target RNA.

IDT Exhibit 1020.9271

FIGURE 7 shows a diagrammatic representation of an enhanced isothermal RNA amplification process using a single composite primer and transcription to generate multiple copies of RNA product comprising sequences complementary to the target RNA.

FIGURES 8A-E show a diagrammatic representation of a linear isothermal RNA amplification process using two composite primers and strand displacement to generate multiple copies of single stranded DNA products comprising sequences complementary to the target RNA and sequences identical to the target RNA.

FIGURE 9 shows a photograph of a gel showing PCR products amplified from double stranded cDNA comprising an appended defined sequence on the second strand cDNA prepared using linear isothermal RNA amplification.

FIGURE 10 shows graphs depicting results of real-time PCR experiments quantifying products generated using linear isothermal RNA amplification.

MODES FOR CARRYING OUT THE INVENTION

Overview of the Invention and Its Advantages

The invention discloses novel methods, compositions and kits for amplification of RNA sequences. The methods provide for isothermal amplification of a single RNA species or a pool of RNA species. Some methods provide for generation of multiple copies of DNA comprising sequences complementary to an RNA sequence of interest. Other methods provide for generation of multiple copies of an RNA sequence of interest. These methods are suitable for, for example, generation of cDNA libraries and subtractive hybridization probes. They generate single stranded DNA or RNA products, which are readily suitable for a variety of uses including expression profiling, e.g., by multiplex analysis by microarray technologies, as well as electrophoresis-based technologies such as differential display. The methods are amenable to automation and do not require thermal cycling.

The methods of the invention are directed to the amplification of one or more species of RNA, such as a pool of RNA sequences, and are especially suitable for the amplification of all RNA (such as mRNA) sequences in a preparation of total RNA from a biological sample. Thus, one of the major advantages of the methods of the invention is the ability to amplify an entire pool of sequences, which is essential for the ability to

IDT Exhibit 1020.9272

analyze the gene expression profile in cells, such as the cells in a biological sample of interest. The methods of the invention may also be used to amplify a specific RNA sequence of interest, or a multiplicity of RNAs, for example, members of a family of related RNAs. The methods of the invention also are suitable for amplifying a large multiplicity, and most preferably all RNA (such as mRNA) sequences in a sample.

Insofar as many mRNAs have a unique poly-A 3'-end, amplification initiated from the 3'-end sequence of mRNAs is commonly performed for preparation of cDNA libraries and subsequent sequence analysis for determination of gene expression profiling or other applications. The methods of the invention are similarly suited for preparation of libraries of amplified 3'-portions of mRNAs. A composite primer used in the methods of invention can be designed to be hybridizable to a multiplicity, or all, of the mRNA species in the sample by using random sequences, according to methods known in the art. Alternatively, if a selected RNA or family of related RNAs are to be amplified, the composite primer will comprise sequence(s) hybridizable to the selected RNA or family of related RNAs.

The methods generally comprise using specially-designed primers, generally one or more RNA/DNA composite primers, to generate a complex of first and second strand cDNAs that comprise a portion with a particular characteristic (e.g., cleavable by an enzyme). As used herein, it is understood that "cDNA" refers to a polynucleotide primer extension product. Generally, the complex comprises an RNA/DNA heteroduplex at an end of the double stranded cDNA complex. The RNA/DNA heteroduplex at an end of the double stranded cDNA complex may comprise a defined end sequence, generally introduced by the RNA portion of the composite primer. The composite primer according to the methods of the invention comprises a 3'-DNA portion that generally is designed to be hybridizable to a target RNA(s). The remaining portion(s) (such as the 5' RNA portion) of the composite primer generally, but not necessarily, comprises a sequence that is not hybridizable to a target RNA (which would constitute a tail when the primer is bound to a target RNA). Thus, and as the description herein makes clear, reference to a primer that hybridizes to a sequence (or hybridization template) encompasses embodiments in which at least a portion of the primer is hybridized, as well as those embodiments in which an entire primer is hybridized.

The double stranded cDNA complex is a substrate for linear amplification as follows: an enzyme which cleaves RNA from an RNA/DNA hybrid (such as RNase H) cleaves RNA sequence from the hybrid, leaving a 3' DNA sequence on the second strand

IDT Exhibit 1020.9273

cDNA available for binding by a composite primer which may or may not be the same as the first composite primer. Extension of a bound composite primer by DNA polymerase produces a primer extension product, which displaces the previously bound cleaved first primer extension product, whereby single stranded DNA product accumulates. The single stranded DNA product is a copy of the complement of the target RNA (or "antisense" DNA). This linear amplification is referred to as "SPIA" (for Single Primer Linear Amplification), and is described in Kurn et al., U.S. Patent No. 6,251,639 B1.

In one aspect, the invention works as follows: a composite RNA/DNA primer forms the basis for replication of the template RNA The composite primer (also referred to herein as "the first primer") hybridizes to template RNA which comprises the RNA sequence(s) of interest, and the composite primer is extended by an RNA-dependent DNA polymerase to form a first primer extension product (interchangeably called "composite primer extension product", or "first-strand cDNA"). After cleavage of the template RNA, a second primer extension product (interchangeably called "second-strand cDNA") is formed (as described below) in a complex with the first primer extension product. The complex of first and second primer extension products is composed of an RNA/DNA hybrid at one end due to the presence of the composite primer in the first primer extension product. An agent such as an enzyme which cleaves RNA from an RNA/DNA hybrid (such as RNase H) cleaves RNA sequence from the hybrid, leaving a sequence on the second primer extension product available for binding by another composite primer, which may or may not be the same as the first composite primer. Another first (composite) primer extension product is produced by DNA polymerase, which displaces the previously bound cleaved first primer extension product, resulting in displaced cleaved first primer extension product.

In some embodiments of the invention, the second primer extension product is formed as follows: following cleavage of the RNA template, a second primer is then hybridized to the first primer extension product and extended to form a second primer extension product in a complex with the first primer extension product. The complex of first and second primer extension products is composed of an RNA/DNA hybrid at one end due to the presence of the composite primer in the first primer extension product. The second primer is any sequence that is hybridizable to the first DNA strand such that it is capable of being extended by a DNA polymerase along a first primer extension product to create a second primer extension product. Thus, in some embodiments, the second primer is an oligonucleotide, which may or may not comprise a random sequence (i.e., a sequence

IDT Exhibit 1020.9274

designed to be hybridizable (under a given set of conditions) to one or more sequences in the sample). In other embodiments, it comprises a sequence of the first DNA strand (generally at the 3' end) that is hybridized to a sequence in the first DNA strand (for example, a hairpin or self-annealed structure).

In another aspect of the amplification methods, one or more fragments of the target RNA serves as the primer of the second primer extension product. The target RNA in the initial complex comprising the target RNA and first primer extension product is cleaved with an agent (such as RNase H) such that at least one fragment of the template RNA remains hybridized to the first primer extension product. In this aspect of the invention, one (or more) template RNA fragment(s) serves as a second "primer" in the manner described above, to generate a fragment extension product which has the same function as the second primer extension product in the amplification methods described above. A suitable RNA fragment in the methods of the invention is long enough such that it does not dissociate from the first strand cDNA, preferably from about 3 to about 30, more preferably from about 5 to about 25, even more preferably from about 10 to about 20, and most preferably from about 12 to about 17, nucleotides in length.

In embodiments involving transcription (referred to herein as "enhanced" methods), the second primer may further comprise a sequence such that displaced first primer extension products (other than the very first composite primer extension product) contain a sequence to which a polynucleotide comprising a propromoter (also referred to herein as "propromoter polynucleotide") is capable of hybridizing. Hybridization of the propromoter polynucleotide to a displaced primer extension product and extension of the 3' end of the displaced first primer extension product (if there is an overhang) results in a double stranded promoter region that drives transcription (via DNA-dependent RNA polymerase) to produce sense RNA products. This "enhanced" approach is described in Kurn et al., U.S. Patent No. 6,251,639 B1.

Accordingly, the invention provides methods of producing at least one copy of a polynucleotide sequence complementary to an RNA sequence of interest comprising combining and reacting the following: (a) a target RNA comprising an RNA sequence of interest; (b) a first (composite) primer comprising an RNA portion and a 3' DNA portion; (c) a second primer that is hybridizable to an extension product of the composite primer; (d) an RNA-dependent DNA polymerase; (e) a DNA-dependent DNA polymerase; (f) an agent (such as an enzyme) that cleaves RNA from an RNA/DNA hybrid; and

IDT Exhibit 1020.9275

(g) deoxyribonucleoside triphosphates or suitable analogs (which may or may not be labeled). In embodiments that include transcription (i.e., the enhanced methods), the following are also included in the amplification reaction (either at the same time as the components listed above or added separately): (h) a propromoter polynucleotide comprising a propromoter and a region which hybridizes to an extension product of the first primer (a displaced primer extension product); (i) an RNA polymerase; and (j) ribonucleoside triphosphates or suitable analogs (which may or may not be labeled). The combination is subjected to suitable conditions for primer hybridization, extension of primers, RNA cleavage, and displacement of the first primer extension product when its RNA is cleaved and another first primer binds in the site vacated by the cleaved RNA. In embodiments that include transcription, conditions employed are also suitable for hybridization of the propromoter polynucleotide to the displaced cleaved first primer extension product, extension of the 3' end of the cleaved first primer extension product (if necessary) to generate a double-stranded promoter region, and RNA transcription driven by the promoter. As described and exemplified herein, the above-described reaction mixture may be subdivided into two or more different reaction mixtures for separate, generally sequential, incubations that correspond to different aspects of the amplification process (see, for example, Example 1).

In another aspect of the amplification methods, fragments of the target RNA serves as a primer of second DNA strand synthesis. The target RNA in the initial complex comprising the target RNA and composite primer extension product is cleaved with an enzyme (such as RNaseH) such that at least one fragment of the template RNA remains hybridized to the composite primer extension product. In this aspect of the invention, one (or more) template RNA fragment(s) serves as a second "primer" in the manner described above, to generate a fragment extension product which has the same function as the second primer extension product in the amplification methods described above.

In some embodiments, the invention provides methods of producing at least one copy of a polynucleotide sequence complementary to an RNA sequence of interest comprising combining and reacting the following: (a) complex of first and second primer extension products; (b) a first (composite) primer comprising an RNA portion and a 3' DNA portion; (c) a DNA-dependent DNA polymerase; (d) an agent (such as an enzyme) that cleaves RNA from an RNA/DNA hybrid; and (e) deoxyribonucleoside triphosphates or suitable analogs (which may or may not be labeled). In embodiments that include

IDT Exhibit 1020.9276

transcription, the following are also included in the amplification reaction (either at the same time as the components listed above or added separately): (f) a propromoter polynucleotide comprising a propromoter and a region which hybridizes to an extension product of the first primer (a displaced primer extension product); (g) an RNA polymerase; and (h) ribonucleoside triphosphates or suitable analogs (which may or may not be labeled). The combination is subjected to suitable conditions for primer hybridization, extension of primers, RNA cleavage, and displacement of the first primer extension product when its RNA is cleaved and another first primer binds in the site on the second primer extension product vacated by the cleaved RNA . In embodiments that include transcription, conditions employed are also suitable for hybridization of the propromoter polynucleotide to the displaced first primer extension product, extension of the 3' end of the first primer extension product (if necessary) to generate a double-stranded promoter region, and RNA transcription driven by the promoter.

In another aspect, the invention provides methods of producing single stranded antisense and sense DNA copies of an RNA sequence of interest using a first composite primer, a second composite primer (termed the "reverse composite" primer), and a target RNA fragment. The method involves the following: (a) formation of a double stranded cDNA comprising a RNA-DNA heteroduplex at each end of the double stranded cDNA; and (b) linear amplification of first strand (sense) cDNA and second strand (antisense) cDNA by primer extension from two composite primers and strand displacement. Single stranded first and second strand cDNA product is produced. This product is useful in, e.g., producing cDNA libraries. As is evident, in this aspect of the invention, the second primer extension product is primed by a composite primer.

The methods of the invention include methods using the amplified products (so-called "applications"). In some embodiments, the invention provides methods of sequencing RNA sequences. For sequencing methods based on methods described herein wherein the amplified product is DNA, the appropriate dNTPs, or analogs thereof, which may be labeled or unlabeled, are used. For sequencing methods based on methods described herein wherein the amplified product is RNA, the appropriate rNTPs, or analogs thereof, which may be labeled or unlabeled, may be used.

In other embodiments, the invention provides methods of detecting nucleic acid sequence mutations. In one embodiment, the amplified products are used to detect and/or identify single strand conformation polymorphisms in a target polynucleotide.

IDT Exhibit 1020.9277

The invention provides methods to characterize (for example, detect presence or absence of and/or quantify) an RNA sequence of interest by generating DNA or RNA products using amplification methods of the invention, and analyzing the products by detection/quantification methods such as those based on array technologies or solution phase technologies. These amplified products may be labeled or unlabeled.

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

In another embodiment, the invention provides methods of generating cDNA libraries, methods of generating subtractive hybridization probes, and methods of generating subtractive hybridization libraries.

Various methods for the detection and quantification of gene expression levels have been developed in recent years. For example, microarrays, in which either defined cDNAs or oligonucleotides are immobilized at discrete locations on, for example, solid or semi-solid substrates, or on defined particles, enable the detection and/or quantification of the expression of a multitude of genes in a given specimen.

Using these previously known methods to detect presence of absence of and/or quantify multiple mRNA species in a sample, which in turn is used as a measure of gene expression profiling, generally requires direct labeling of cDNA, which requires a large amount of input total RNA, in part because mRNA represents only a small subset of the total RNA pool. Thus, when using total RNA preparations from a given cell or tissue sample, the analysis of gene expression in the sample using methods such as arrays requires a substantial amount of input RNA, which generally ranges from 50 to 200 μg. Similarly, 2 to 5 μg of mRNA purified from a total RNA preparation would generally be required for a single analysis of gene expression profiling using array technologies. This is a clear limitation of methods such as those based on array technology, insofar as the number of cells, or size of tissue specimen required is very large, and these cells or tissue specimens are often scarcely available for testing or are too precious. This limitation is especially severe in the study of clinical specimens, where the cells to be studied are rare and/or difficult to cultivate *in vitro*, and in high throughput screening of libraries of effector molecules. Also, previous transcription-based methods of amplification of mRNA (described in, for example, Lockhart et al, *Nature Biotechnology* (1996), 14, 1675-1680);

IDT Exhibit 1020.9278

van Gelder et al., U.S. Patent No. 5,716,785), are limited to the amplification efficiency of DNA-dependent RNA polymerases and some of these methods require multiple steps. Moreover, the process by which the polymerase promoter sequence is incorporated is prone to result in non-specific amplification.

The methods of the invention offer the ability to efficiently amplify mRNA under conditions that provide for high specificity of target amplification and which is generally reflective of the distribution in the input RNA. Thus, the utility of the detection/quantification methods which can be used with the amplification products of the invention, such as those based on the powerful array technology, real time PCR, quantitative TaqMan, quantitative PCR using molecular beacons and the like, should be greatly enhanced.

The linear aspect of the amplification methods of the invention significantly increases the specificity of target amplification. Since generation of multiple copies of a sequence of interest is not dependent nor based on cyclical, exponential, amplification of amplification products, the specificity of products obtained is greatly enhanced. The distribution of the various species of amplified products is generally reflective of the distribution in the input RNA.

The methods of the invention do not require thermocycling and all of the steps can be performed isothermally, although the various steps may be carried out a different temperatures. This feature provides numerous advantages, including facilitating automation and adaptation for high through-put procedures. The isothermal reaction is faster than that afforded by thermal cycling and is suitable for performing the methods of the invention in miniaturized devices.

The intermediate double stranded cDNA complex comprising an RNA/DNA heteroduplex provides a substrate for linear amplification. Cleavage of the RNA portion of the RNA/DNA heteroduplex permits further amplification without the need to denature the double stranded cDNA intermediate complex. Moreover, since the cleaved double stranded cDNA complex is mostly double stranded, it is less likely that the secondary structure of a single stranded template will interfere with subsequence amplification.

Finally, most of the methods of the invention produce products that are single stranded, thus rendering them more accessible to binding to probes, either in a homogeneous manner, i.e. in solution, or binding to probes immobilized on solid supports.

IDT Exhibit 1020.9279

The double stranded products of the methods of the invention are useful for, e.g., production of cDNA libraries.

The ability to efficiently amplify mRNA (or any other desired RNA species or population) under conditions that provides for high specificity of target amplification and which will generally not alter the representation of the various mRNA species in the preparation, will greatly enhance the utility of the detection/quantification methods such as those based on the powerful array technology.

General Techniques

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

Primers, oligonucleotides and polynucleotides employed in the invention can be generated using standard techniques known in the art.

Definitions

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

"Amplification," or "amplifying", as used herein, generally refers to the process of producing multiple copies of a desired sequence. "Multiple copies" mean at least 2 copies. A "copy" does not necessarily mean perfect sequence complementarity or identity to the template sequence. For example, copies can include nucleotide analogs such as

IDT Exhibit 1020.9280

deoxyinosine, intentional sequence alterations (such as sequence alterations introduced through a primer comprising a sequence that is hybridizable, but not complementary, to the template, or a non-target sequence introduced through a primer), and/or sequence errors that occur during amplification. "Amplifying" a sequence may generally refer to making copies of a sequence or its complement, with the understanding that, as stated above, copying does not necessarily mean perfect sequence complementarity or identity with respect to the template sequence.

"Polynucleotide," or "nucleic acid," as used interchangeably herein, refer to polymers of nucleotides of any length, and include DNA and RNA. The nucleotides can be deoxyribonucleotides, ribonucleotides, modified nucleotides or bases, and/or their analogs, or any substrate that can be incorporated into a polymer by DNA or RNA polymerase. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and their analogs. If present, modification to the nucleotide structure may be imparted before or after assembly of the polymer. The sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component. Other types of modifications include, for example, "caps", substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as, for example, those with uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoamidates, cabamates, etc.) and with charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), those containing pendant moieties, such as, for example, proteins (e.g., nucleases, toxins, antibodies, signal peptides, ply-L-lysine, etc.), those with intercalators (e.g., acridine, psoralen, etc.), those containing chelators (e.g., metals, radioactive metals, boron, oxidative metals, etc.), those containing alkylators, those with modified linkages (e.g., alpha anomeric nucleic acids, etc.), as well as unmodified forms of the polynucleotide(s). Further, any of the hydroxyl groups ordinarily present in the sugars may be replaced, for example, by phosphonate groups, phosphate groups, protected by standard protecting groups, or activated to prepare additional linkages to additional nucleotides, or may be conjugated to solid supports. The 5' and 3' terminal OH can be phosphorylated or substituted with amines or organic capping groups moieties of from 1 to 20 carbon atoms. Other hydroxyls may also be derivatized to standard protecting groups. Polynucleotides can also contain analogous forms of ribose or deoxyribose sugars that are generally known in the art, including, for example, 2'--O-methyl-, 2'-O-allyl, 2'-fluoro- or 2'-azido-ribose,

IDT Exhibit 1020.9281

carbocyclic sugar analogs, α-anomeric sugars, epimeric sugars such as arabinose, xyloses or lyxoses, pyranose sugars, furanose sugars, sedoheptuloses, acyclic analogs and abasic nucleoside analogs such as methyl riboside. One or more phosphodiester linkages may be replaced by alternative linking groups. These alternative linking groups include, but are not limited to, embodiments wherein phosphate is replaced by P(O)S("thioate"), P(S)S ("dithioate"), "(O)NR_2 ("amidate"), P(O)R, P(O)OR', CO or CH_2 ("formacetal"), in which each R or R' is independently H or substituted or unsubstituted alkyl (1-20 C) optionally containing an ether (-O-) linkage, aryl, alkenyl, cycloalkyl, cycloalkenyl or araldyl. Not all linkages in a polynucleotide need be identical. The preceding description applies to all polynucleotides referred to herein, including RNA and DNA.

A "labeled dNTP," or "labeled rNTP," as used herein, refers, respectively, to a dNTP or rNTP, or analogs thereof, that is directly or indirectly attached with a label. For example, a "labeled" dNTP or rNTP, may be directly labeled with, for example, a dye and/or a detectable moiety, such as a member of a specific binding pair (such as biotin-avidin). A "labeled" dNTP or rNTP, may also be indirectly labeled by its attachment to, for example, a moiety to which a label is/can be attached. A dNTP or rNTP, may comprise a moiety (for example, an amine group) to which a label may be attached following incorporation of the dNTP or rNTP into an extension product. Useful labels in the present invention include fluorescent dyes (e.g., fluorescein isothiocyanate, Texas red, rhodamine, green fluorescent protein and the like), radioisotopes (e.g., ^{3}H, ^{35}S, ^{32}P, ^{33}P, ^{125}I, or ^{14}C), enzymes (e.g., LacZ, horseradish peroxidase, alkaline phosphatase,), digoxigenin, and 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 dNTP or rNTP, as used herein, refers to the particular base of a nucleotide, namely adenine, cytosine, thymine, 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 invention include the composite primer and propromoter polynucleotide (such as the PTO). The terms

IDT Exhibit 1020.9282

"oligonucleotide" and "polynucleotide" are not mutually exclusive. The description above for polynucleotides is equally and fully applicable to oligonucleotides.

A "primer," as used herein, refers to a nucleotide sequence, generally with a free 3'-OH group, that hybridizes with a template sequence (such as a target RNA, or a primer extension product) and is capable of promoting polymerization of a polynucleotide complementary to the template. A "primer" can be, for example, an oligonucleotide. It can also be, for example, a sequence of the template (such as a primer extension product or a fragment of the template created following RNase cleavage of a template-DNA complex) that is hybridized to a sequence in the template itself (for example, as a hairpin loop), and that is capable of promoting nucleotide polymerization. Thus, a primer can be an exogenous (e.g., added) primer or an endogenous (e.g., template fragment) primer.

A "random primer," as used herein, is a primer that comprises a sequence that is designed not necessarily based on a particular or specific sequence in a sample, but rather is based on a statistical expectation (or an empirical observation) that the sequence of the random primer is hybridizable (under a given set of conditions) to one or more sequences in the sample. The sequence of a random primer (or its complement) may or may not be naturally-occurring, or may or may not be present in a pool of sequences in a sample of interest. The amplification of a plurality of RNA species in a single reaction mixture would generally, but not necessarily, employ a multiplicity, preferably a large multiplicity, of random primers. As is well understood in the art, a "random primer" can also refer to a primer that is a member of a population of primers (a plurality of random primers) which collectively are designed to hybridize to a desired and/or a significant number of target sequences. A random primer may hybridize at a plurality of sites on a nucleic acid sequence. The use of random primers provides a method for generating primer extension products complementary to a target polynucleotide which does not require prior knowledge of the exact sequence of the target.

"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.

IDT Exhibit 1020.9283

A "propromoter polynucleotide," as used herein, refers to a polynucleotide comprising a propromoter sequence. Example of a propromoter polynucleotide is a propromoter template oligonucleotide (PTO).

"Propromoter template oligonucleotide (PTO)"and "promoter template oligonucleotide (PTO)" as used herein, refer to an oligonucleotide that comprises a propromoter sequence and a portion, generally a 3' portion, that is hybridizable (under a given set of conditions) to the 3' region of a primer extension product. The propromoter sequence and the hybridizable portion may be the same, distinct or overlapping nucleotides of an oligonucleotide.

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). An example of a complex is a nucleic acid duplex comprising a first primer extension product and a second primer extension product.

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 Figures 1A and 2A.

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. "A" fragment means one or more fragments.

"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

IDT Exhibit 1020.9284

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 sequence, due to mechanisms such as substitution, transversion, 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 refer 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, result 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, polystyrene), polyacrylamide, nitrocellulose, silicon, optical fiber or any other suitable solid or semi-solid support, and configured in a planar (e.g., glass plates, silicon chips) or three-dimensional (e.g., pins, fibers, beads, particles, microtiter wells, capillaries) configuration. Probes forming the arrays may be attached to the substrate by any number of ways including (i) *in situ* synthesis (e.g., high-density oligonucleotide arrays) using photolithographic techniques (see, Fodor et al., *Science* (1991), 251:767-773; Pease et al., *Proc. Natl. Acad. Sci. U.S.A.* (1994), 91:5022-5026; Lockhart et al., *Nature Biotechnology*

IDT Exhibit 1020.9285

(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 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 50, more preferably from about 10 to about 40, even more preferably from about 20 to about 30 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 50, more preferably from about 10 to about 40, even more preferably from about 20 to about 30 nucleotides.

IDT Exhibit 1020.9286

As is well understood by those skilled in the art, a "tail" sequence of a primer is a sequence not hybridizable to the target sequence under conditions in which other region(s) or portion(s) of the primer hybridizes to the target.

"Absent" or "absence" of product, and "lack of detection of product" as used herein includes insignificant, or *de minimus* levels, generally due to lack of significant accumulation of product due to cycling.

Amplification methods of the invention

The following are examples of the amplification methods of the invention. It is understood that various other embodiments may be practiced, given the general description provided above. For example, reference to using a composite primer means that any of the composite primers described herein may be used.

In one aspect of the invention, a method for generating multiple copies (amplifying) of a polynucleotide (DNA) sequence complementary to an RNA sequence of interest using a composite primer and a second primer is provided. In this method, isothermal linear nucleic acid sequence amplification is achieved using two primers (a composite primer and a second primer) and strand displacement. In some embodiments, a transcription step is included (i.e., an "enhanced" method), and amplified product in the form of sense RNA is produced.

In another aspect of the invention, a method for generating multiple copies (amplifying) of a DNA sequence complementary to an RNA sequence of interest using a single primer (which is a composite primer) is provided. In this method, isothermal linear nucleic acid amplification is achieved using a single composite primer, a fragment of target RNA (which serves as an endogenous primer), and strand displacement. In some embodiments, a transcription step is included (i.e., the "enhanced" method), and amplified product in the form of sense RNA is produced.

As described herein, the amplification methods of the invention may further include a transcription step. If a primer extension product that is to be transcribed comprises a propromoter sequence, a double stranded promoter region is generally generated by hybridizing a polynucleotide comprising a propromoter (also referred to herein as "propromoter polynucleotide") to the primer extension product. If a primer extension product does not comprise a desired propromoter sequence, the transcription step is

IDT Exhibit 1020.9287

generally dependent on the incorporation of an RNA polymerase propromoter sequence, by use of a propromoter polynucleotide such as a promoter sequence oligonucleotide (PTO). A propromoter polynucleotide such as the PTO can serve as a template for extension of a single stranded primer extension product and formation of a partial duplex comprising a double stranded promoter at one end. The ability to hybridize the single stranded product to the propromoter polynucleotide (such as a PTO) is generally achieved by creating a primer extension product with a defined 3' end sequence, which is complementary to the 3' end sequence of the propromoter polynucleotide.

One aspect of the methods of the invention includes the design of primers which are able to hybridize to RNA sequences, such as a plurality of RNA sequences, for initiation of primer extension and production of amplification substrates and products.

Linear nucleic acid sequence amplification using a composite primer and a second primer

The invention provides methods of amplifying an RNA sequence of interest by using a composite primer and a second primer, and strand displacement. Amplification by these methods is linear and can be achieved isothermally. In embodiments that do not include a transcription step, amplified products are single stranded DNA comprising sequences complementary to the RNA sequence of interest in the target RNA. In embodiments that include transcription, amplified products are sense RNA copies of the RNA sequence of interest in the target RNA.

A schematic description of one embodiment of the composite primer, second primer and strand displacement-based methods of the invention is given in Figures 1A-B and 2A-C. Figures 1A-B illustrate a non-enhanced linear method. Figures 2A-C illustrate an enhanced (i.e., including a transcription step) method. The methods involve the following steps: (a) formation of a second strand cDNA which is the same sense as the input RNA; (b) linear amplification of the complement of a second strand cDNA strand by primer extension (from a composite primer) along the second strand cDNA and strand displacement; and, in the enhanced method, (c) transcription of the product of the linear amplification step to produce multiple copies of sense RNA products. As illustrated, all steps are isothermal, although the temperatures for each of the steps may or may not be the same.

IDT Exhibit 1020.9288

The embodiment illustrated in Figures 1A-B employs two oligonucleotides: a composite primer, (labeled 1), used for the amplification; and a second primer (labeled 2), used for the formation of the sense complement DNA (cDNA) (interchangeably called "second primer extension product" or "second strand cDNA). The embodiment illustrated in Figures 2A-C employs three oligonucleotides: a composite primer, (labeled 1) , used for the amplification; a second primer (labeled 2), used for the formation of the second strand (sense) cDNA ; and a polynucleotide comprising a propromoter sequence of a DNA-dependent RNA polymerase, i.e., a promoter template oligonucleotide (PTO) (labeled 3).

The 3' portion of the composite primer can be designed in any of a number of ways (in terms of sequence), depending on which type, class, population, and/or species of RNA is desired to be amplified. In some embodiments, the 3' portion of composite primer 1, as illustrated in Figures 1 and 2, comprises a sequence complementary to the poly-A tail of mRNA, and may further comprise additional random sequences (generally not complementary to a poly-A sequence) at the 3' end of the 3' portion. In other embodiments, the 3' portion of composite primer 1 is a random primer comprising sequences which are hybridizable to a multiplicity of RNA species (which may range from 2 or more to many hundred or thousands or more). Random primers are known in the art, for example, they have been used extensively in the preparation of cDNA libraries using PCR-based procedures. As is well understood in the art, a "random primer" can 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.

In other embodiments, the 3' portion of composite primer 1 can comprise a sequence complementary or hybridizable to a specific RNA or family of RNAs (or portions thereof).

In some embodiments as illustrated in Figures 1 and 2, the 5' portion of composite primer 1 can be a sequence not hybridizable to the target sequence (under conditions in which the 3' portion hybridizes to RNA target), e.g., a sequence forming a "tail" when the primer is hybridized to a target. This "tail" sequence generally is incorporated into the first primer extension product (first strand cDNA), and the complement of this tail will be incorporated at the 3' end of the second primer extension product (second strand cDNA).

IDT Exhibit 1020.9289

Accordingly, in some embodiments, composite primer 1 is a mixtures of composite primers which comprise the same 5' RNA portion and a multiplicity of 3' DNA portions selected to amplify a multiplicity (which can be small to very large) of RNA sequences of interest. In other embodiments, the 5' portion of the composite primer 1 can be hybridizable to the target RNA. Although Figure 1 and 2 depict a 3' portion hybridized to the target sequence that is a DNA portion and a 5' portion non hybridized to the target that is an RNA portion, it is understood that DNA portions can comprise part of a "tail" and conversely that RNA portions can be hybridizable to target RNA. For example, the 5' RNA portion of a composite primer may be partially hybridizable and partially nonhybridizable, or the DNA portion of a composite primer may be partially hybridizable and partially nonhybridizable, or both.

As illustrated in these figures, the composite primer 1 comprises a DNA portion, A, at its 3' end, and an RNA portion, B, at its 5' end. As discussed herein, it is also possible to employ a composite primer in which the 3' DNA portion is followed, in the direction of its 5', by an RNA portion, which is followed by a portion which is DNA. The length of each of these sections is generally determined for maximum efficiency of the amplification. Only the two-portion (i.e., 3'-DNA-RNA-5') composite primer is shown in Figures 1A-B and Figures 2A-C.

Primer 2 as illustrated in Figures 1A-B and 2A-C can be, but is not necessarily, composed of DNA and can comprise two sections (interchangeably called "portions" or "regions"). The 3' portion, F, of primer 2 can be selected for random priming of many, most and/or all possible mRNA sequences in a biological sample. Random primers are known in the art, for example, they have been used extensively in the preparation of cDNA libraries using PCR-based procedures. The 5' portion, E, of primer 2 can be a sequence which is not complementary and not substantially hybridizable to a specific target sequence, i.e., it would not hybridize (under conditions in which the 3' portion hybridizes to RNA target) and would constitute a tail. The "tail" sequence would generally be incorporated into the second primer extension product, and the 3' end sequence of the DNA product of the linear amplification steps would generally comprise the complement of this sequence. In other embodiments, if enhanced amplification is not desired (as described below), the 5' end portion, E, of primer 2 can be hybridizable to the target sequence in the first primer extension product.

IDT Exhibit 1020.9290

As illustrated in Figures 2A-C, a promoter template oligonucleotide, 3 (PTO), can be designed as follows: the 3' portion is the same as portion E of primer 2. This design enables the PTO to hybridize to the 3' end of the linear amplification product. The 5'-most portion of the PTO is a promoter sequence for a DNA-dependent RNA polymerase, which, as described above, is used in certain (namely, the "enhanced" method) embodiments of the amplification methods of the invention. Generally, the sequence between these two sections is designed for optimal transcription by the polymerase of choice. Criteria for selection and design of this sequence are known in the art.

For convenience, only one first primer extension product (first strand cDNA) and second strand primer extension product (second strand cDNA) are described and illustrated in Figures 1 and 2. It is understood that the methods of the invention are useful not only for producing large amounts of one specific nucleic acid sequence, but also for amplifying simultaneously more than one different specific nucleic acid sequence located on the same or different target nucleic acid molecules. For example, the methods of the invention are useful for amplifying all mRNA in a sample, or for amplifying a multiplicity of specific RNA species or family of RNA species in a sample.

As illustrated in Figures 1A-B, in one embodiment, the process of the amplification methods of the invention resulting in generation of DNA products comprising sequences complementary to an RNA sequence(s) of interest based on RNA is as follows:

A) Formation of a double stranded cDNA substrate for linear amplification

1. Primer 1 binds to an RNA species in a sample by hybridization of the random sequence portion A (which can be based at least in part on the poly-A sequence of the mRNA), to form complex I (Figure 1A).
2. A reverse transcriptase, (indicated as "RT"), extends the hybridized primer 1 along the target RNA strand to which it is hybridized, to form an RNA/DNA duplex. RNase H degrades the target RNA strand of the hybrid duplex to generate a single stranded first strand cDNA (labeled "II"). The 5' end of II is primer 1.
3. Primer 2 binds to first strand cDNA, II, by hybridization of sequence F, to form complex III.
4. Primer 2 is extended along the cDNA strand II by a DNA polymerase to form a double stranded product (labeled "IV"). Primer extension along the 5' RNA

IDT Exhibit 1020.9291

portion of II by an RNA-dependent DNA polymerase such as a reverse transcriptase results in formation of an RNA/DNA hybrid portion at one end of complex IV.

5. RNase H degrades the RNA portion of the RNA/DNA hybrid at one end of complex IV, to create a partial double stranded complex (labeled "V") with a 3' DNA single stranded end, which has a sequence which is the complement of portion B of the composite primer 1. The RNase H activity may be supplied by the RNA-dependent DNA polymerase (such as reverse transcriptase) or may be provided in a separate enzyme. Reverse transcriptases useful for this method may or may not have RNase H activity.

B) Isothermal linear amplification

1. Primer 1 binds to complex V by hybridization of the RNA portion, B, to the single stranded DNA end, which is complementary to it, to form complex VI. The 3' DNA sequence of primer 1 is not hybridized.
2. The 3' end of bound primer 1 in complex VI, and the 5' end of the DNA strand immediately upstream are the same, and would compete for hybridization to the opposite strand. Without wishing to be bound by theory, the high affinity of the DNA polymerase to hybridized 3' end of a primer would be expected to push the equilibrium of the two competing structures towards hybridization of the 3' end of the new primer and displacement of the 5' end of the previous primer extension product (labeled "VII"). Primer extension along the second strand (sense) cDNA strand results in displacement of the previous second primer extension product (VII), and replicates the E sequence of primer 2, to form complex VIII.
3. Complex VIII has, at one end, an RNA/DNA hybrid composed of sequence B and its complement. The RNA segment of the hybrid is degraded by RNase H, to form complex IX, which results in the formation of a single stranded 3' end to which a new primer 1 can be bound by its 5' B portion.
4. The process of hybridization of the 3' end sequence A of the bound primer 1, by displacement of the 5'-most end of the previous primer extension product in the duplex, primer extension and displacement of the previous product continues as shown in Figures 1A-B, and results in the accumulation of multiple copies of anti-sense single stranded DNA products (labeled "XII"). These products have at their

IDT Exhibit 1020.9292

3' ends a sequence complementary to portion E of primer 2, and at their 5' ends a sequence substantially identical (generally, identical) to sequence A of the composite primer. Kurn, U.S. Patent No. 6,251,639 B1.

One embodiment of the enhanced method is illustrated in Figures 2A-C. In this embodiment, subsequent to the generation of DNA product comprising complementary sequences of an RNA sequence of interest, the following steps are performed.

C) Transcription of the DNA products

1. A PTO (see Figure 2C), binds to DNA product XII by hybridization of its 3' end sequence to the 3' end sequence of DNA product XII to form complex XIII. The 3' end of the PTO is preferably, but not necessarily, blocked so that it cannot be extended by a DNA polymerase.
2. DNA polymerase extends the 3' end of product XII in complex XIII, along the PTO template to form complex XIV comprising a double stranded promoter sequence at one end.
3. DNA-dependent RNA polymerase binds to the double stranded promoter in complex XIV to transcribe the anti-sense single stranded DNA product to yield multiple copies of sense RNA products (labeled "XV"). Multiple species of product XV would represent multiple copies of multiple sequences from a pool of input RNA. Kurn, U.S. Patent No. 6,251,639 B1.

The methods of the invention may be used for generation of multiple copies of a plurality of RNA sequences in the sample. For example, the composite primer can comprise a poly-dT sequence, which would be expected to hybridize to the poly-A tails of all mRNA is a sample, or the composite primer can generally comprise at least a 3' portion that is hybridizable to random or partially random sequences (e.g. comprising a poly-dT sequence and a random sequence, which would be expected to hybridize to the beginning of the poly-A tails of mRNAs). In another aspect, the methods of the invention may be used for generation of multiple copies of a specific RNA species or class of RNA species (e.g., a family or superfamily of RNA species). In this latter case, the composite primer

IDT Exhibit 1020.9293

generally comprises a 3'-portion which is complementary to a sequence of a specific RNA target (or family of RNA targets).

The 5'-RNA portion of the composite primer may or may not be related to the specific RNA target sequence, and may or may not hybridize to the RNA target, for example, it may form a tail as further described herein.

Although only one composite primer is described in the embodiments above, it is understood that a different composite primer may be used in step (b), above. The different composite primer comprises sequences hybridizable to the single stranded DNA portion of the complex IX, described above. The second composite primer may further comprise sequences hybridizable to a portion of the second primer extension product sequences immediately 5' to the single stranded DNA portion of complex IX. The second composite primer generally comprises overlapping sequences with the first composite primer. The second composite primer is hybridized to the second primer extension product and extended by primer extension. Cleavage of the DNA-RNA heteroduplex by RNAse permits binding of another second composite primer , extension and strand displacement, whereby multiple copies of the single stranded product are produced.

For convenience, only one first primer extension product (first strand cDNA) and second strand primer extension product (second strand cDNA) are described and illustrated in Figures 1 and 2. It is understood that the methods of the invention are useful not only for producing large amounts of one specific nucleic acid sequence, but also for amplifying simultaneously more than one different specific nucleic acid sequence located on the same or different target nucleic acid molecules. For example, the methods of the invention are useful for amplifying all mRNA in a sample, or for amplifying a multiplicity of specific RNA species (in which case a multiplicity of first composite primer each comprising a 3' portion hybridizable to specific sequences of specific RNA species could be used) or family of RNA species in a sample.

Linear nucleic acid sequence amplification using a single composite primer and a target RNA fragment

The invention also provides methods of amplifying an RNA sequence of interest by using a single primer (which is a composite primer), a target RNA fragment, and strand displacement. Amplification by these methods is linear and can be achieved isothermally.

IDT Exhibit 1020.9294

In embodiments that do not include a transcription step, amplified products are DNA comprising sequences complementary to the RNA sequence of interest in the target RNA. In embodiments that include transcription, amplified products are sense RNA copies of the RNA sequence of interest in the target RNA.

A schematic description of a non-enhanced (linear) embodiment of the single composite primer and strand displacement-based methods of the invention is provided in Figures 3A-B. The method involves the following steps: (a) primer extension to form an RNA/DNA heteroduplex of a target RNA and a first strand cDNA; (b) incomplete degradation of the RNA target of the heteroduplex to form a complex of the first strand cDNA and at least one RNA fragment that can in turn serve as a primer; (c) formation of a second strand cDNA that is of the same sense as the input target RNA; (d) linear amplification of the complement of a second strand DNA by primer extension (from a composite primer) along the second strand cDNA and strand displacement to produce multiple copies of antisense single stranded DNA products. As illustrated, all steps are isothermal, although the temperatures for each of the steps may or may not be the same. A schematic description of an embodiment of enhanced single composite primer and strand displacement-based methods of the invention is provided in Figures 4A-B. The method involves the following steps: (a) primer extension to form an RNA/DNA heteroduplex of a target RNA and a first strand cDNA; (b) incomplete degradation of the RNA target of the heteroduplex to form a complex of the first strand cDNA and at least one RNA fragment that can in turn serve as a primer; (c) formation of a second strand cDNA that is of the same sense as the input target RNA; (d) linear amplification of the complement of a second strand DNA by primer extension (from a composite primer) along the second strand cDNA and strand displacement to produce multiple copies of antisense single stranded DNA products; and (e) transcription of the product of the linear amplification step to produce multiple copies of sense RNA products.

A composite primer, as described herein, is used for the amplification in these methods. As illustrated in Figures 3A and 4A, the single (composite) primer (labeled "1") can be composed of a 3'-DNA portion (labeled "A") which is complementary to a sequence on the target RNA, and a 5'-RNA portion (labeled "B") which comprises a non-target related sequence (i.e., it is not complementary/hybridizable (under a given set of conditions) to a sequence on the target RNA). The 3'-DNA portion of the composite primer may comprise poly-dT nucleotides, which would render it

IDT Exhibit 1020.9295

complementary/hybridizable (under a given set of conditions) to the poly-A 3' end of mRNA derived from a eukaryotic cell.

A composite primer that is hybridized to a target RNA is extended by an RNA-dependent DNA polymerase, such as a reverse transcriptase, to form an RNA/DNA heteroduplex of the target RNA and a first strand cDNA. Degradation of the target RNA of the heteroduplex is then achieved using a ribonuclease such as RNase H, to form a complex of the first strand cDNA and one or more RNA fragments (oligonucleotides). The RNase H activity may be supplied by the RNA-dependent DNA polymerase (such as reverse transcriptase) or may be provided in a separate enzyme. Reverse transcriptases useful for this method may or may not have RNase H activity. The fragments are a result of incomplete degradation of the target RNA in the heteroduplex. These fragments function as primers for a DNA-dependent DNA polymerase to form the second strand cDNA. Okayama & Berg, *Molecular and Cell Biology* (1982), 2:161; and Gubler & Hoffman, *Gene* (1983), 25:263. Reverse transcriptase then extends the 3'-end of the second strand cDNA in the duplex along the 5'-RNA sequence of the composite primer extension product (the first strand cDNA), to form an RNA/DNA heteroduplex at the end of the double stranded cDNA product.

The heteroduplex at the end of the double stranded cDNA is a substrate for RNase H. The RNase H degrades the 5'-RNA portion of the first strand cDNA, to create a site for hybridization of the composite primer, which hybridizes by its 5'-RNA portion to the 3'-end of the second strand cDNA in the duplex cDNA. The 3' DNA portion of the new primer displaces the 5' end of the first strand cDNA, by hybridization to its complementary sequence on the second strand cDNA. A DNA polymerase with strong strand displacement activity then extends the new primer along the second strand cDNA, and displaces the previously formed cDNA product. An RNase H degrades the 5'-portion of the new primer extension product in the heteroduplex, to create a free site for hybridization of a new composite primer, thus resulting in continuous linear amplification of the target sequence and generation of multiple copies of single stranded DNA product which is anti sense to the target RNA sequence.

As depicted in Figures 4A-B, and as is evident from the description herein, the single composite primer amplification methods can also include a transcription step employing a propromoter polynucleotide in the same manner as that described with respect to the enhanced methods of amplification using a composite primer and a second primer.

IDT Exhibit 1020.9296

The methods of the invention may be used for generation of multiple copies of a plurality of RNA sequences in the sample or of a specific RNA species or group of species. In the latter cases, the composite primer generally comprises a 3'-portion which is complementary and/or hybridizable to a sequence of a specific RNA target or group of targets (e.g., homologous RNA targets or targets that are members of a family or superfamily of sequences). For example, the primer may comprise a collection of primer sequences, such as where more than one target sequence exists.

Another application of the methods of the invention is in detection of variant regions flanking a common sequence. By designing a first composite primer that recognizes a commonly shared sequence, single stranded DNA or RNA product is produced that contains not only the common region recognized by the primer, but also 5'-flanking sequence useful in detecting sequence variants. Thus, for example, single stranded DNA or RNA product can be produced from limited amounts of clinical material to allow pathogen-specific sequences (such as those distinguishing viral types) to be identified, genetic polymorphisms to be detected, or alternate splicing variants to be characterized, all in accordance with standard techniques. In other embodiments, single stranded DNA or RNA product is produced that contains not only the common region recognized by the primer, e.g., a conserved or functional sequence motif, but also 5'-flanking sequences permitting identification of groups of RNA species comprising similar sequence motifs.

The 5'-RNA portion of the composite primer may or may not be related to the specific RNA target sequence, and may or may not hybridize to the RNA target. Methods of the invention using "tailed" composite primers are described further herein.

Linear mRNA amplification using a single composite primer and a target RNA fragment

As is described herein, total mRNA may be amplified using a composite primer comprising a poly(T) of sufficient length to hybridize with substantially an entire population of messages (i.e., poly(T)n, wherein n is typically from about 5 to 50 or more). Figure 5 exemplifies a schematic description of a non-enhanced embodiment of the strand displacement-based methods of the invention comprising use of a single composite primer comprising a 3'-DNA portion comprising a poly-dT sequence, and further comprising a random sequence 3' the poly-dT sequence. Composite primer 1 further comprises a 5'-

IDT Exhibit 1020.9297

RNA portion that is not substantially hybridizable to the RNA target sequences (i.e. a tail under conditions in which the poly-dT sequence hybridizes). Optionally, a second composite primer may be used as described below.

The method involves the following steps: (a) primer extension to form an RNA/DNA heteroduplex of a target RNA and a first strand cDNA; (b) formation of a second primer extension product (second strand cDNA) that is of the same sense as the input target RNA. The second primer extension product generally comprises a 3'-portion that is complementary to the 5'-RNA portion of the composite primer (i.e., the tail); (c) linear amplification of the complement of the second strand cDNA by primer extension (from a composite primer) and strand displacement to produce multiple copies of antisense single stranded DNA products (that are complementary to the RNA sequence of interest).

Figure 6 exemplifies a schematic description of an non-enhanced embodiment of the strand displacement-based methods of the invention comprising use of a single composite primer comprising a 3'-DNA portion comprising a random sequence and further comprising a 5'-RNA portion that is not substantially hybridizable to the RNA target sequences (i.e. a tail under conditions in which the random sequence hybridizes). Optionally, a second composite primer may be used as described below.

The method involves the following steps: (a) primer extension to form RNA/DNA heteroduplexes of a target RNA and a first strand cDNA; (b) formation of second primer extension products (second strand cDNA) that is of the same sense as the input target RNA. The second primer extension product generally comprises a 3'-portion that is complementary to the 5'-RNA portion of the composite primer (i.e., the tail); (c) linear amplification of the sense DNA strand by primer extension from a composite primer and strand displacement to produce multiple copies of antisense single stranded DNA products that are complementary to the RNA sequence of interest.

As illustrated in these embodiments, all steps are isothermal, although the temperatures for each of the steps may or may not be the same. It is understood that various other embodiments may be practiced, given the general description provided above. It is further understood that the formation of a second primer extension product may be accomplished by any method known in the art or described herein (e.g., extension of a hybridized second primer or RNA fragment).

As is evident from the description herein, the amplification methods can also include a transcription step employing a propromoter polynucleotide in the same manner as

IDT Exhibit 1020.9298

that described with respect to the enhanced methods of amplification using a composite primer and a second primer. The single stranded RNA products of the enhanced methods of amplification would generally comprise a 3'-region that is complementary to the 5'-RNA portion of the first composite primer.

Figure 7 exemplifies a schematic description of an enhanced embodiment of the strand displacement-based methods of the invention comprising use of (a) a single composite primer comprising a 3'-DNA portion comprising a random sequence and further comprising a 5'-RNA portion that is not substantially hybridizable to the RNA target sequences (i.e. a tail under conditions in which the random sequence hybridizes); and (b) a propromoter oligonucleotide). Optionally, a second composite primer may be used as described herein.

The method involves the following steps: (a) primer extension to form RNA/DNA heteroduplexes of a target RNA and a first strand cDNA; (b) formation of second primer extension products (second strand cDNA) that is of the same sense as the input target RNA and cleavage of RNA present in a RNA/DNA heteroduplex by an agent (such as RNase H) capable of cleaving RNA present in an RNA/DNA heteroduplex, whereby a double stranded complex of first and second primer extension product comprising a 3' single stranded portion is generated (complex IX in Figure 8, which correspond to complex IX as illustrated in Figure 1); and (c) a PTO binds to double stranded product IX to form complex X, as shown in Figure 8. DNA polymerase extends the 3' end of the second primer extension product in complex XIII, along the PTO template to form complex XI comprising a double stranded promoter sequence at one end; and (d) transcription using DNA-dependent RNA polymerase to produce multiple copies of sense RNA products. As illustrated, all steps are isothermal, although the temperatures for each of the steps may or may not be the same.

As illustrated in Figure 8, a promoter template oligonucleotide, 3 (PTO), can be designed as follows: the 3' portion is the same as the 5' RNA portion of the first composite primer (that hybridizes to template RNA). This design enables the PTO to hybridize to the single stranded 3' portion of the second primer extension product (which is present in a complex with the first primer extension product). The 5'-most portion of the PTO is a promoter sequence for a DNA-dependent RNA polymerase, which, as described above, is used in certain (namely, the other "enhanced" methods) embodiments of the amplification methods of the invention.

IDT Exhibit 1020.9299

Linear mRNA amplification using a first composite primer, a second composite primer, and a target RNA

The invention provides methods of producing single stranded antisense and sense polynucleotide, generally DNA, copies of an RNA sequence of interest using a first composite primer, a second composite primer which is used to generate the second strand cDNA, and a target RNA. In this aspect of the invention, a first composite primer is used to generate first extension product (generally cDNA), which is a substrate for the linear amplification using a composite primer, as described above. In addition, a second composite primer is used to generate a second strand cDNA which is a substrate for linear amplification, resulting in production of single stranded polynucleotide copies of the RNA sequence of interest.

The method involves the following: (a) formation of a double stranded cDNA comprising a RNA-DNA heteroduplex at each end of the cDNA; and (b) linear amplification of first strand (sense) cDNA and second strand (antisense) cDNA by primer extension from the first composite primer and from the second composite primer (which binds to the first strand cDNA), and strand displacement. Single stranded first and second strand cDNA product is produced, which is useful for, for example, producing cDNA libraries. As is evident, in this aspect of the invention, the second primer extension product is primed by a composite primer.

Figure 8 illustrates one embodiment of the invention. Two composite primers comprising different "tail" sequences are used to generate a double stranded cDNA comprising a RNA-DNA heteroduplex at each end of the cDNA. Cleavage of RNA by an agent that cleaves RNA from an RNA/DNA heteroduplex permits binding of another first composite primer, another second composite primer (that hybridizes to the first primer extension product), extension and strand displacement, whereby multiple copies of an antisense single stranded product and multiple copies of a sense single stranded DNA product are produced. Combination of sense and antisense single stranded cDNA product is capable of producing double stranded cDNA. The process of the amplification methods of the invention resulting in generation of single stranded cDNA products comprising sequences complementary to an RNA sequence(s) of interest and sequences comprising an RNA sequence(s) of interest is as follows (an embodiment of which is illustrated in Figure 7):

IDT Exhibit 1020.9300

A) Formation of a double stranded cDNA substrate for linear amplification

1. Composite primer 1 binds to an RNA target in a sample by hybridization of the primer portion A (which can be based at least in part on the poly-A sequence of the mRNA), to form complex I.
2. A reverse transcriptase extends the hybridized primer 1 along the target RNA strand to which it is hybridized, to form an RNA/DNA duplex, labeled II. An agent (such as RNase H) degrades the target RNA strand of the hybrid duplex to generate a single stranded first strand cDNA (labeled "III"). The 5' end of III is primer 1.
3. Composite primer 2, binds to the first strand cDNA, III, by hybridization of sequence F, to form complex IV.
4. Composite primer 2 is extended along the cDNA strand III by a DNA polymerase to form a double stranded product (labeled "V") which consists of first and second strand cDNA. Primer extension along the 5' RNA portion of IV by an RNA-dependent DNA polymerase such as a reverse transcriptase results in formation of an RNA/DNA hybrid portion at one end of complex V.
5. An agent (such as RNase H) degrades the RNA portion of the RNA/DNA hybrid at one end of complex V, to create a partial double stranded complex (labeled "VI") with a 3' DNA single stranded end, which has a sequence which is the complement of portion B of the composite primer 1.
6. Composite primer 1 binds to complex VI by hybridization of the RNA portion to the single stranded DNA end, which is complementary to it, to form complex VII.
7. Primer extension of bound primer 1 in complex VII along the sense cDNA strand results in displacement of the previous primer extension product (VII), and replicates portion "G" of the second composite primer, to form complex VIII.
8. An agent (such as RNase H) cleaves the RNA portions of the RNA/DNA heteroduplexes, forming complex VIII. Complex VIII has two RNA/DNA heteroduplexes comprised of composite primer 1 and the complement of composite primer 1 at one end, and the second composite primer and the complement of the second composite primer at the other end. Complex VIII is a substrate for subsequence reactions denoted "A" and "B".

IDT Exhibit 1020.9301

B) Isothermal linear amplification

9. In reaction A, a first composite primer binds to complex VIII. Primer extension and displacement produces first displacement product A. RNase cleavage creates a site for binding of a first composite primer, and subsequence primer extension, whereby single stranded antisense DNA product accumulates.
10. In reaction B, a second composite primer binds to complex VIII (or to first displacement product A). Primer extension and displacement produces single stranded sense DNA products.

The single stranded products can be annealed to form a double stranded complex of first and second strand cDNA, or can be prevented from annealing (or subsequently denatured) to produce a mixture of single stranded first and second strand cDNA.

Components and reaction conditions used in the methods of the invention

Template nucleic acid

The RNA target to be amplified includes RNAs from any source in purified or unpurified form, which can be RNA such as total RNA, tRNA, mRNA, rRNA, mitochondrial RNA, chloroplast RNA, DNA-RNA hybrids, or mixtures thereof, from any source and/or species, including human, animals, plants, and microorganisms such as bacteria, yeasts, viruses, viroids, molds, fungi, plants, and fragments thereof. RNAs can be obtained and purified using standard techniques in the art. Amplification of a DNA target (including genomic DNA target) would require initial transcription of the DNA target into RNA form, which can be achieved using methods disclosed in Kurn, U.S. Patent No. 6,251,639 B1, and by other techniques (such as expression systems) known in the art. Amplification of a DNA-RNA hybrid would require denaturation of the hybrid to obtain a ssRNA, or denaturation followed by transcription of the DNA strand to obtain an RNA. The target RNA 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 target RNA can be known or unknown and may contain more than one desired specific nucleic acid sequence of interest, each of which may be the same or different from

IDT Exhibit 1020.9302

each other. Therefore, the amplification process is useful not only for producing large amounts of one specific nucleic acid sequence, but also for amplifying simultaneously more than one different specific nucleic acid sequence located on the same or different nucleic acid molecules.

The initial step of the amplification of a target RNA sequence is rendering the target single stranded. If the target nucleic acid is double stranded (e.g., RNA/DNA hybrid) the initial step could be target denaturation. Denaturation may also be carried out to remove secondary structure present in a RNA target molecule. The denaturation step may be thermal denaturation or any other method known in the art.

Composite primer

The methods of the invention employ a composite primer that is composed of RNA and DNA portions. As described herein, when used for hybridizing and initiating the methods of RNA amplification as described herein, the composite primer generally comprises a DNA portion which hybridizes to the RNA target (which, as described herein, can have any of a number of sequence permutations, depending on the nature of the RNA (whether a species or a population) designed to be amplified). When used to amplify a cDNA strand produced by the methods of the invention described herein, the composite 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. It is understood that, in the following section that generally describes aspects of the composite primers used in the methods of the invention, characteristics described may be applicable to the primers if used for hybridizing and initiating the RNA amplification (production of first extension product) and/or for linear displacement amplification.

Composite primers for use in the methods and compositions of the invention comprise a sequence capable of hybridizing to a target RNA. The sequence that is capable of hybridizing to the target RNA can be based on the particular sequence of a specific target RNA (for e.g., the mRNA of a particular gene), or be based on a more general sequence type known to be present in a plurality of RNA species, such as the poly-A tail

IDT Exhibit 1020.9303

sequence generally believed in the art to be present in all eukaryotic mRNA. In addition, the sequence that is capable of hybridizing to the target RNA may comprise a sequence complementary to the poly-A tail of mRNA, and may further comprise an additional random sequence (generally not complementary to a poly-A sequence) at the 3' end of the 3' portion (or a population of random sequences).

The sequence that is capable of hybridizing to the target RNA may also comprise a random sequence. Random primers are well known in the art, see, e.g., and include at least the following: primers hybridizable to two or more sequences in a sample; and primers comprising poly-dT sequences that are hybridizable to a multiplicity of RNAs in a sample (such as all mRNA). For convenience, a single random composite primer is discussed above. However, it is understood that the term "random primer" can refer to a primer that is a member of a population of primers which are collectively designed to a desired and/or significant population of target sequences.

It is also understood that the amplification of a plurality of mRNA species in a single reaction mixture may, but not necessarily, employ a multiplicity of primers (from two to many more). Thus, the invention contemplates the use of a multiplicity of different composite primers (random or non-random) when amplifying a plurality of mRNA species in a single reaction mixture.

In some embodiments, a first composite primer is used in the methods of the invention, including those steps which involve linear displacement amplification (SPIA) of the second cDNA strand. In other embodiments, a first and second, different, composite primer are used in the methods of the invention. The second composite primer is used for the linear displacement amplification (SPIA) step, and may comprise some or all of the sequence of the first composite primer, and the first composite primer may comprise some or all of the sequence of the second composite primer. In some embodiments, the second composite primer comprises a different sequence than the first composite primer.

In some embodiments, a composite primer is designed such that the entire primer hybridizes to the target RNA. In other embodiments, a composite primer comprises a sequence, preferably at the 5' end, that is not hybridizable (under a given set of conditions) to the target (for example, a non-hybridized 5' portion that would constitute a tail when the primer is bound to the target). Individual DNA and RNA portions of the composite primer may be completely or partially hybridizable to the target RNA. For example, the 5' RNA portion of a composite primer may be partially hybridizable and partially nonhybridizable,

IDT Exhibit 1020.9304

or the DNA portion of a composite primer may be partially hybridizable and partially nonhybridizable, or both. Put another way, DNA portions can constitute part of a "tail" and RNA portions can be partially or completely hybridizable to target RNA. For example, the 5' RNA portion of a composite primer may be partially hybridizable and partially nonhybridizable, or the DNA portion of a composite primer may be partially hybridizable and partially nonhybridizable, or both.

For use in linear displacement amplification, a composite primer comprises at least one RNA portion that is capable of (a) binding (hybridizing) to a sequence on the second strand cDNA (interchangeably called "second primer extension product" or "composite primer extension product") independent of hybridization of the DNA portion(s) to a sequence on the same second strand cDNA; and (b) being cleaved with an agent such as a ribonuclease when hybridized to the second primer or fragment extension product. The composite primers bind to the second strand cDNA to form a partial heteroduplex in which only the RNA portion of the primer is cleaved upon contact with an agent which cleaves RNA in an RNA/DNA hybrid, such as an enzyme, such as a ribonuclease (such as RNase H), while the second strand cDNA remains intact, thus enabling annealing of another composite primer.

When used for the linear displacement amplification described herein, the composite primers also comprise a 3' DNA portion that is capable of hybridization to a sequence on the second strand cDNA such that its hybridization to the second strand cDNA is favored over that of the nucleic acid strand that is displaced from the second strand cDNA by the DNA polymerase. Such primers can be rationally designed based on well known factors that influence nucleic acid binding affinity, such as sequence length and/or identity, as well as hybridization conditions. In a preferred embodiment, hybridization of the 3' DNA portion of the composite primer to its complementary sequence in the second strand cDNA favored over the hybridization of the homologous sequence in the 5' end of the displaced strand to the second strand cDNA.

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

IDT Exhibit 1020.9305

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 one embodiment, the composite primer has an RNA portion of about 14 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 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.

IDT Exhibit 1020.9306

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, 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

IDT Exhibit 1020.9307

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.

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'-

IDT Exhibit 1020.9308

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 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).

IDT Exhibit 1020.9309

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 RNA 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 nucleic acid.

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

Second primer

The second primer in the methods of the invention (which primes generation of the second primer extension product, interchangeably referred to as second strand cDNA) comprises a sequence (which may or may not be the whole of the primer) that is hybridizable (under a given set of conditions) to a first strand cDNA (interchangeably called first primer extension product) at a site on the first strand cDNA such that the second strand cDNA would include the RNA 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 first strand cDNA. In other embodiments, the hybridizable sequence is based on random sequences, for example, known in the art to be suitable for random priming of first strand cDNAs generated from a plurality of RNA species. In other embodiments, the second primer comprises a strand switch oligonucleotide, described in U.S. Patent Nos. 5,962,271 and 5,962,272, which is hybridizable to the Cap sequences present on mRNA and causes the reverse transcriptase to switch from the mRNA template to the switch oligonucleotide, permitting generation of a second strand cDNA primed by the "switch oligonucleotide".. Alternatively, a homopolymeric tail is added to the 3'

IDT Exhibit 1020.9310

terminus of the first primer extension product, and the second primer comprises the complement of the homopolymeric tail.

In some embodiments, the second primer comprises DNA. In other embodiments, the second primer consists of DNA. In other embodiments, as described herein, the second primer is a fragment of the target RNA, with the fragment being generated by cleavage of the RNA target.

In some embodiments, the second primer (which primes generation of the second strand cDNA) is a composite primer (as described above). In these embodiments, the method involves the following: (a) formation of a double stranded cDNA comprising a RNA-DNA heteroduplex at one end of the cDNA; and (b) linear amplification of first strand (sense) cDNA, whereby multiple copies of single stranded first strand cDNA is generated.

To achieve hybridization to a first strand cDNA (which, as is well known and understood in the art, depends on other factors such as, for example, ionic strength and temperature), the sequence of the second primer that is hybridizable to the first strand cDNA is preferably of at least about 60%, more preferably at least about 75%, even more preferably at least about 90%, and most preferably at least about 95% complementarity to the first strand cDNA.

In certain embodiments (typically, but not necessarily, ones that include transcription), the second primer may also comprise a sequence, preferably a sequence at the 5' end (which generally includes the 5' most nucleotide), that is not hybridizable to a first strand cDNA under a given set of conditions. This sequence enables the creation of a defined end sequence for the 5' end of the second strand cDNA (and thus, subsequently the 3' end of the single stranded DNA products). Having a defined end sequence at the 3' end of the single stranded DNA products is particularly advantageous with respect to hybridization (in embodiments that include a transcription step) of a propromoter polynucleotide to displaced first primer extension products in subsequent steps. In certain embodiments, a 5' non-hybridizable sequence comprises a sequence the complement of which is hybridizable by a propromoter polynucleotide. Single stranded DNA products comprising a 3' defined end sequence are also useful for hybridization to a complementary oligonucleotide attached to a binding partner or substrate, for example, a generic microarray as described herein. Accordingly, the invention provides methods of making these products with a 3' defined end, as described herein.

IDT Exhibit 1020.9311

In one embodiment, the second primer comprises DNA. In another embodiment, the second primer comprises 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 composite primer extension product. In these embodiments, a sequence at the 3' end of the composite primer extension product hybridizes to another sequence in the composite primer extension product itself, for example as described in U.S. Pat. No. 6,132,997. In these embodiments, said sequence at the 3' of the composite primer extension product is generally cleaved (for example, with S1 nuclease) following its hybridization to the composite primer extension product and/or its extension along the composite primer extension product. U.S. Pat. No. 6,132,997.

In some embodiments, the second primer is provided by one or more target RNA fragments. Such a target RNA fragment can be generated as a result of incomplete degradation of a target RNA in a complex of target RNA and first primer extension product by an agent (such as an enzyme) that cleaves RNA in an RNA/DNA hybrid, such that one or more RNA fragments remain bound to the first primer extension product.

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.

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 an embodiment of the invention, the portion capable of hybridizing to the 3' end of a primer

IDT Exhibit 1020.9312

extension product comprises a sequence the complement of which is hybridizable to a defined end sequence of the second primer extension product (and thus, subsequently the 3' end of the single stranded DNA products). In another embodiment, the portion capable of hybridizing to the 3' end of a primer extension product comprises a random sequence. In another embodiment, the portion capable of hybridizing to the 3' end of a primer extension product comprises a sequence the complement of which is capable of hybridizing to sequences found at the 3' end of a multiplicity of first strand cDNAs.

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 displaced primer extension product, typically comprising all or part of the propromoter sequence), DNA polymerase fills in the overhang to create a double stranded promoter capable of effecting transcription by a suitable RNA polymerase.

Promoter sequences that allow transcription of a template DNA are known in the art and have been discussed above. Preferably, the promoter sequence is selected to provide optimal transcriptional activity of the particular RNA polymerase used. Criteria for such selection, i.e., a particular promoter sequence particularly favored by a particular RNA polymerase, is also known in the art. For example, the sequences of the promoters for transcription by T7 DNA dependent RNA polymerase and SP6 are known in the art. The promoter sequence can be from a prokaryotic or eukaryotic source.

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

IDT Exhibit 1020.9313

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.

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.

IDT Exhibit 1020.9314

DNA polymerase, an agent capable of cleaving an RNA-DNA hybrid, and RNA polymerase

The amplification methods of the invention employ the following enzymes: an RNA-dependent DNA polymerase, a DNA-dependent DNA polymerase, an agent capable of cleaving an RNA strand of an RNA-DNA hybrid (for example, a ribonuclease such as RNase H), and, in some aspects a DNA-dependent RNA polymerase. One or more of these activities may be found and used in a single enzyme. For example, RNase H activity may be supplied by an RNA-dependent DNA polymerase (such as reverse transcriptase) or may be provided in a separate enzyme. Reverse transcriptases useful for this method may or may not have RNase H activity.

One aspect of the invention is the formation of double stranded cDNA from a primer-RNA complex. This process generally utilizes the enzymatic activities of an RNA-dependent DNA polymerase, a DNA-dependent DNA polymerase and a agent capable of cleaving an RNA/DNA hybrid (such as RNase H).

RNA-dependent DNA polymerases for use in the methods and compositions of the invention are capable of effecting extension of a primer according to the methods of the invention. Accordingly, a preferred RNA-dependent DNA polymerase is one that is capable of extending a nucleic acid primer along a nucleic acid template that is comprised at least predominantly of ribonucleotides. Suitable RNA-dependent DNA polymerases for use in the methods and compositions of the invention include reverse transcriptase. Many reverse transcriptases, such as those from avian myeoloblastosis virus (AMV-RT), and Moloney murine leukemia virus (MMLV-RT) comprise more than one activity (for example, polymerase activity and ribonuclease activity) and can function in the formation of the double stranded cDNA molecules. However, in some instances, it is preferable to employ a reverse transcriptase which lacks the RNase H activity. Reverse transcriptase devoid of RNase H activity are known in the art, including those comprising a mutation of the wild type reverse transcriptase where the mutation eliminates the RNase H activity. In these cases, the addition of an RNase H from other sources, such as that isolated from *E. coli*, can be employed for the formation of the double stranded cDNA.

DNA-dependent DNA polymerases for use in the methods and compositions of the invention are capable of effecting extension of the composite primer according to the methods of the invention. Accordingly, a preferred polymerase is one that is capable of

IDT Exhibit 1020.9315

extending a nucleic acid primer along a nucleic acid template that is comprised at least predominantly of deoxynucleotides. The formation of the double stranded cDNA can be carried out by reverse transcriptase which comprises both RNA-dependent DNA polymerase and DNA-dependent DNA polymerase activities. Amplification of an RNA sequence according to methods of the invention involves the use of a DNA polymerase that is able to displace a nucleic acid strand from the polynucleotide to which the displaced strand is bound, and, generally, the more strand displacement capability the polymerase exhibits (i.e., compared to other polymerases which do not have as much strand displacement capability) is preferable. Preferably, the DNA polymerase has high affinity for binding at the 3'-end of an oligonucleotide hybridized to a nucleic acid strand. Preferably, the DNA polymerase does not possess substantial nicking activity. Generally, the DNA polymerase preferably has little or no 5'->3' exonuclease activity so as to minimize degradation of primer, or primer extension polynucleotides. Generally, this exonuclease activity is dependent on factors such as pH, salt concentration, whether the template is double stranded or single stranded, and so forth, all of which are familiar to one skilled in the art. Mutant DNA polymerases in which the 5'->3' exonuclease activity has been deleted, are known in the art and are suitable for the amplification methods described herein. Mutant DNA polymerases which lack both 5' to 3' nuclease and 3' to 5' nuclease activities have also been described , for example, exo$^{-/-}$Klenow DNA polymerase. It is preferred that the DNA polymerase displaces primer extension products from the template nucleic acid in at least about 25%, more preferably at least about 50%, even more preferably at least about 75%, and most preferably at least about 90%, of the incidence of contact between the polymerase and the 5' end of the primer extension product. In some embodiments, the use of thermostable DNA polymerases with strand displacement activity is preferred. Such polymerases are known in the art, such as described in U.S. Pat. No. 5744312 (and references cited therein). Preferably, the DNA polymerase has little to no proofreading activity

Suitable DNA polymerases for use in the methods and compositions of the invention include those disclosed in U.S. Pat. Nos. 5648211 and 5744312, which include exo$^-$ Vent (New England Biolabs), exo$^-$ Deep Vent (New England Biolabs), Bst (BioRad), exo$^-$ Pfu (Stratagene), Bca (Panvera), sequencing grade Taq (Promega), exo$^{-/-}$Klenow DNA polymerase, and thermostable DNA polymerases from thermoanaerobacter thermohydrosulfuricus.

IDT Exhibit 1020.9316

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 in an RNA/DNA hybrid regardless of the identity and type of nucleotides adjacent to the ribonucleotide to be cleaved. It is preferred that the ribonuclease cleaves independent of sequence identity. Examples of suitable ribonucleases for the methods and compositions of the invention are well known in the art, including ribonuclease H (RNase H) including Hybridase.

The DNA-dependent RNA polymerases for use in the methods and compositions of the invention are known in the art. Either eukaryotic or prokaryotic polymerases may be used. Examples include T7, T3 and SP6 RNA polymerases. Generally, the RNA polymerase selected is capable of transcribing from the promoter sequence provided by the propromoter polynucleotides as described herein. Generally, the RNA polymerase is a DNA-dependent polymerase, which is preferably capable of transcribing from a single stranded DNA template so long as the promoter region is double stranded.

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

IDT Exhibit 1020.9317

the total ionic strength of the medium. For example, the range of a salt such as KCl is preferably from about 0 to about 125 mM, more preferably from about 0 to about 100 mM, and most preferably from about 0 to about 75 mM. The reaction medium can further include additives that could affect performance of the amplification reactions, but that are not integral to the activity of the enzyme components of the methods. Such additives include proteins such as BSA, single strand binding proteins (for e.g., T4 gene 32 protein), and non-ionic detergents such as NP40 or Triton. Reagents, such as DTT, that are capable of maintaining enzyme activities can also be included. Such reagents are known in the art. Where appropriate, an RNase inhibitor (such as Rnasin) that does not inhibit the activity of the RNase employed in the method can also be included. Any aspect of the methods of the invention can occur at the same or varying temperatures. Preferably, the amplification reactions (particularly, primer extension other than the first and second strand cDNA synthesis steps, and strand displacement) are performed isothermally, which avoids the cumbersome thermocycling process. The amplification reaction is carried out at a temperature that permits hybridization of the oligonucleotides (primer and/or PTO) of the invention to the template polynucleotide and primer extension products, and that does not substantially inhibit the activity of the enzymes employed. The temperature can be in the range of preferably about 25°C to about 85°C, more preferably about 30°C to about 80°C, and most preferably about 37°C to about 75°C. In some embodiments that include RNA transcription, the temperature for the transcription steps is lower than the temperature(s) for the preceding steps. In these embodiments, the temperature of the transcription steps can be in the range of preferably about 25°C to about 85°C, more preferably about 30°C to about 75°C, and most preferably about 37°C to about 70°C.

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,

IDT Exhibit 1020.9318

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.

The oligonucleotide components of the amplification reactions of the invention are generally in excess of the number of target nucleic acid sequence to be amplified. They can be provided at about or at least about any of the following: 10, 10^2, 10^4, 10^6, 10^8, 10^{10}, 10^{12} times the amount of target nucleic acid. Composite primers and PTO can each be provided at about or at least about any of the following concentrations: 50 nM, 100 nM, 500 nM, 1000 nM, 2500 nM, 5000 nM.

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 RNA, as determined by their thermal stability and/or other considerations known to the person of skill in the art. The first strand cDNA (composite primer extension product) and the second strand cDNA (second primer extension product) synthesis reactions can be performed consecutively, followed by the amplification steps (binding by another composite primer, primer extension and strand displacement). In these embodiments, the reaction conditions and components may be varied between the different reactions.

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 strand cDNA synthesis. Another timepoint is at the end of second strand cDNA synthesis. Methods for stopping the reactions are known in the art, including, for example, cooling the reaction mixture to a temperature that inhibits enzyme activity or heating the reaction mixture to a temperature that destroys an enzyme. Methods for resuming the reactions are also known in the art, including, for example, raising the temperature of the reaction mixture to a temperature that permits enzyme activity or replenishing a destroyed (depleted) enzyme. In some embodiments, one or more of the

IDT Exhibit 1020.9319

components of the reactions is replenished prior to, at, or following the resumption of the reactions. For example, it may be necessary to replenish the composite primer prior to beginning the linear amplification reaction if the same composite primer is being used. Alternatively, the reaction can be allowed to proceed (i.e., from start to finish) without interruption.

The reaction can be allowed to proceed without purification of intermediate complexes, for example, to remove primer. Products can be purified at various timepoints, which can be readily identified by a person of skill in the art. One timepoint is at the end of first strand cDNA synthesis. Another timepoint is at the end of second strand cDNA synthesis. We have observed that routine purification of the complex of first and second cDNA results in slightly higher amplification efficiency in subsequent linear amplification steps.

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 a composition comprising a composite primer and a second primer, wherein the second primer is a random primer. In some embodiments, the second primer comprises DNA. In other embodiments, the second primer consists of DNA. In still another example, the composition comprises a composite primer and a second primer that comprises a non-target sequence that is included for the purpose of generating displaced primer extension products

IDT Exhibit 1020.9320

to which a propromoter polynucleotide can hybridize. This second primer may also be a random primer.

In some embodiments, the composite primer comprises an RNA portion adjacent to the DNA portion. In another embodiment, the composite primer comprises 5'- and 3'-DNA portions with at least one intervening RNA portion. In other embodiments, the composite primer comprises a poly-dT portion. In another example, the invention comprises a composite primer that is a random primer. In some embodiments, the random composite primer or composite primer comprising a poly-dT portion further comprise a portion not hybridizable to a target (under conditions where a portion of the primer hybridizes to target). In other examples, the invention provides a composition comprising a composite primer that is further derivatized by attachment of a moiety capable of effecting attachment of a polynucleotide comprising the composite primer to a solid substrate used in preparing nucleic acid microarrays. In some embodiments, the composite primer is further derivatized by attachment of a positively charged moiety such as an amine.

In some embodiments, the invention provides a composition comprising a composite primer and a polynucleotide comprising a propromoter sequence, such as a PTO (i.e., any of those embodiments described herein). With respect to compositions containing a random primer, these compositions may also contain a plurality of random primers (i.e., a population of random primers having different sequences).

In certain embodiments, the composition comprises (a) a composite primer; (b) a second primer (which can be a random primer); and (c) a reverse transcriptase. In yet other embodiments, the composition comprises (a) a composite primer; (b) a second primer (which can be a random primer); (c) a reverse transcriptase; and (d) a DNA polymerase. In some embodiments, the composition comprises (a) a composite primer; (b) a second primer (which can be a random primer); and (c) a polynucleotide comprising a propromoter sequence (which can be a PTO). Any of the above compositions may further comprise target RNA (which comprises an RNA sequence of interest) and/or any of the enzymes described herein (such as DNA polymerase, for example, reverse transcriptase, RNase H, and/or RNA polymerase). The compositions are generally in lyophilized or 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 DNA or RNA

IDT Exhibit 1020.9321

molecules which are copies or the complement of a target sequence, which are produced by any of the methods described herein (or compositions comprising the products). .

In another aspect, the invention provides a population of sense polynucleotide (preferably, DNA) molecules and antisense polynucleotide (preferably, DNA) molecules which are copies and complements of a target sequence, which are produced by any of the methods described herein. The invention also includes compositions and various configurations (such as arrays) of these populations, which may be homogeneous (same sequence) or heterogeneous (different sequence). These populations may be any assembly of sequences obtained from the methods described herein, including based on mRNA, as well as certain species or classes of mRNA.

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 an RNA sequence; sequencing of an RNA sequence of interest; detection of sequence mutation based on amplifying an RNA sequence (e.g., genotyping or nucleic acid mutation detection); determining presence or absence of a sequence of interest; methods of expression profiling; methods of subtractive hybridization; methods of preparing a subtractive hybridization probe; methods of differential amplification; methods of preparation of libraries (including cDNA and differential expression libraries); methods of preparation of an immobilized nucleic acid (which can be a nucleic acid immobilized on a microarray), and methods of characterizing amplified nucleic acid products generated by the methods of the invention.

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. A kit may comprise any of the composite primers described herein. In some embodiments, a kit comprises two or more composite primers and second primers, which may or may not be separately packaged. A kit may comprise a composite primer and a polynucleotide comprising a propromoter sequence (which may be a PTO). A kit may further comprise a second primer (which can be a random primer). The composite primer may be labeled or unlabeled. Kits may also optionally further include any of one or more

IDT Exhibit 1020.9322

of the enzymes described herein (for example, RNA-dependent DNA polymerase such as reverse transcriptase, and ribonuclease such as RNase H), as well as deoxynucleoside triphosphates (labeled or unlabeled) and/or ribonucleoside triphosphates (labeled or unlabeled). Kits may also include one or more suitable buffers (as described herein). Kits useful for nucleic acid sequencing may optionally include labeled or unlabeled nucleotide analogs that upon incorporation into a primer extension product or RNA transcript effect termination of nucleotide polymerization. One or more reagents in the kit can be provided as a dry powder, usually lyophilized, including excipients, which on dissolution will provide for a reagent solution having the appropriate concentrations for performing any of the methods described herein. Each component can be packaged in separate containers or some components can be combined in one container where cross-reactivity and shelf life permit.

The kits of the invention may optionally include a set of instructions, generally written instructions, although electronic storage media (*e.g.*, magnetic diskette or optical disk) containing instructions are also acceptable, relating to the use of components of the methods of the invention for the intended nucleic acid amplification, and/or, as appropriate, for using the amplification products for purposes such as nucleic acid sequencing and detection of sequence mutation. The instructions included with the kit generally include information as to reagents (whether included or not in the kit) necessary for practicing the methods of the invention, instructions on how to use the kit, and/or appropriate reaction conditions. For example, kits of the invention can comprise: a composite primer (which can comprise a poly-dT portion and/or can be a random primer), a second primer (which can be a random primer), and instructions for using the primers to amplify RNA according to methods of the invention. In another example, kits of the invention comprise a composite primer (which can comprise a poly-dT portion and/or can be a random primer), and instructions for using the primers to amplify RNA according methods of the invention. In another example, kits of the invention comprise: a composite primer, a second primer (which can be a random primer), and instructions for generating double stranded complementary DNA from an RNA target and/or amplifying RNA according to methods of the invention. In yet another example, any of these kits further comprises a propromoter polynucleotide, and instructions for producing a duplex of primer extension product and the propromoter polynucleotide such that a double stranded promoter region is generated and/or amplifying RNA according to methods of the invention. In another example, kits of

IDT Exhibit 1020.9323

the invention comprise a composite primer (which can comprise a poly-dT portion, and/or can be a random primer) capable of generating a first strand cDNA , a second composite primer capable of hybridizing to a first strand cDNA, and instructions for using the primers to generate double stranded cDNA according to any of the methods of the invention. In another example, the kits of the invention comprise a double stranded cDNA complex (comprising first and second strand cDNA) comprising a 3' single stranded DNA portion. In yet another example, any of these kits further comprises one or more controls (which can be, for example, RNA template , composite primers, and/or double stranded cDNA complex (comprising first and second strand cDNA) comprising a 3' single stranded DNA portion).

The component(s) of the kit may be packaged in any convenient, appropriate packaging. The components may be packaged separately, or in one or multiple combinations. Where kits are provided for practicing amplification methods of the invention that involve transcription, the RNA polymerase (if included) is preferably provided separately from the components used in the steps prior to the transcription steps.

The relative amounts of the various components in the kits can be varied widely to provide for concentrations of the reagents that substantially optimize the reactions that need to occur to practice the methods disclosed herein and/or to further optimize the sensitivity of any assay.

The invention also provides systems for effecting the methods described herein. These systems comprise various combinations of the components discussed above. For example, in some embodiments, the invention provides a system suitable for producing target polynucleotide sequence (or amplifying target polynucleotide sequence) comprising (a) a composite primer (any of those described herein), (b) DNA polymerase; and (c) ribonuclease. In some embodiments, the system further comprises a polynucleotide comprising a propromoter sequence (which may be a PTO) and a DNA-dependent RNA polymerase. In other embodiments, the system further comprises an RNA-dependent DNA polymerase. Any of the systems embodiments may also comprise a template (target) sequence, as described herein. A system generally includes one or more apparatuses for performing the amplification methods of the invention. Such apparatuses include, for example, heating devices (such as heating blocks or water baths) and apparatuses which effect automation of one or more steps of the methods described herein. The methods of the invention are particularly suitable for use with miniaturized devices, as thermal cycling

IDT Exhibit 1020.9324

is not required for any of the steps. A non-limiting example of suitable devices includes the BioAnalyzer (Agilant and Caliper) and the eSensor.

The invention also provides reaction mixtures (or compositions comprising reaction mixtures) which contain various combinations of components described herein. Examples of reaction mixtures have been described. In some embodiments, the invention provides reaction mixtures comprising (a) a target RNA; (b) a composite primer comprising a 3' DNA portion and an RNA portion; (c) a second primer; and (d) DNA polymerase. As described herein, any of the composite primers may be in the reaction mixture (or a plurality of composite primers), including a composite primer that comprises a 5' RNA portion which is adjacent to the 3' DNA portion. The reaction mixture could also further comprise an enzyme which cleaves RNA from an RNA/DNA hybrid, such as RNase H. A reaction mixture of the invention can also further comprise a polynucleotide comprising a propromoter sequence as described herein. Another example of a reaction mixture is (a) a displaced primer extension product (and, as such, contains at its 5' end a sequence complementary to the 3' DNA portion of the composite primer, but not a sequence complementary to the RNA portion of the composite primer); (b) a polynucleotide comprising a propromoter sequence (for example, a PTO); and (c) RNA polymerase. 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 1-8. As an example, one complex of the invention is a complex comprising: (a) a target RNA strand; and (b) a composite primer, said composite primer comprising a 3' DNA portion and an RNA portion. The composite primer may have an RNA portion which is 5' and adjacent to the 3" DNA portion. As another example, a complex of the invention is a complex comprising: (a) a composite primer extension product; and (b) a target RNA strand. In still another example, a complex of the invention is a complex comprising: (a) a first primer extension product, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion; and (b) a second primer. In again another example, a complex of the invention is a complex comprising: (a) a first primer extension product, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion; and (b) a second primer extension product. In yet another example, a complex of the invention is a complex comprising: (a) a displaced primer extension product, wherein the primer is a composite

IDT Exhibit 1020.9325

primer comprising an RNA portion and a 3' DNA portion; and (b) a propromoter polynucleotide (such as a PTO).

In yet another example, a complex of the invention is a double stranded cDNA complex further comprising a RNA/DNA portion at one end, prepared by any of the methods described herein. In some embodiments, the double stranded cDNA complex further comprises a second RNA/DNA portion at a second end. In yet another example, the complex of the invention is a first and second primer extension product comprising a 3' single stranded DNA portion comprising a 3' single stranded DNA portion produced by any of the methods described herein. In some embodiments, the composition further comprises a second 3' single stranded region. In another example, the complex of the invention is (a) a complex of first and second primer extension product comprising a 3' single stranded DNA portion, and (b) a composite primer hybridized to second primer extension product. In another example, the complex of the invention is a complex of a first strand cDNA and a second strand cDNA (that is generated by extension along first strand cDNA of a primer). In some embodiments, the primer comprises a fragment of template RNA hybridized to the first strand cDNA. In some embodiment, the primer is DNA.

Methods 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 detection), determining the presence or absence of a sequence of interest, preparation of an immobilized nucleic acid (which can be a nucleic acid immobilized on a microarray), and characterizing nucleic acids using the amplified nucleic acid products generated by the methods of the invention, are described. Methods of expression profiling, methods of subtractive hybridization and the preparation of probes for subtractive hybridization, and methods of preparing libraries (which can be cDNA and/or differential hybridization libraries) are also described.

Sequencing of RNA targets using the methods of the invention

The amplification methods of the invention are useful, for example, for sequencing of an RNA sequence of interest. The sequencing process is carried out by amplifying a target RNA containing the sequence of interest by any of the methods described herein.

IDT Exhibit 1020.9326

Addition of nucleotides during primer extension is analyzed using methods known in the art, for example, incorporation of a terminator nucleotide or sequencing by synthesis (*e.g.* pyrosequencing).

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

In embodiments wherein the end product is in the form of RNA products, sequencing can be based on premature (deliberate) termination of RNA transcription. The inclusion of rNTP analogs, which may be labeled or unlabeled, that upon incorporation into an RNA transcript effects termination of rNTP polymerization in the reaction mixture, will result in production of truncated RNA products, which result from blocking of the RNA polymerase at sites of incorporation of the analogs.

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

IDT Exhibit 1020.9327

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 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.

Mutation detection, including mutation detection based on single stranded conformation polymorphism utilizing the amplification methods of the invention

The DNA or RNA amplification products generated according to the methods of the invention are also suitable for analysis for the detection of any alteration in the target nucleic acid sequence, as compared to a reference nucleic acid sequence which is identical to the target nucleic acid sequence other than the sequence alteration. The sequence alterations may be sequence alterations present in the genomic sequence or may be sequence alterations which are not reflected in the genomic DNA sequences, for example, alterations due to post transcriptional alterations, and/or mRNA processing, including

IDT Exhibit 1020.9328

splice variants. Sequence alterations (interchangeably called "mutations") include deletion, substitution, insertion and/or transversion of one or more nucleotide.

The DNA or RNA products of the amplification methods are suitable for single stranded conformation polymorphism (SSCP or 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 DNA or 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 DNA or RNA product using nucleases which recognize sequence dependent secondary structures may be useful for the determination of sequence specific conformation polymorphism. Such nucleases are known in the art, such as the Cleavase assay (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.

IDT Exhibit 1020.9329

Other art recognized methods of analysis for the detection of any alteration in the target nucleic acid sequence, as compared to a reference nucleic acid sequence, are suitable for use with the single stranded nucleic acid products of the amplification methods of the invention. Such methods are well-known in the art, and include various methods for the detection of specific defined sequences including methods based on allele specific primer extension, allele specific probe ligation, differential probe hybridization, and limited primer extension. See, for example, Kurn et al, U.S. Patent No. 6,251,639 B1; U.S. Patent Nos. 5,888,819; 6,004,744; 5,882,867; 5, 854, 033; 5,710,028; 6,027,889; 6,004,745; 5,763,178; 5,011,769; 5,185,243; 4,876,187; 5,882,867; 5,731,146; WO US88/02746; WO 99/55912; WO92/15712; WO 00/09745; WO 97/32040; WO 00/56925; and 5,660,988. Thus, the invention also provides methods for detection of a mutation in an RNA sequence of interest comprising a single nucleotide polymorphism, comprising: (a) amplifying a target RNA using any of the methods described herein; and (b) analyzing the amplification products for presence of an alteration (mutation) as compared to a reference single stranded polynucleotide.

Methods of determining the presence or absence of a sequence of interest

The unique properties of the second 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 (defined in the sense that location of the mutation is defined), 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.

The RNA portion(s) of the composite primer is designed to be hybridizable to the sequence of the test target RNA in which the presence of a sequence alteration is suspected. Stated alternatively, the primer comprises an RNA portion(s) that comprises a sequence that is hybridizable to the reference RNA sequence (for example, a wild type sequence) against which the sequence in the test target RNA 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.

IDT Exhibit 1020.9330

In another embodiment, the RNA portion(s) of the composite primer is designed to be hybridizable to the altered sequence suspected to be present in the test target RNA. Stated alternatively, the primer comprises an RNA portion(s) that comprises a sequence that is hybridizable to the test target RNA, and thus is not hybridizable to the reference sequence (for example, a wild type sequence) against which the sequence in the test target RNA 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 hybridizable 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 hybridizable 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 methods, such that a double stranded complex of first and second primer extension products comprising RNA/DNA heteroduplex. A ribonuclease, such as RNase H, then cleaves the RNA portion of theRNA/DNA heteroduplex. 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 another 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 second primer extension product. Another possibility is that the mismatch will result in lower efficiency of cleavage of the RNA portion of the primer thus 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

IDT Exhibit 1020.9331

the present invention can be based on absence or presence of single stranded amplification products, or quantitative comparisons of amount of accumulated primer extension product. 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 hybridizable 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 primer extension 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 hybridizable 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 amplification methods of the present invention are suitable for detection of mutation as described above.

Accordingly, the invention provides a method of determining presence or absence of a sequence of interest, said method comprising (i) amplifying a target RNA containing the sequence of interest, said amplification comprising extending a composite primer hybridized to cleaved complex of first and second primer extension product prepared by any of the methods described herein, 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 hybridizable to the RNA portion of the composite primer indicates that the second primer extension product does not comprise a sequence hybridizable to the RNA portion of the composite primer and is a sequence variant with respect to the sequence hybridizable to the

IDT Exhibit 1020.9332

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 hybridizable to the RNA portion of the composite primer indicates that the second primer extension product comprises a sequence hybridizable to the RNA portion of the composite primer and is not a sequence variant with respect to the sequence hybridizable to the RNA portion of the composite primer.

Method of preparing nucleic acids immoblized to a substrate, including a microarray of nucleic acids

The single stranded products of some of the amplification methods of the invention are suitable for immobilizing to a surface. The single stranded products are particularly suitable for preparing microarrays comprising the single stranded amplification products.

Single stranded 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., polystyrene, polypropylene, nylon), polyacrylamide, nitrocellulose, or other materials.

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.* (1995) 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,

IDT Exhibit 1020.9333

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, polystyrene, 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, 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 amenable to further characterization. The single stranded nature of some products of the methods facilitates characterization.. The methods of the invention producing single stranded products are particularly amenable to quantitative analysis, as sufficient single stranded DNA and RNA products are produced which generally accurately reflect the representation of the various mRNA in the starting material.

IDT Exhibit 1020.9334

The amplified polynucleotide products, either DNA or RNA (i.e., products of any of the amplification methods described herein), can be analyzed using, for example, probe hybridization techniques known in the art, such as Southern and 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. In addition, the single stranded DNA and RNA products may serve as starting material for other starting material for other analytical and/or quantification methods known in the art, such as real time PCR, quantitative TaqMan, quantitative PCR using molecular beacons, methods described in Kurn, U.S. Patent NO. 6,251,639, etc.. Thus, the invention includes those further analytical and/or quantification methods as applied to any of the products of the methods herein.

In one embodiment, the amplification methods of the invention are utilized to generate multiple copies of single stranded products, and analyzing single stranded products by contact with a probe.

In one embodiment, the amplification methods of the invention are utilized to generate multiple copies of single stranded polynucleotide (generally, DNA) products that are labeled by using composite primers that are labeled (in the portion(s) that is not cleaved). In another embodiment, the amplification methods of the invention are utilized to generate multiple copies of single stranded polynucleotide (DNA or RNA) products that are labeled by the incorporation of labeled nucleotides during DNA or RNA polymerization. For example, amplification according to the methods of the invention can be carried out with suitable labeled dNTPs or rNTPs. These labeled nucleotides can be directly attached to a label, or can comprise a moiety which could be attached to a label. The label may be attached covalently or non-covalently to the amplification products. Suitable labels are known in the art, and include, for example, a ligand which is a member of a specific binding pair which can be detected/quantified using a detectable second member of the binding pair. Thus, amplification of total mRNA according to the methods of the invention in the presence of, for example, Cy3-dUTP or Cy5-dUTP results in the incorporation of these nucleotides into the amplification products.

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

IDT Exhibit 1020.9335

methods to characterize (for example, detect and/or quantify) an RNA sequence of interest by generating labeled polynucleotide (generally, DNA or 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 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 RNA present in a sample. Hybridization of the labeled products is detectable, and the amount of specific labels that are detected is proportional to the amount of the labeled amplification products of a specific RNA sequence of interest. This measurement is useful for, for example, measuring the relative amounts of the various RNA species in a sample, which are related to the relative levels of gene expression, as described herein. 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 RNA species in the sample.

In another aspect, the invention provides a method of quantitating single stranded polynucleotide (generally, DNA or RNA) comprising use of an oligonucleotide (probe) of defined sequence (which may be immobilized, for example, on a microarray). In this aspect of the invention, labeled single stranded polynucleotide (generally, DNA or RNA) products comprising defined sequences at the 5' and/or 3' ends (introduced using tailed first or second primers, as described herein) are hybridizable to a defined oligonucleotides, wherein the oligonucleotide comprises the complement of the defined sequence introduced at the 5' and/or 3' end). In some embodiments, specific mRNA species are amplified using a composite and/or second primer tailed with a defined sequence that is hybridizable to a sequence immobilized on the array (depending whether the defined sequence is introduced in the composite or second primer). For example, in one embodiment, a first composite primer comprises a 3' portion which is hybridizable to a sequence of a specific RNA species, and a 5' portion that is not hybridizable to a specific RNA template, but is

IDT Exhibit 1020.9336

hybridizable to a defined oligonucleotide. In another embodiment, a second primer comprises a 3' portion which is hybridizable to a sequence of a first primer extension product, and a 5' portion that is not hybridizable to a first primer extension product, but comprises a sequence of a defined oligonucleotide. Multiple copies of single stranded labeled DNA or RNA products are produced which are hybridizable to oligonucleotide. It is understood that although a single RNA species is discussed above, multiple species may be amplified simultaneously, each with a composite primer or second primer comprising a tail hybridizable to a different defined oligonucleotide.

Determination of gene expression profile

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

Accordingly, the invention provides methods of determining gene expression profile in a sample, said method comprising: amplifying single stranded product from at least one RNA sequence of interest in the sample, using any of the methods described herein ; and determining amount of amplification products of each RNA sequence of interest, wherein each said amount is indicative of amount of each RNA sequence of interest in the sample, whereby the expression profile in the sample is determined. Generally, labeled products are generated. In one embodiment, the target RNA is mRNA. In yet another embodiment, the composite primer comprises a poly-dT sequence (such that mRNA in a sample is amplified). It is understood that amount of amplification product may be determined using quantitative and/or qualitative methods. Determining amount of amplification product includes determining whether amplification product is present or

IDT Exhibit 1020.9337

absent. Thus, an expression profile can includes information about presence or absence of one or more RNA sequence of interest. "Absent" or "absence" of product, and "lack of detection of product" as used herein includes insignificant, or *de minimus* levels.

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

Method of preparing a library

The single stranded DNA and RNA products of the methods of the invention are useful in preparing libraries, including cDNA libraries and subtractive hybridization libraries. Using the methods of the invention, libraries may be prepared from limited amount of starting material, for example, mRNA extracted from limited amount of tissue or even single cells. Accordingly, in one aspect, the methods of the invention provides preparing a library from the single stranded DNA or RNA products of the invention. In another aspect, the invention provides methods of preparing a library from the double stranded cDNA produced by the methods of the invention comprising two composite primers. Method for preparing libraries from double stranded cDNA are well known in the art. In still another aspect, the invention provides methods for making a library, said method comprising: preparing a subtractive hybridization probe using any of the methods described herein.

In some embodiments, the first composite primer is hybridizable to the poly-A sequence found in essentially all mRNAs. In other embodiments, the first composite primer is a random primer.

IDT Exhibit 1020.9338

Methods of subtractive hybridization

The amplification methods of the invention are particularly suitable for use in subtractive hybridization methods, in which (at least) a first and second target RNA population is compared, since the methods described herein are capable of amplifying multiple target RNAs in the same sample, and the methods of the invention are suitable for producing large amounts of single stranded antisense nucleic acid suitable for use as "driver" in subtractive hybridization. For example, two nucleic acid populations, one sense and one antisense, can be allowed to mix together with one population present in molar excess ("driver"). Sequence present in both populations will form hybrids, while sequences present in only one population remain single-stranded. Thereafter, various well known techniques are used to separate the unhybridized molecules representing differentially expressed sequences. See, e.g., Hamson et al., U.S. Patent No. 5,589,339; Van Gelder, U.S. Patent No. 6,291,170.

Accordingly, the invention provides methods for performing subtractive hybridization, said methods comprising: (a) preparing multiple DNA copies of the complement of at least one RNA sequences of interest from a first RNA population using any of the amplification methods described herein; and (b) hybridizing the multiple copies to a second mRNA population, whereby a subpopulation of the second mRNA population forms a complex with a nucleotide DNA copy. The invention also provides methods for performing subtractive hybridization, said methods comprising: hybridizing multiple copies of the complement of at least one RNA sequences of interest from a first RNA population using any of the amplification methods described herein to a second mRNA population, whereby a subpopulation of the second mRNA population forms a complex with a copy. In some embodiments, "driver" single stranded anti-sense DNA product of the methods of the invention is combined with tester (sense) mRNA species. In some embodiments, "driver" single stranded antisense nucleic acid (generally, DNA) product is produced using the methods of the invention described herein, and a first composite primer hybridizable to the poly-A sequence (amplifying essentially all mRNA species). In other embodiment, the first composite primer is a random primer.

In another aspect, the invention provides methods of differential amplification in which single stranded driver (antisense) DNA sequences that hybridize with tester mRNA sequence are subjected to cleavage by an agent that cleaves RNA present in a DNA/RNA

IDT Exhibit 1020.9339

hybrid, such as RNase H. Cleavage of the mRNA results in the inability to generate single stranded DNA product from the test mRNA strands. Conversely, non-cleaved tester (i.e., tester mRNA that did not hybridize to driver DNA molecules) may serve as a substrate for subsequent amplification. Amplified differentially expressed products have many uses, including as a differential expression probe, to produce differential expression libraries Accordingly, the invention provides methods for differential amplification of one or more RNA sequence of interest, said method comprising: (a) preparing multiple polynucleotide (generally, DNA) copies of the complement of at least one RNA sequences of interest from a first RNA population using any of the amplification methods described herein; (b) hybridizing the multiple copies to a second mRNA population, whereby a subpopulation of the second mRNA population forms a complex with a DNA copy; (c) cleaving RNA in the complex of step (b) with an enzyme that cleaves RNA from an RNA/DNA hybrid; and (d) amplifying an unhybridized subpopulation of the second mRNA population, whereby multiple copies of single stranded DNA complementary to the unhybridized subpopulation of the second mRNA population are generated. In some embodiments, step (d) is performed using any of the amplification methods described herein. In some embodiments, the methods comprise hybridizing multiple polynucleotide (generally, DNA) copies of the complement of at least one RNA sequences of interest from a first RNA population using any of the amplification methods described herein to a second mRNA population, whereby a subpopulation of the second mRNA population forms a complex with a DNA copy; (b) cleaving RNA in the complex of step (a) with an enzyme that cleaves RNA from an RNA/DNA hybrid; and (c) amplifying an unhybridized subpopulation of the second mRNA population, whereby multiple copies of single stranded DNA complementary to the unhybridized subpopulation of the second mRNA population are generated.

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

EXAMPLES

Example 1: Amplification of total poly-A mRNA

Poly-A mRNA from MOLT-4 cell line (CLONTECH 6587-1) was used as a target for amplification. The process of amplification was in three steps: 1) synthesis of first

IDT Exhibit 1020.9340

cDNA strand; 2) synthesis of second cDNA strand to produce a double stranded cDNA from the total mRNA of the sample; and 3) amplification of the total mRNA. The double stranded cDNA product comprises at one end an RNA/DNA heteroduplex, which is a substrate for RNase H. The sequence of the two strands of this heteroduplex portion is not related to the target, and is incorporated through utilization of a composite (first) primer.

Primer sequences:

MTA1: *GACGGAUGCGGUCU*TTTTTTT
MTA2: *GACGGAUGCGGUCU*TTTTTTTN
MTA3: *GACGGAUGCGGUCU*TTTTTTTNN

wherein italicized nucleotides denote ribonucleotides and "N" denotes a degenerate nucleotide (i.e., it can be A, T, C or G).

Step 1: Synthesis of the first strand cDNA from poly A mRNA

0.1 μg of total poly-A mRNA was mixed with the following reagents in a total volume of 10 ul:

0.2 μl primer MTA3 (100 μM)
0.5 μl dNTPs (25 mM)
0.1 μl Rnasin
0.1 μl DTT
2 μl 5X AMV reverse transcriptase reaction buffer
DEPC treated water to 10 μl total volume

The reaction mixture was incubated for 2 min at 75°C, and then cooled to 37°C. 1 μl AMV reverse transcriptase (USB 70041Y, 15U/μl) was added to each reaction and the reaction mixture was further incubated at this temperature for 60 min.

Step 2: Second strand cDNA synthesis

IDT Exhibit 1020.9341

The first strand cDNA reaction mixture was mixed with 10 µl of the second strand cDNA synthesis mixture containing the following:

1 µl 10X Klenow reaction buffer
0.1 µl dNTPs (25 mM)
0.5 µl Klenow (USB 2141Y 5U/µl) DNA polymerase
8.4 µl water

The reaction mixture was incubated for 30 min at 37°C, followed by heating to 75°C for 5 min to stop the reactions by inactivating the enzymes.

Step 3: Amplification of total cDNA

Two composite primers were tested -- MTA1 and MTA2.

The reactions were carried out in a total volume of 20 µl, comprising the following:
1 µl cDNA reaction
0.2 µl MTA1 or MTA2 primer (both are at 100µM)
0.2 µl 25 mM dNTPs
0.1 µl Rnasin
0.1 µl DTT
17.2 µl water

The mixture above was incubated at 94°C for 20 seconds and then cooled to 50°C. A mixture of 2U BCA, 0.02 U Hybridase (RNase H) and 0.4 µg T4 Gene 32 protein (single stranded DNA binding protein) was added, and the reaction mixture was incubated at 50°C for 60 min.

5 µl of each reaction mixture was analyzed by electrophoresis on 5-20% PAGE (Novex). Successful amplification was indicated by the reaction products of the amplified total mRNA appearing as a smear, which was expected due to amplification of a plurality of mRNA species. No product was observed in reactions carried out without one of the

IDT Exhibit 1020.9342

following components: a. input double stranded cDNA; b. primer for first strand synthesis; and c. input mRNA for the first strand cDNA synthesis.

Example 2: Characterization of products of step 2 and step 3 reactions of Example 1

In the amplification reactions of Example 1, a "unique" sequence (i.e., a sequence not hybridizable to the RNA template) is expected to be created at the 3'-end of the second strand cDNA due to the "unique" sequence of the 5' RNA portion of the composite primer used. This sequence (of the 3'-end of the second strand cDNA) is complementary to the 5'-RNA portion of the composite primer and is not related to sequences in the target RNA. To determine the presence of this sequence in the second strand cDNA that is obtained, PCR amplification of the reaction products (as found in reaction mix of step 2 of Example 1) was performed using a primer which is complementary to the expected sequence at the 3'-end of the second strand cDNA, as a forward primer, and a G3PDH-specific primer as a reverse PCR primer. This primer pair would be expected to amplify a specific product from a double stranded cDNA that has the "unique" sequence. It would not, however, be expected to generate a specific product from PCR amplification of the anti-sense DNA products (as found in the reaction mix of step 3 of Example 1), because these products would not be expected to contain the "unique" sequence (which is introduced by the RNA portion of the composite primer which is cleaved by RNase H). Since the reaction mix of step 3 of Example 1 contains predominantly amplified DNA products (that should not contain the "unique" sequence), PCR amplification of this reaction mix would be expected to be much less efficient (and thus generate substantially less products) than PCR amplification of the reaction mix of step 2 (which contains primarily double stranded cDNA product).

PCR reactions were carried out as follows:

Each 50 μl of PCR reaction contains:

0.4 μM of each primer (Biosource International)

100 μM of each dNTP (Epicenter)

2 mM Magnesium chloride (Epicenter)

IDT Exhibit 1020.9343

1-2 units Polymerase (either MasterAmp taq or MasterAmp Tfl, both from Epicenter)

5 µl 10X buffer as supplied with the enzyme.

Either 0.5 µl of a linear amplification reaction from the third step in Example 1, or a 1:20 dilution of the cDNA generated in step 2 in Example 1.

The PCR amplification cycles were 94°C for 30 seconds, 51°C for 30 seconds, and 72°C for 30 seconds. Generally, the samples were cycled 20 or 25 times. There was a final 5-minute extension at 72°C before the samples were held at 4°C.

Similar experiments were carried out with primer specific to the T-cell receptor specific mRNA (TCR) which is expressed by the MOLT4 cell line.

Expected PCR product size (base pairs) using the G3PDH primers.

REV PRIMER	FORWARD PRIMER G3PDH3	FORWARD PRIMER dMTA1
G3PDH5-2	18	62
G3PDH5-3	110	156
G3PDH5-4	157	203
G3PDH5	253	299
G3PDH5-6	309	354
G3PDH5-7	361	405

Primer sequences

G3PDH5: 5' TTT CCT GGT ATG ACA ACG AA

G3PDH5-4: 5' CCA GCA AGA GCA CAA GAG GA

G3PDH3: 5' GAT GGT ACA TGA CAA GGT

dMTA1: 5' GAC GGA TGC GGT CTT TTT TTT

IDT Exhibit 1020.9344

Expected PCR product size (base pairs) using T-Cell receptor primers

	TCR3	DMTA1
TCR5-2	160	Approx. 440
TCR5	238	Approx. 500

Primer sequences

TCR5: 5' CCC GCA ACC ACT TCC GCT GTC

TCR5-2: 5' CAA ACC CGT CAC CCA GAT CGT

TCR3: 5' CAA CAC AAG GGC GCT GAC C

The results show that the unique sequence is incorporated into the second strand cDNA, as indicated by the presence of a product that was about 250 base pairs in length when the step 2 reaction mix was PCR amplified using primers DMTA1 and G3PDH5 (amplification of a sequence of G3PDH mRNA), and a product of about 400 base pairs in length when using primers DMTA1 and TCR5-2 (amplification of a sequence of TCR beta chain mRNA). PCR amplification of the step 3 reaction mix with the same primer pairs, on the other hand, showed a greatly reduced amount of amplification products. Thus, the results demonstrated the incorporation of the "unique" sequence (of the RNA portion of the composite primer used in Example 1) into the double stranded cDNA products generated, and the absence of the sequence in the final amplified DNA products (due to cleavage of the RNA portion).

IDT Exhibit 1020.9345

Example 3: Amplification of total mRNA starting with a total RNA preparation

The ability to amplify total mRNA from a preparation of total RNA greatly simplifies the process by eliminating the mRNA purification step. Experimental demonstration of amplifying total mRNA from a total RNA preparation using methods of the invention was carried out using commercial total RNA preparation from breast cancer tumor (CLONTECH; cat. no.: 64015-1). The process of amplification of total mRNA was carried out in three steps as described in the following.

Primer sequence:

MTB2:*GAC GGA UGC GGU CU*TTTTTTTTTTTTTTNN
BA5: AAC TAC CTT CAA CTC CAT CA
BA3: GGA CTC GTC ATA CTC CTG C

wherein italicized nucleotides denote ribonucleotides and "N" denotes a degenerate nucleotide (i.e., it can be A, T, C or G).

Step 1: First strand cDNA synthesis

Each reaction mixture comprised the following:

4 μl of a 5X buffer (250 mM Tris-HCl, pH 8.3; 375 mM KCl, 15 mM MgCl2)
MTB2 primer @1μM
25 mM dNTPs
0.2 μl RNasin Ribonuclease Inhibitor (Promega N2511, 40u/μl)
1 μl 0.1 M DTT
5 μg, 1 μg, 0.2 μg or 40 ng of total RNA per reaction
DEPC- treated water to a total volume of 19 μl

IDT Exhibit 1020.9346

The reaction mixtures were incubated at 75°C for 2 minutes, and then cooled down to 42°C. SuperScript II RNase H⁻ Reverse Transcriptase (200 U, BRL 18064-022) was added to each reaction, and the reactions were incubated at 42°C for 50 minutes.

Step 2: Synthesis of second strand cDNA

10 µl of the first strand cDNA synthesis reaction mixture was aliquoted to individual reaction tubes. 20 µl of second strand synthesis stock reaction mixture was added to each tube. The second strand syntheis stock reaction mixture contained the following:

2 µl of 10X Klenow reaction buffer (10X buffer: 500 mM Tris-HCl, pH 8.0; 100 mM $MgCl_2$, 500 mM NaCl)

2U Klenow DNA polymerase (BRL 18012-021)

0.1 µl of AMV reverse transcriptase (BRL 18020-016, 25U/µl)

0.2 µl of *E coli* Ribonuclease H (BRL 18021-014, 4 U/µl)

0.2 µl (25 mM) dNTPs

0 or 0.2 µl of Ecoli DNA ligase (BRL 18052-019, 10U/µl)

The reaction mixtures were incubated at 37°C for 30 minutes. The reactions were stopped by heating to 75°C for 5 minutes to inactivate the enzymes.

Step 3: Amplification of total cDNA

Amplification was carried out using 1 µl of the second strand cDNA reaction mixture above, using the MTA1 composite primer in the presence of T4 gene 32 protein at 50°C for 60 min.

Each reaction mixture contained the following:

2 µl of 10X buffer (200 mM Tris-HCl, pH 8.5, 50 mM $MgCl_2$, 1% NP-40)

0.2 µl of dNTPs (25 mM)

IDT Exhibit 1020.9347

0.2 µl of MTA1 (100 µM)

1 µl of the second strand cDNA synthesis mixture

0.1 µl Rnasin

0.1 µl DTT (0.1M)

DEPC-treated water to a total volume of 18.8 µl

The reaction mixtures were incubated at 94°C for 20 seconds, and then cooled down to 50°C. 2U Bca (Takara Cat.# 2710A), 0.02U Hybridase Thermostable Rnase H (Epicentre H39100), and 0.4 µg T4 Gene 32 Protein (USB 70029Z)were added, and the reactions were further incubated at this temperature for 60 min.

The step 3 reaction mix (expected to contain amplified DNA products) was analyzed by gel electrophoresis (5-20% PAGE, Novex). Successful amplification was indicated by the amplification products of the total mRNA appearing as a smear, which was expected due to amplification of a plurality of mRNA species in the sample.

The incorporation of a "unique" (defined) sequence (complementary to the 5'-end RNA portion of the composite primer used) into the second strand cDNA was demonstrated by PCR amplification using specific primer pairs. Aliquots of the step 2 and step 3 reaction mixes were subjected to PCR amplification using primers G3PDH5-4/G3PDH3 or BA5/BA3 (beta actin), using conditions as described in Example 2. PCR amplification of step 2 reaction mixes resulted in substantial amounts of products of the correct size, whereas amplification of step 3 reaction mixes resulted in substantial smaller amounts of the same products. Thus, the results demonstrated the incorporation of the "unique" sequence (of the RNA portion of the composite primer used in this Example) into the double stranded cDNA products, and the absence of the sequence in the final amplified DNA products (due to cleavage of the RNA portion).

IDT Exhibit 1020.9348

Example 4: Preparation of double stranded cDNA comprising an appended defined sequence in the second strand cDNA from total RNA preparation and purified mRNA

Total RNA (1 ug) prepared from the HCT116 cell line, or mRNA (100ng) prepared from MOLT4 cell line (Clontech) was used as a target for production of the intermediate double stranded cDNA product comprising an appended defined sequence in the second strand cDNA. The appended sequence is incorporated through utilization of a composite (first) primer.

The process of preparing the first and second strand cDNA was carried out essentially as described in Example 1 and 3, and generally included the following steps: (1) synthesis of first cDNA strand; (2) synthesis of second cDNA strand to produce a double stranded cDNA comprising at one end an RNA/DNA heteroduplex which is a substrate for RNase H. Double stranded cDNA intermediate products are expected to comprise cDNA copies of multiple RNA from the target RNA sample, each cDNA with the same appended defined sequence. The sequence of the appended defined sequence is expected to be the complement of the sequence of the 5' RNA portion of the composite (first) primer that hybridizes to target RNA.

PCR experiments were performed to confirm the presence of second strand cDNA comprising (a) second strand cDNA copy of a sequence of the GAPDH mRNA known to be represented in the mRNA of both RNA target samples and (b) the defined sequence (i.e., the complement of the 5' RNA portion of the first composite primer) at the 3' end. The PCR primer pairs used were as follows:

1) A primer complementary to the unique sequence (DMTA1) and a primer complementary to the sequence of the GAPDH mRNA (GAPDH5-4) for generation of a 203 bp product that is dependent on appending of the sequence at the 3' end of the second cDNA strand.

2) Two primers complementary to the sequence of the GAPDH mRNA, GAPDH3 and primer GAPDH5-4, used for generation of a 157 bp product specific for GAPDH and is independent of appending the unique sequence at the 3' end of the second cDNA strand.

PCR was performed as described in Examples 1 and 2 using two separate preparation of cDNA from each starting template, as described above. PCR reactions were analyzed using gel electrophoresis. The results are shown in Figure 9. Lanes correspond to the reaction mixtures containing the following templates and primer pairs:

IDT Exhibit 1020.9349

1. marker	
2. cDNA from HCT116	GAPDH3/GAPDH5-4
3. cDNA from HCT116	GAPDH3/GAPDH5-4
4. cDNA from MOLT4	GAPDH3/GAPDH5-4
5. cDNA from MOLT4	GAPDH3/GAPDH5-4
6. no template	GAPDH3/GAPDH5-4
7. cDNA from HCT116	dMTA1/GAPDH5-4
8. cDNA from HCT116	dMTA1/GAPDH5-4
9. cDNA from MOLT4	dMTA1/GAPDH5-4
10. cDNA from MOLT4	dMTA1/GAPDH5-4
11. no template	dMTA1/GAPDH5-4

Arrows mark the position of expected pcr product.

As expected, a longer product was produced in HCT116 and MOLT4 samples amplified using primer pair (1), and a sorter product was generated in HCT116 and MOLT4 samples amplified using primer pair (2). No product was produced in control samples lacking template. This example demonstrates efficient appending of a defined sequence at the 3' end of the second cDNA using the methods described herein.

Example 5: Amplification of total polyA mRNA and quantification of products using Real Time PCR

200 ng of total RNA from Human Colon Tumor Total RNA, (Clontech Catalog No. 64014-1), was used as a target for amplification. The preparation of first and second strand cDNA and subsequent amplification step was carried out essentially as described in Example 3 with the following modifications:

(1) The reaction mixture for second strand cDNA synthesis contained Klenow DNA polymerase (which lacks 3' and 5' exonuclease activities), and lacked ligase.

(2) Amplification of the resultant cDNA was carried out using Bst polymerase (4 units, NEB) instead of Bca polymerase

Quantification was determined for the cDNA intermediates (second strand cDNA) and the antisense amplification products using four different primer pairs corresponding to four different mRNAs, and Real Time PCR according to the following protocol:

cDNA or amplification products were diluted 1:10 or 1:100 in TE buffer.

IDT Exhibit 1020.9350

Reaction mixtures for Real Time PCR were set to a total volume of 20 ul, as follows:

For each reaction

10 µl of 2X ABI SYBR Green master mix (ABI Cat # 4309155)

0.6 µl of 10 µM forward primer

0.6 µl of 10 µM reverse primer

1 µl template (dilution of either cDNA or amplification products specified above)

7.8 µl of H_2O

The following primer pairs were used for quantification of four specific expressed genes in either cDNA or amplification products generated as described above:

G6PD		
	G6PD5	5′ AGGCAGCCTCTCTGCTATAAGAAA 3′
	G6PD3	5′ GCAGGGCATTGAGGTTGG 3′
LGALS1		
	LGALS15	5′ ATGGCAGCTGACGGTGACTT 3′
	LGALS13	5′ CATGGGCTGGCTGATTT 3′
MT2A		
	MT2A5	5′ CGCCTGATGCTGGGACAG 3′
	MT2A3	5′ GTTGTACATAAAAAATCCAGGTTTGTG 3′
RPL27		
	RPL275	5′ GATCCTGCTCTTAAACGCAAGG 3′
	RPL273	5′ TGCCTGTCTTGTATCTCTCTTCAAAC 3′

PCR reaction were performed in an iCycler (BioRad), using the following thermocycling protocol:

94°C for 10 min to activate the DNA polymerase

40 cycles of: 94°C for 30 seconds followed by 60°C for 30 seconds.

Data analysis was carried out as recommended by the manufacturer.

Figure 10 show 4 traces of fluorescence reading as a function of cycle number for PCR reaction quantifying cDNA product (by amplifying the second strand cDNA) or the corresponding SPIA amplification products (by amplifying accumulated second strand

IDT Exhibit 1020.9351

cDNA). The panels depict the results of quantification experiments performed using (from top to bottom):MTA2, RPL27, LGALS1, and G6PD primer pairs, respectively. Each panel shows the results of 6 experiments using amplification product preparation (labeled "SPIA") and 2 experiments using cDNA product preparation (labeled "cDNA"). The X axis is PCR cycles and the Y axis is PCR baseline subtracted RFU.

The level of amplification of each of the gene products using the method of the invention is defined by the different number of PCR cycles required for generation of fluorescence signal (termed "CT") above a defined threshold, between reaction carried out using cDNA product template and reactions carried out using the corresponding amplification products as template. Table 1 shows a calculation of a "delta CT" value for each gene product (reflecting the comparison between CT values corresponding to reactions carried out using cDNA product template and reactions carried out using the corresponding amplification products as template), andwhich revealed that regardless of their expression level in the input total RNA, mRNAs corresponding to the four gene products in the sample are equally amplified by the method of the invention.

Table 1: Calculation of delta CT value for each gene product

Gene	cDNA CT	SPIA CT	delta CT
MT2A	37	26	11
RPL27	30	19	11
LGAL	31	21	10
G6PD	35	26	9

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

IDT Exhibit 1020.9352

CLAIMS

What is claimed is:

1. A method of generating multiple copies of a polynucleotide sequence complementary to an RNA sequence of interest, said method comprising the steps of:

(a) extending a first primer hybridized to a target RNA with an RNA-dependent DNA polymerase, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, whereby a complex comprising a first primer extension product and the target RNA is produced;

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

(c) extending a second primer hybridized to the first primer extension product with a DNA-dependent DNA polymerase, whereby a second primer extension product is produced to form a complex of first and second primer extension products;

(d) cleaving RNA from the composite primer in the complex of first and second primer extension products with an enzyme that cleaves RNA from an RNA/DNA hybrid such that a composite primer hybridizes to the second primer extension product, wherein the composite primer comprises an RNA portion and a 3' DNA portion;

(e) extending the composite primer hybridized to the second primer extension product with a DNA-dependent DNA polymerase;

whereby said first primer extension product is displaced, and whereby multiple copies of a polynucleotide sequence complementary to the RNA sequence of interest are generated.

2. The method of claim 1, wherein the second primer comprises a fragment of the target RNA hybridized to the primer extension product, said fragment generated by

IDT Exhibit 1020.9353

cleaving RNA in the complex of step (b) with an enzyme that cleaves RNA from an RNA/DNA hybrid.

3. The method of claim 1, wherein the second primer comprises DNA.

4. The method of claim 1, wherein the RNA portion of the composite primer that hybridizes to target RNA is 5' with respect to the 3' DNA portion.

5. The method of claim 4, wherein the 5' RNA portion is adjacent to the 3' DNA portion.

6. The method of claim 1, wherein the composite primer that hybridizes to target RNA comprises a 5' portion that is not hybridizable to the target RNA under conditions which the composite primer hybridizes to the target RNA.

7. The method of claim 1, wherein the composite primer that hybridizes to target RNA comprises a poly-dT sequence.

8. The method of claim 7, wherein the target RNA is mRNA.

9. The method of claim 1, wherein the composite primer that hybridizes to target RNA comprises a random sequence.

10. The method of claim 1, wherein the target RNA is mRNA, and the composite primer that hybridizes to target RNA comprises a poly-dT sequence, and further comprises a 5' portion that is not hybridizable to the target mRNA under conditions which the composite primer hybridizes to the target mRNA.

11. The method of claim 1, wherein a plurality of different composite primers are used for hybridizing to the target RNA.

12. The method of claim 1, wherein the second primer is a random primer.

IDT Exhibit 1020.9354

13. The method of claim 1, wherein the second primer comprises a 5' portion that is not hybridizable to the first primer extension product under conditions which the second primer hybridizes to the first primer extension product.

14. The method of claim 1, wherein the enzyme that cleaves RNA from an RNA/DNA hybrid is RNase H.

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

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

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

18. 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.

19. The method of claim 1, wherein at least one type of dNTP used is a labeled dNTP, whereby labeled products are generated.

20. The method of claim 1, wherein the composite primer that hybridizes to target RNA and the composite primer that hybridizes to second primer extension product are the same.

21. The method of claim 1, wherein the composite primer that hybridizes to target RNA and the composite primer that hybridizes to second primer extension product are different.

IDT Exhibit 1020.9355

22. The method of claim 1, wherein said method comprises generating multiple copies of a polynucleotide sequence complementary to two or more different sequences of interest.

23. The method of claim 22, wherein said method comprises at least two different composite primers that hybridize to target RNA.

24. The method of claim 22 or 23, wherein the composite primer that hybridizes to target RNA comprises a poly-dT portion.

25. The method of claim 22 or 23, wherein the composite primer that hybridizes to target RNA is a random primer.

26. A method of generating multiple copies of an RNA sequence of interest, said method comprising the steps of:

(a) extending a first primer hybridized to a target RNA with an RNA-dependent DNA polymerase, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, whereby a complex comprising a first primer extension product and the target RNA is produced;

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

(c) extending a second primer hybridized to the first primer extension product with a DNA-dependent DNA polymerase, whereby a second primer extension product is produced to form a complex of first and second primer extension products;

(d) cleaving RNA from the composite primer in the complex of first and second primer extension products with an enzyme that cleaves RNA from an RNA/DNA hybrid such that a composite primer hybridizes to the second primer extension product, wherein the composite primer comprises an RNA portion and a 3' DNA portion;

IDT Exhibit 1020.9356

(e) extending said composite primer hybridized to the second primer extension product with a DNA-dependent DNA polymerase, whereby said first primer extension product is displaced;

(f) hybridizing the displaced first primer extension product with a polynucleotide comprising a propromoter and a region which is hybridizable to the displaced first primer extension product under conditions which allow transcription to occur by RNA polymerase, such that RNA transcripts are produced comprising sequences complementary to the displaced first primer extension product,

whereby multiple copies of the RNA sequence of interest are generated.

27. The method of claim 26, wherein the second primer comprises a fragment of the target RNA hybridized to the primer extension product, said fragment generated by cleaving RNA in the complex of step (b) with an enzyme that cleaves RNA from an RNA/DNA hybrid.

28. The method of claim 26, wherein the second primer comprises DNA.

29. The method of claim 26, wherein the RNA portion of the composite primer that hybridizes to target RNA is 5' with respect to the 3' DNA portion.

30. The method of claim 29, wherein the 5' RNA portion is adjacent to the 3' DNA portion.

31. The method of claim 26, wherein the composite primer that hybridizes to target RNA comprises a 5' portion that is not hybridizable to the target RNA under conditions which the composite primer hybridizes to the target RNA.

32. The method of claim 26, wherein the composite primer that hybridizes to target RNA comprises a poly-dT sequence.

33. The method of claim 26, wherein the target RNA is mRNA.

IDT Exhibit 1020.9357

34. The method of claim 26, wherein the composite primer that hybridizes to target RNA comprises a random sequence.

35. The method of claim 26, wherein the target RNA is mRNA, and the composite primer that hybridizes to target RNA comprises a poly-dT sequence, and further comprises a 5' portion that is not hybridizable to the target mRNA under conditions which the composite primer hybridizes to the target mRNA.

36. The method of claim 26, wherein a plurality of different composite primers are used for hybridizing to the target RNA.

37. The method of claim 26, wherein the second primer is a random primer.

38. The method of claim 26, wherein the second primer comprises a 5' portion that is not hybridizable to the first primer extension product under conditions which the second primer hybridizes to the fist primer extension product.

39. The method of claim 26, wherein the enzyme that cleaves RNA from an RNA/DNA hybrid is RNase H.

40. The method of claim 26, wherein the RNA -dependent DNA polymerase and DNA-dependent DNA polymerase are the same enzyme.

41. The method of claim 26, wherein the RNA-dependent DNA polymerase and the enzyme that cleaves RNA from an RNA/DNA hybrid are the same enzyme.

42. The method of claim 26, wherein the DNA-dependent DNA polymerase and the enzyme that cleaves RNA from an RNA/DNA hybrid are the same enzyme.

43. The method of claim 26, 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.

IDT Exhibit 1020.9358

44. The method of claim 26, wherein at least one type of dNTP used is a labeled dNTP, whereby labeled products are generated.

45. The method of claim 26, wherein the composite primer that hybridizes to target RNA and the composite primer that hybridizes to second primer extension product are the same.

46. The method of claim 26, wherein the composite primer that hybridizes to target RNA and the composite primer that hybridizes to second primer extension product are different.

47. The method of claim 26, wherein the propromoter polynucleotide comprises a region at the 3' end which hybridizes to the displaced primer extension product, whereby DNA polymerase extension of displaced primer extension product produces a double stranded promoter from which transcription occurs.

48. The method of claim 26, wherein the propromoter polynucleotide is a propromoter template oligonucleotide (PTO).

49. The method of claim 26, wherein at least one type of rNTP used is a labeled rNTP, whereby labeled products are generated.

50. The method of claim 26, wherein said method comprises generating multiple copies of two or more different sequences of interest.

51. The method of claim 50, wherein said method comprises at least two different composite primers that hybridize to target RNA.

52. The method of claim 50 or 51, wherein the composite primer that hybridizes to target RNA comprises a poly-dT portion.

IDT Exhibit 1020.9359

53. The method of claim 50 or 51, wherein the composite primer that hybridizes to target RNA is a random primer.

54. A method of amplifying an RNA sequence of interest, comprising incubating a reaction mixture, said reaction mixture comprising:

(a) the complex of first and second primer extension products of step **(c)** of claim 1;

(b) a composite primer that is hybridizable to the second primer extension product, wherein the composite primer comprises an RNA portion and a 3' DNA portion;

(c) a DNA-dependent DNA polymerase; and

(d) an enzyme that cleaves RNA from an RNA/DNA hybrid;

wherein the incubation is under conditions that permit primer hybridization, RNA cleavage, and 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 second primer extension product in the complex and is extended, whereby multiple copies of a polynucleotide sequence complementary to the RNA sequence of interest are generated.

55. The method of claim 54, wherein the RNA portion of the composite primer that hybridizes to target RNA is 5' with respect to the 3' DNA portion, and the 5' RNA portion is adjacent to the 3' DNA portion.

56. The method of claim 54, wherein the target RNA is mRNA.

57. The method of claim 54, wherein at least one type of dNTP used is a labeled dNTP, wherein labeled products are generated.

58. A method of amplifying an RNA sequence of interest comprising incubating a reaction mixture, said reaction mixture comprising:

IDT Exhibit 1020.9360

(a) a first complex, wherein the first complex is the complex of first and second primer extension products of step (c) of claim 1;

(b) a composite primer that is hybridizable to the second primer extension product, wherein the composite primer comprises an RNA portion and a 3' DNA portion;

(c) a DNA-dependent DNA polymerase;

(d) an RNA polymerase;

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

(f) an enzyme that cleaves RNA from an RNA/DNA hybrid;

wherein the incubation is under conditions that permit primer hybridization, RNA cleavage, displacement of the first primer extension product from the first complex when its RNA is cleaved and a composite primer binds to the second primer extension product in the first complex, hybridization of the propromoter polynucleotide to the displaced first primer extension product to form a second complex comprising the displaced primer extension product and the propromoter polynucleotide, and RNA transcription from said second complex, whereby multiple copies of the RNA sequence of interest are generated.

59. The method of claim 58, wherein the RNA portion of the composite primer that hybridizes to target RNA is 5' with respect to the 3' DNA portion, and the 5' RNA portion is adjacent to the 3' DNA portion.

60. The method of claim 58, wherein the enzyme that cleaves RNA from an RNA/DNA hybrid is RNase H.

61. The method of claim 58, wherein the propromoter polynucleotide comprises a region at the 3' end which hybridizes to the displaced primer extension product, whereby

IDT Exhibit 1020.9361

DNA polymerase extension of displaced primer extension product produces a double stranded promoter from which transcription occurs.

62. The method of claim 58, wherein at least one type of rNTP used is a labeled rNTP, whereby labeled products are generated.

63. The method of claim 58, wherein the target RNA is mRNA.

64. A method of generating multiple copies of (amplifying) an RNA sequence of interest comprising:

(a) incubating a reaction mixture, said reaction mixture comprising:

(i) a target RNA;

(ii) a first primer that is hybridizable to a target RNA, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion;

(iii) a RNA-dependent DNA polymerase; and

(iv) an enzyme capable of cleaving RNA from an RNA/DNA hybrid;

wherein the incubation is under conditions that permit primer hybridization, formation of a complex comprising a first primer extension product and the target RNA and cleavage of RNA in the complex comprising a first primer extension product and the target RNA;

(b) incubating a reaction mixture, said reaction mixture comprising:

(i) the first primer extension product;

(ii) a DNA-dependent DNA polymerase; and

(iii) optionally, an enzyme capable of cleaving RNA from an RNA/DNA hybrid;

wherein the incubation is under conditions permitting formation of a complex comprising the first primer extension product and a second primer extension product, and optionally, cleavage of RNA in the complex comprising a first primer extension product and the target RNA;

(c) incubating a reaction mixture, said reaction mixture comprising:

(i) the complex comprising the first primer extension product and a second primer extension product;

(ii) an enzyme capable of cleaving RNA from an RNA/DNA hybrid;

IDT Exhibit 1020.9362

(iii) a composite primer, wherein the composite primer comprises a RNA portion and a 3' DNA portion;

(iv) the cleaved complex comprising the first primer extension product and a second primer extension product; and

(v) DNA-dependent DNA polymerase, wherein the incubation is under conditions that permit cleavage of RNA in the complex comprising the first primer extension product and a second primer extension product, composite primer hybridization, and displacement of the first primer extension product from the complex comprising the first primer extension product and a second primer extension product;

whereby multiple copies of a polynucleotide sequence complementary to the RNA sequence of interest are generated.

65. The method of claim 64, wherein step (b) includes incubating a reaction mixture further comprising: (iv) a second primer; and wherein the incubation is under conditions permitting hybridization of the second primer.

66. The method of claim 65, wherein the second primer comprises DNA.

67. The method of claim 64, wherein the RNA portion of the composite primer that hybridizes to target RNA is 5' with respect to the 3' DNA portion.

68. The method of claim 67, wherein the 5' RNA portion is adjacent to the 3' DNA portion.

69. The method of claim 64, wherein the composite primer that hybridizes to target RNA comprises a 5' portion that is not hybridizable to the target RNA under conditions which the composite primer hybridizes to the target RNA.

70. The method of claim 64, wherein the composite primer that hybridizes to target RNA comprises a poly-dT sequence.

71. The method of claim 64, wherein the target RNA is mRNA.

IDT Exhibit 1020.9363

72. The method of claim 64, wherein the composite primer that hybridizes to target RNA comprises a random sequence.

73. The method of claim 64, wherein the target RNA is mRNA, and the composite primer that hybridizes to target RNA comprises a poly-dT sequence, and further comprises a 5' portion that is not hybridizable to the target mRNA under conditions which the composite primer hybridizes to the target mRNA.

74. The method of claim 64, wherein a plurality of different composite primers are used for hybridizing to the target RNA.

75. The method of claim 64, wherein the second primer is a random primer.

76. The method of claim 64, wherein the second primer comprises a 5' portion that is not hybridizable to the first primer extension product under conditions which the second primer hybridizes to the first primer extension product.

77. The method of claim 64, wherein the enzyme that cleaves RNA from an RNA/DNA hybrid is RNase H.

78. The method of claim 64, wherein the RNA -dependent DNA polymerase and DNA-dependent DNA polymerase are the same enzyme.

79. The method of claim 64, wherein the RNA-dependent DNA polymerase and the enzyme that cleaves RNA from an RNA/DNA hybrid are the same enzyme.

80. The method of claim 64, wherein the DNA-dependent DNA polymerase and the enzyme that cleaves RNA from an RNA/DNA hybrid are the same enzyme.

81. The method of claim 64, 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.

IDT Exhibit 1020.9364

82. The method of claim 64, wherein at least one type of dNTP used is a labeled dNTP, whereby labeled products are generated.

83. The method of claim 64, wherein the composite primer that hybridizes to target RNA and the composite primer that hybridizes to second primer extension product are the same.

84. The method of claim 64, wherein the composite primer that hybridizes to target RNA and the composite primer that hybridizes to second primer extension product are different.

85. The method of claim 64, wherein said method comprises generating multiple copies of a polynucleotide sequence complementary to two or more different sequences of interest.

86. The method of claim 85, wherein said method comprises at least two different composite primers that hybridize to target RNA.

87. The method of claim 85 or 86, wherein the composite primer that hybridizes to target RNA comprises a poly-dT portion.

88. The method of claim 85 or 86, wherein the composite primer that hybridizes to target RNA is a random primer.

89. A method of generating multiple copies of a polynucleotide sequence complementary to an RNA sequence of interest, said method comprising the steps of:

(a) cleaving RNA from a complex of first and second primer extension products with an enzyme that cleaves RNA from an RNA/DNA hybrid such that a composite primer hybridizes to the second primer extension product, wherein the composite primer comprises an RNA portion and a 3' DNA portion, wherein the first primer extension product is produced by extension of a first primer hybridized to a target RNA with a RNA-dependent DNA polymerase, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion;

IDT Exhibit 1020.9365

(b) hybridizing a composite primer to the second primer extension product and extending the composite primer with a DNA-dependent DNA polymerase;

whereby said first primer extension product is displaced, and whereby multiple copies of a polynucleotide sequence complementary to the RNA sequence of interest are generated.

90. The method of claim 89, wherein the second primer comprises DNA.

91. The method of claim 89, wherein the RNA portion of the composite primer that hybridizes to target RNA is 5' with respect to the 3' DNA portion.

92. The method of claim 91, wherein the 5' RNA portion is adjacent to the 3' DNA portion.

93. The method of claim 89, wherein the composite primer that hybridizes to target RNA comprises a 5' portion that is not hybridizable to the target RNA under conditions which the composite primer hybridizes to the target RNA.

94. The method of claim 89, wherein the composite primer that hybridizes to target RNA comprises a poly-dT sequence.

95. The method of claim 94, wherein the target RNA is mRNA.

96. A method of generating multiple copies of a polynucleotide sequence complementary to an RNA sequence of interest, said method comprising the step of:

extending a composite primer in a complex comprising:

(i) a complex of a first and second primer extension products, wherein the first primer extension product is produced by extension of a first primer hybridized to a target RNA with a RNA-dependent DNA polymerase, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, wherein the second primer extension product is generated by extension of a second primer hybridized to the first primer extension product, and wherein RNA from the complex of first and second

IDT Exhibit 1020.9366

primer extension products is cleaved with an enzyme that cleaves RNA from an RNA/DNA hybrid; and

(ii) a composite primer, said composite primer comprising an RNA portion and a 3' DNA portion, wherein the composite primer is hybridized to the second primer extension product, and wherein the composite primer may be the same or different from the first primer;

whereby said first primer extension product is displaced, and whereby multiple copies of a polynucleotide sequence complementary to the RNA sequence of interest are generated.

97. The method of claim 96, wherein the second primer comprises DNA.

98. The method of claim 96, wherein the RNA portion of the composite primer that hybridizes to target RNA is 5' with respect to the 3' DNA portion.

99. The method of claim 98, wherein the 5' RNA portion is adjacent to the 3' DNA portion.

100. The method of claim 96, wherein the composite primer that hybridizes to target RNA comprises a 5' portion that is not hybridizable to the target RNA under conditions which the composite primer hybridizes to the target RNA.

101. The method of claim 96, wherein the composite primer that hybridizes to target RNA comprises a poly-dT sequence.

102. The method of claim 101, wherein the target RNA is mRNA.

103. A method of sequencing an RNA sequence of interest, said method comprising (a) amplifying a target RNA containing the sequence of interest by any of the methods of claims 1, 54, 64, 89, and 96 in the presence of a mixture of dNTPs and dNTP analogs such that primer extension is terminated upon incorporation of a dNTP analog; and (b) analyzing the amplification products to determine sequence.

IDT Exhibit 1020.9367

104. The method of claim 103, wherein the target RNA is mRNA.

105. A method of sequencing an RNA sequence of interest, said method comprising (a) amplifying a target RNA containing the sequence of interest by any of the methods of claims 26 and 58 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.

106. The method of claim 105, wherein the target RNA is mRNA.

107. A method of detecting a mutation in a target RNA by single stranded conformation polymorphism, comprising (a) amplifying the target RNA by any of the methods of claims 1, 26, 54, 58, 64, 89, and 96; 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.

108. The method of claim 107, wherein the target RNA is mRNA.

109. A method of determining presence or absence of a sequence of interest, said method comprising

(i) amplifying a target RNA containing the sequence of interest, said amplifying comprising extending a composite primer hybridized to the complex of step (d) 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 hybridizable to the RNA portion of the composite primer indicates that the second primer extension product does not comprise a sequence hybridizable to the RNA portion of the composite primer and is a sequence variant with respect to the sequence hybridizable to the RNA portion of the composite primer; or

IDT Exhibit 1020.9368

(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 hybridizable to the RNA portion of the composite primer indicates that the second primer extension product comprises a sequence hybridizable to the RNA portion of the composite primer and is not a sequence variant with respect to the sequence hybridizable to the RNA portion of the composite primer.

110. The method of claim 109, wherein the target RNA is mRNA.

111. A method of producing a nucleic acid immobilized to a substrate, comprising (a) amplifying a target RNA by any of the methods of claims 1, 26, 54, 58, 64, 89, and 96; and (b) immobilizing the amplification products on a substrate.

112. The method of claim 111, wherein the target RNA is mRNA.

113. The method of claim 111, wherein the substrate is a microarray.

114. A method of characterizing an RNA sequence of interest, comprising (a) amplifying a target RNA by any of the methods of claims 1, 54, 64, 89, and 96; and (b) analyzing the DNA products.

115. The method of claim 114, wherein the target RNA is mRNA.

116. The method of claim 114, wherein the DNA products are labeled.

117. The method of claim 114, wherein step (b) of analyzing the DNA products comprises determining amount of said products, whereby the amount of the RNA sequence of interest present in a sample is quantified.

118. The method of claim 114, wherein step (b) comprises contacting the DNA products with at least one probe.

IDT Exhibit 1020.9369

119. The method of claim 118, wherein the DNA products are labeled, and wherein the at least one probe is provided as a microarray.

120. The method of claim 119, wherein the microarray comprises at least one probe immobilized on a substrate fabricated from a material selected from the group consisting of paper, glass, ceramic, plastic, polypropylene, polystyrene, nylon, polyacrylamide, nitrocellulose, silicon, and optical fiber.

121. The method of claim 120, 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.

122. A method of characterizing an RNA sequence of interest, comprising (a) amplifying a target RNA by the methods of any of claims 26 and 58; and (b) analyzing the RNA products.

123. The method of claim 122, wherein the target RNA is mRNA.

124. The method of claim 122, wherein the RNA products are labeled.

125. The method of claim 122, wherein step (b) of analyzing the RNA products comprises determining amount of said products, whereby the amount of the RNA sequence of interest present in a sample is quantified.

126. The method of claim 122, wherein step (b) comprises contacting the labeled RNA products with at least one probe.

127. The method of claim 122, wherein the RNA products are labeled, and wherein the at least one probe is provided as a microarray.

128. The method of claim 127, wherein the microarray comprises at least one probe immobilized on a substrate fabricated from a material selected from the group

IDT Exhibit 1020.9370

consisting of paper, glass, ceramic, plastic, polypropylene, polystyrene, nylon, polyacrylamide, nitrocellulose, silicon, and optical fiber.

129. The method of claim 128, 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.

130. A method of determining gene expression profile in a sample, said method comprising: (a) amplifying at least one RNA sequence of interest in the sample using the method of any of claims 1, 54, 64, 89, and 96; and (b) determining amount of amplification products of each RNA sequence of interest, wherein each said amount is indicative of amount of each RNA sequence of interest in the sample, whereby the gene expression profile in the sample is determined.

131. The method of claim 130, wherein each target RNA is mRNA.

132. A method of preparing a library, said method comprising: amplifying at least one RNA sequences of interest using the method of any of claims 1, 26, 54, 58, 64, 89, and 96.

133. The method of claim 132, wherein the first primer that hybridizes to target RNA is a random primer

134. The method of claim 132, wherein the first primer that hybridizes to target RNA comprises a poly-dT portion.

135. A method of preparing a subtractive hybridization probe, said method comprising generating multiple DNA copies of the complement of at least one RNA sequence of interest from a first RNA population using the methods of any of claims 1, 54, 64, 89, and 96.

136. A method of performing subtractive hybridization, said method comprising: (a) generating multiple DNA copies of the complement of at least one RNA sequence of

IDT Exhibit 1020.9371

interest from a first RNA population using the methods of any of claims 1, 54, 64, 89, and 96; and (b) hybridizing the multiple copies to a second mRNA population, whereby a subpopulation of the second mRNA population forms a complex with a DNA copy.

137. The method of claim 136, said method further comprising: (c) cleaving RNA in the complex of step (b) with an enzyme that cleaves RNA from an RNA/DNA hybrid; and (d) amplifying an unhybridized subpopulation of the second mRNA population, whereby multiple copies of single stranded DNA complementary to the unhybridized subpopulation of the second mRNA population are generated.

138. A method for differential amplification of one or more RNA sequence of interest, said method comprising:

(a) hybridizing multiple polynucleotide copies of the complement of one or more RNA sequence of interest from a first RNA population using the methods of any of claims 1, 54, 64, 89, and 96 to a second mRNA population, whereby a subpopulation of the second mRNA population hybridizes to the polynucleotide copies to form complexes;

(b) cleaving RNA in the complexes of step (b) with an enzyme that cleaves RNA from an RNA/DNA hybrid; and

(c) amplifying an unhybridized subpopulation of the second mRNA population, whereby multiple copies of single stranded DNA complementary to the unhybridized subpopulation of the second mRNA population are generated.

139. A method for making a cDNA library, said method comprising: preparing a subtractive hybridization probe according to claim 84.

140. A composition comprising a composite primer, wherein the composite primer comprises an RNA portion and a 3' DNA portion, and a second primer.

141. The composition of claim 140, wherein the second primer comprises DNA.

142. The composition of claim 141, wherein the second primer is a random primer.

IDT Exhibit 1020.9372

143. The composition of claim 140, wherein the second primer comprises a fragment of the target RNA hybridized to the primer extension product.

144. The composition of claim 140, further comprising an RNA-dependent DNA polymerase.

145. The composition of claim 140, wherein the composite primer further comprises a 5' region that is not hybridizable to the RNA sequence of interest under conditions which the composite primer hybridizes to the target RNA.

146. A composition comprising a composite primer and a second primer that comprises a sequence that is not hybridizable to a first primer extension product under conditions which the composite primer hybridizes to the target RNA.

147. A composition comprising: (a) a composite primer; (b) a second primer; and (c) a propromoter polynucleotide.

148. The composition of claim 147, wherein the second primer is a random primer.

149. A composition comprising a complex of (a) a first primer extension product, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion; and (b) a propromoter polynucleotide.

150. A composition comprising a complex of (a) a first primer extension product, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion; and (b) a second primer extension product, wherein the second primer comprises DNA.

151. A composition comprising a complex of (a) a first primer extension product, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA

IDT Exhibit 1020.9373

portion; and (b) a second primer extension product, wherein the second primer comprises a fragment of the RNA target.

152. A composition comprising a complex of (a) a cleaved primer extension product, wherein the primer is a composite primer comprising an RNA portion and a 3' DNA portion; (b) a second primer extension product; and (c) a composite primer that hybridizes to second primer extension product.

153. The method of claim 152, wherein the composite primer that hybridizes to target RNA and the composite primer that hybridizes to second primer extension product are the same.

154. The method of claim 152, wherein the composite primer that hybridizes to target RNA and the composite primer that hybridizes to second primer extension product are different.

155. A reaction mixture comprising (a) a target RNA; (b) a composite primer comprising a 3' DNA portion and an RNA portion; (c) a second primer; and (d) a DNA polymerase.

156. The reaction mixture of claim 155, further comprising: (e) an enzyme which cleaves RNA from an RNA/DNA hybrid.

157. The reaction mixture of claim 155, further comprising: (e) a propromoter polynucleotide.

158. A kit for amplifying a target RNA, comprising: (a) a composite primer comprising a 3' DNA portion and an RNA portion; (b) a second primer; and (c) instructions for amplifying RNA according to any of methods 1, 26, 54, 58, 64, 89, and 96.

159. The kit of claim 158, further comprising: (d) a propromoter polynucleotide.

IDT Exhibit 1020.9374

160. The kit of claim 157 or 158, further comprising an enzyme which cleaves RNA from an RNA/DNA hybrid.

161. A method of making a complex of first and second primer extension products comprising a 3' single stranded portion comprising:

(a) extending a first primer hybridized to a target RNA with an RNA-dependent DNA polymerase, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, whereby a complex comprising a first primer extension product and the target RNA is produced;

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

(c) extending a second primer hybridized to the first primer extension product with a DNA-dependent DNA polymerase, whereby a second primer extension product is produced to form a complex of first and second primer extension products;

(d) cleaving RNA from the composite primer in the complex of first and second primer extension products with an enzyme that cleaves RNA from an RNA/DNA hybrid; '

whereby a complex of first and second primer extension products comprising a 3' single stranded portion is generated;

whereby the 3' single stranded portion is complementary to the RNA portion of the composite primer.

IDT Exhibit 1020.9375

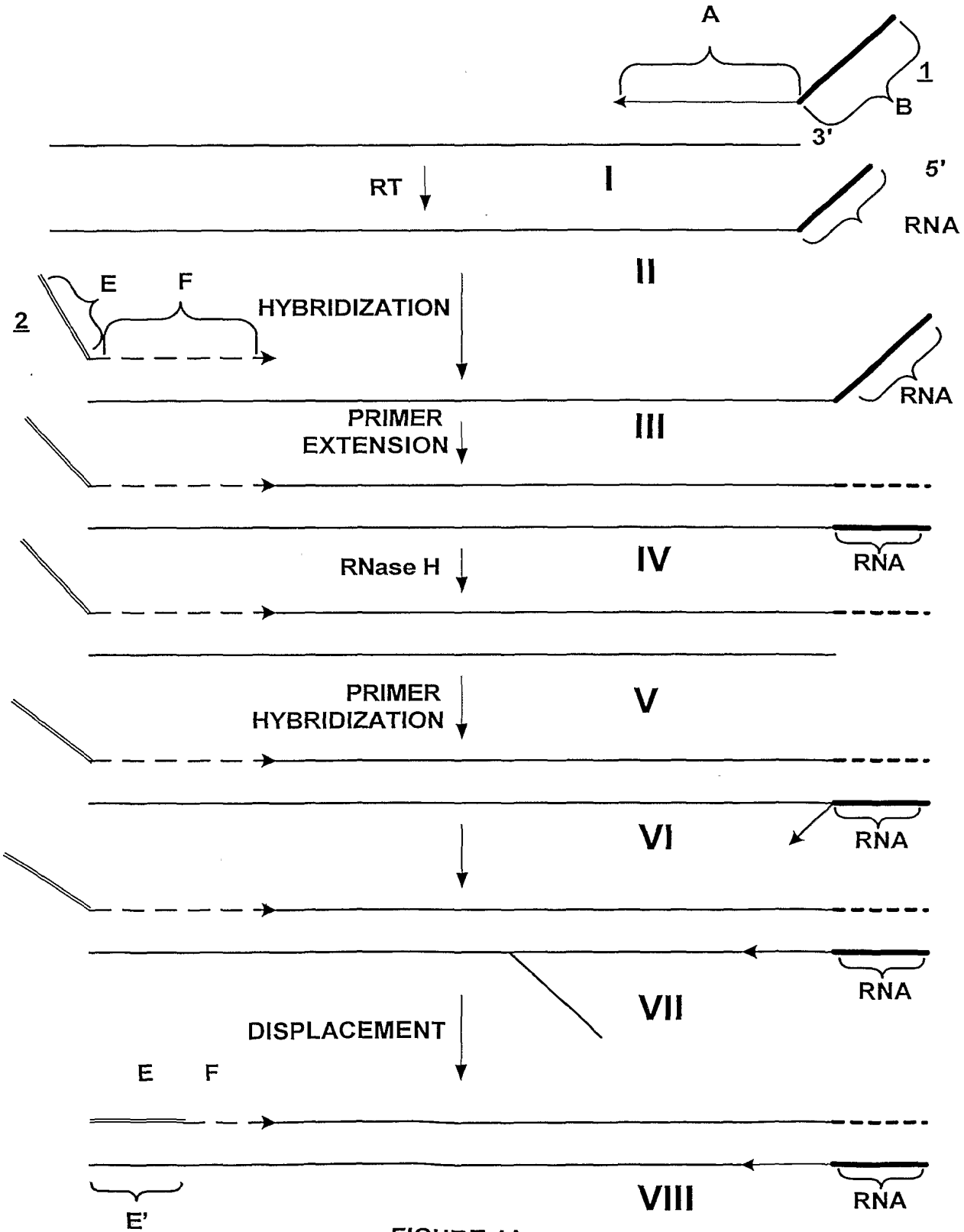


FIGURE 1A

IDT Exhibit 1020.9376

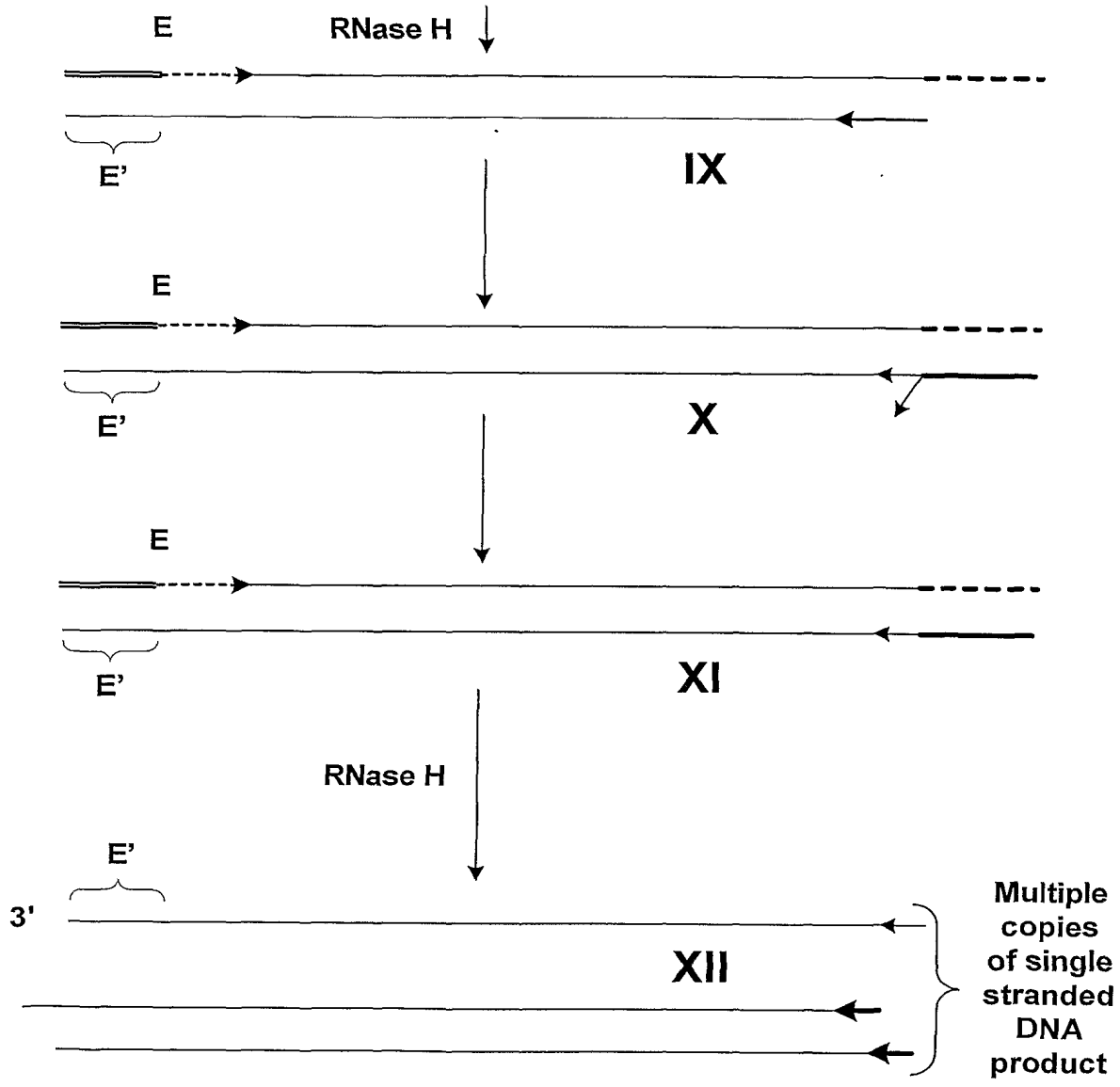


FIGURE 1B

IDT Exhibit 1020.9377

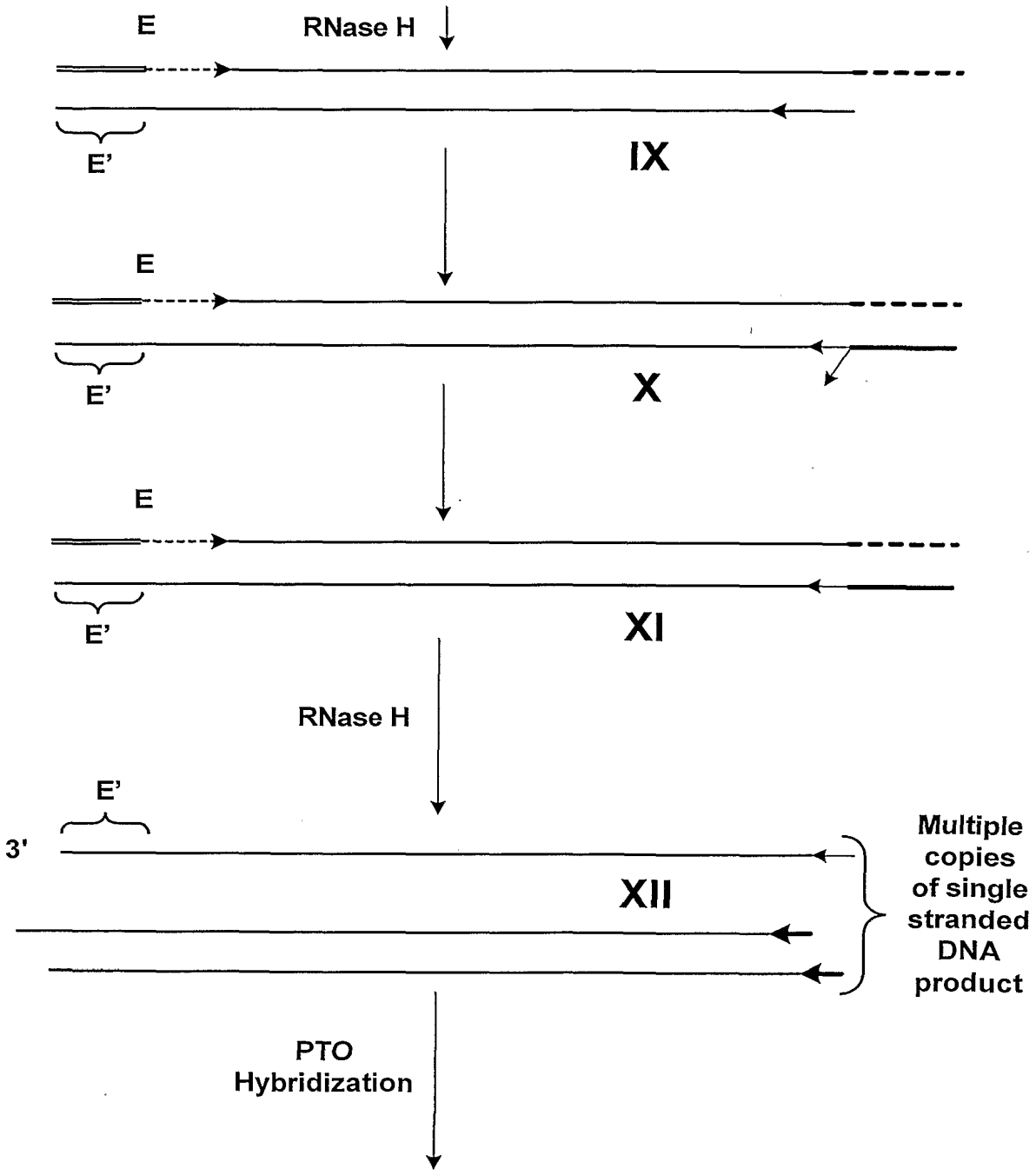


FIGURE 2B

IDT Exhibit 1020.9378

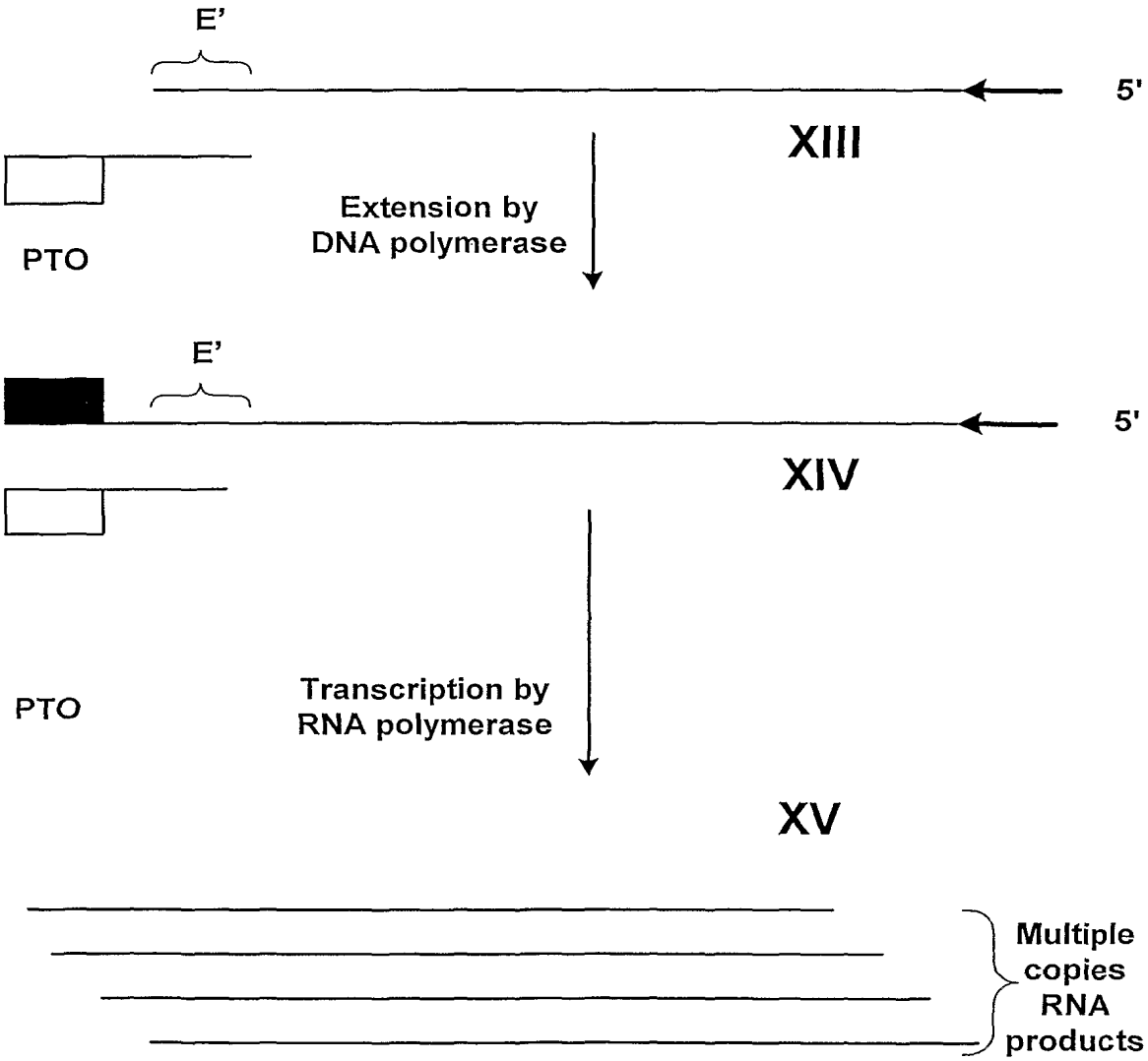


FIGURE 2C

IDT Exhibit 1020.9379

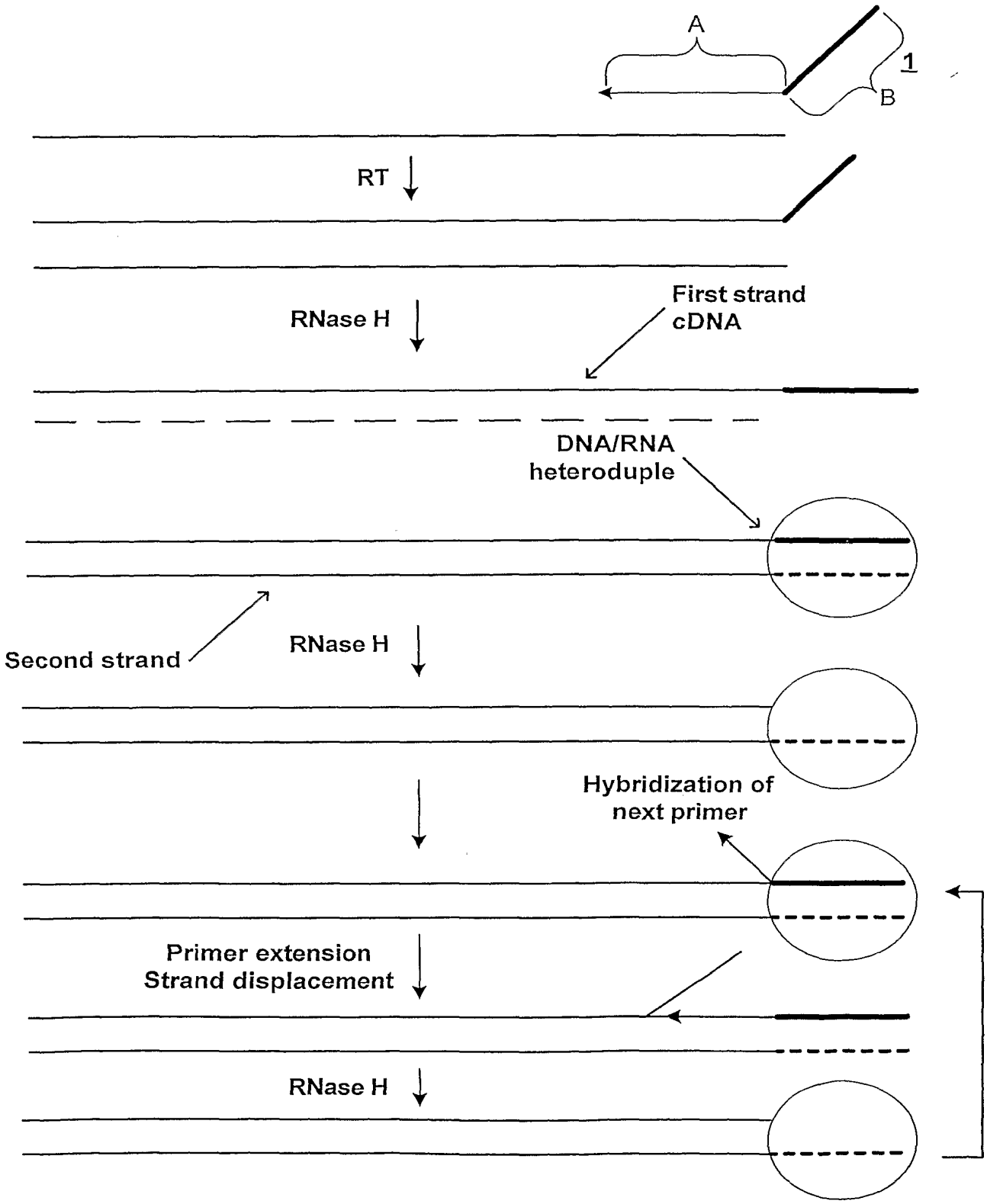


FIGURE 3A

IDT Exhibit 1020.9380

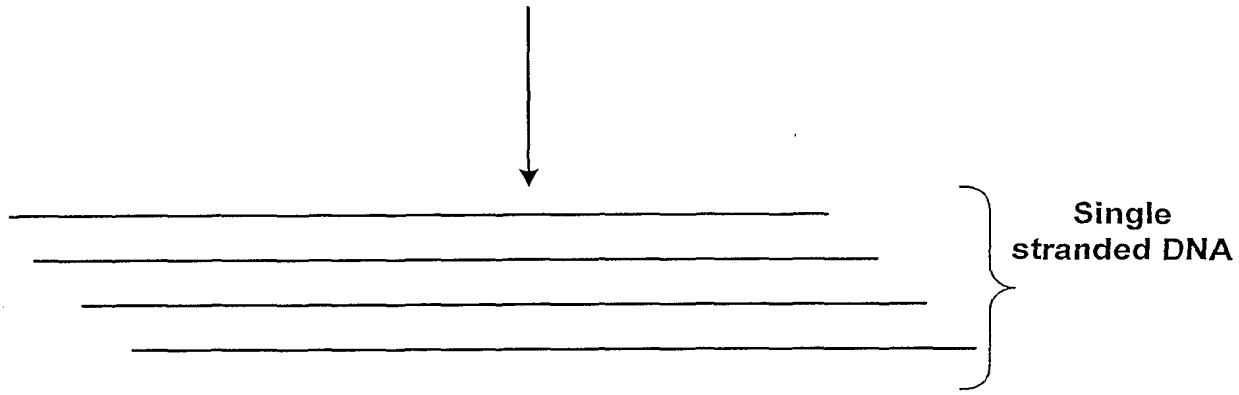


FIGURE 3B

IDT Exhibit 1020.9381

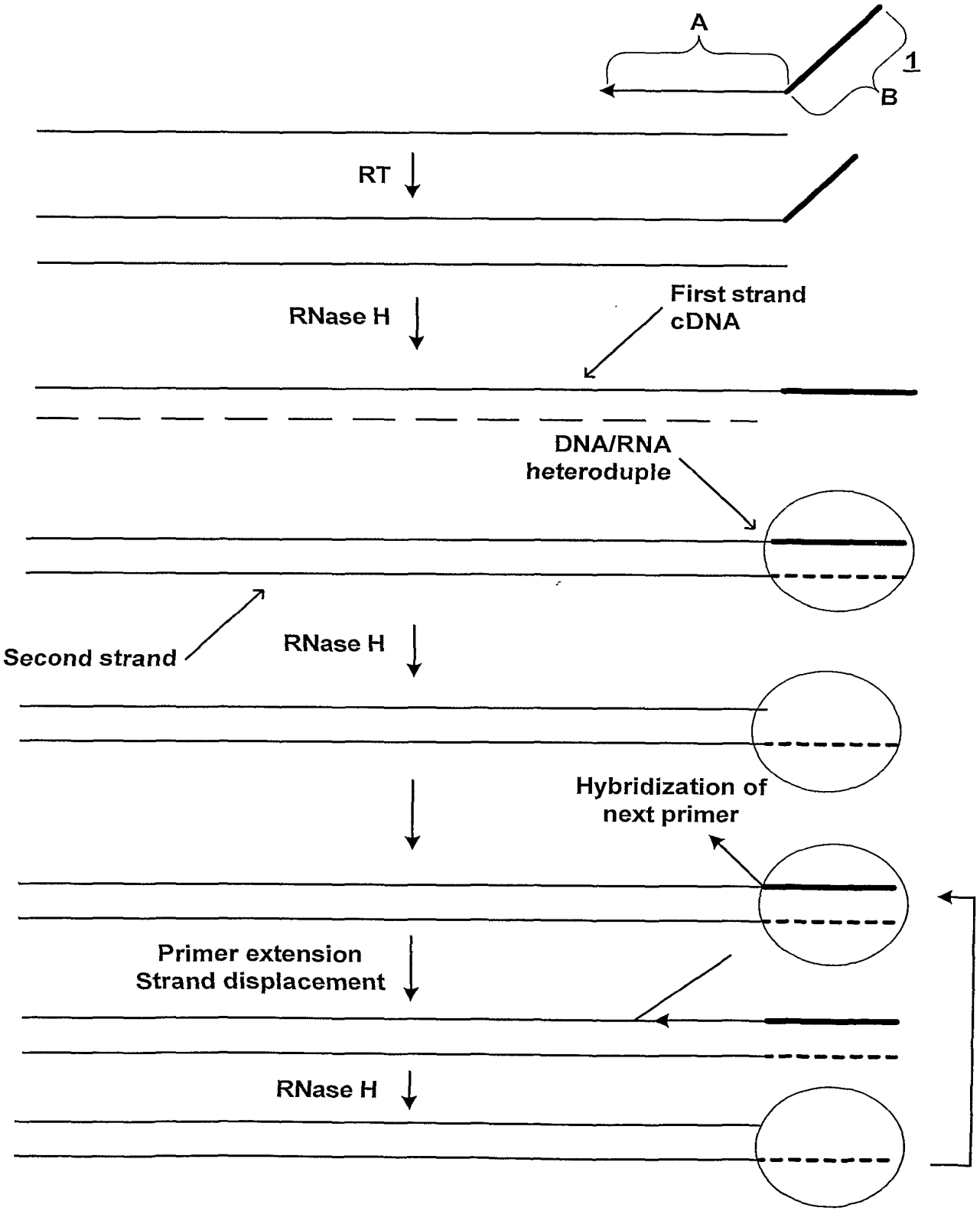


FIGURE 4A

IDT Exhibit 1020.9382

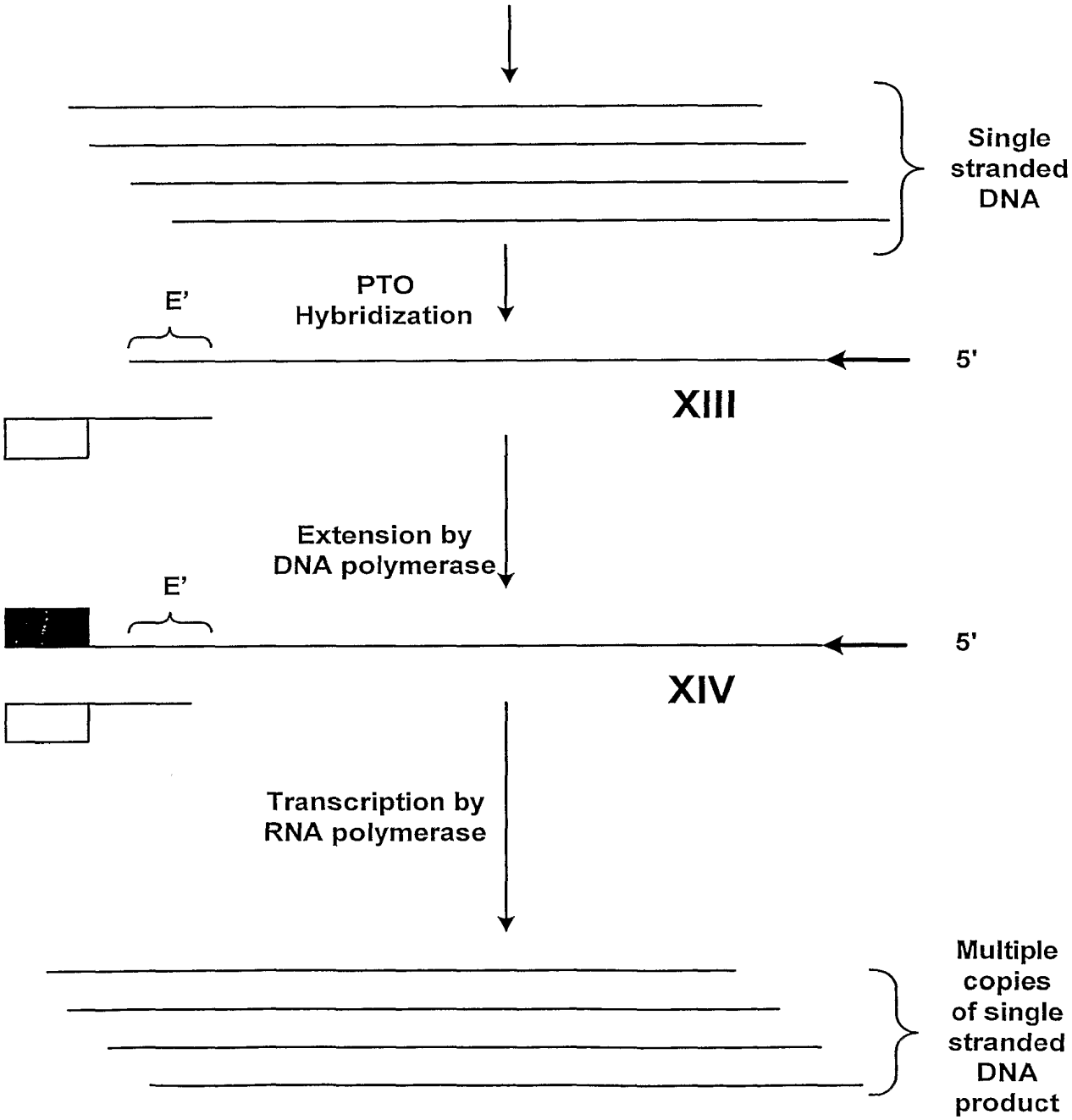


FIGURE 4B

IDT Exhibit 1020.9383

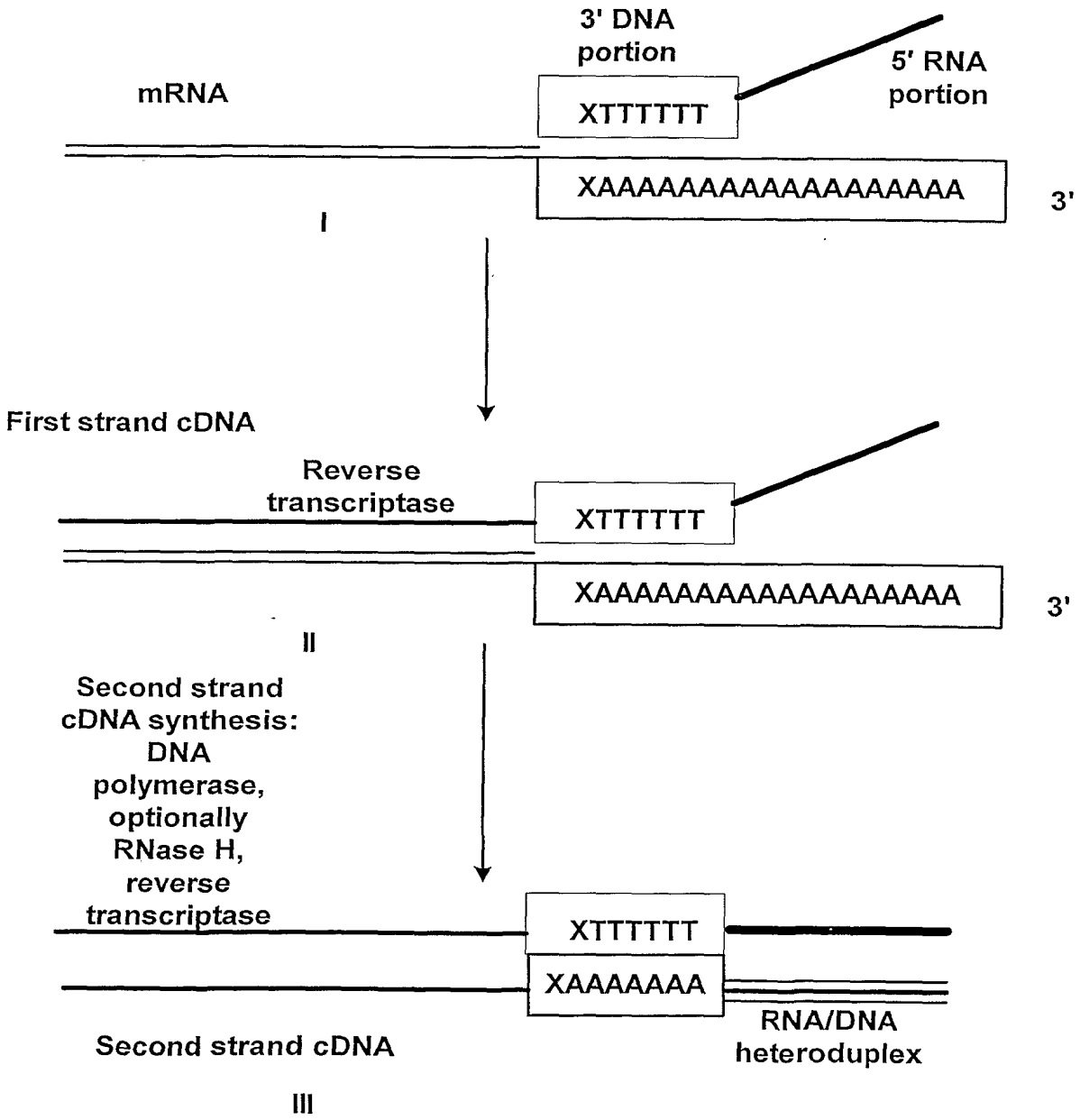


FIGURE 5A

IDT Exhibit 1020.9384

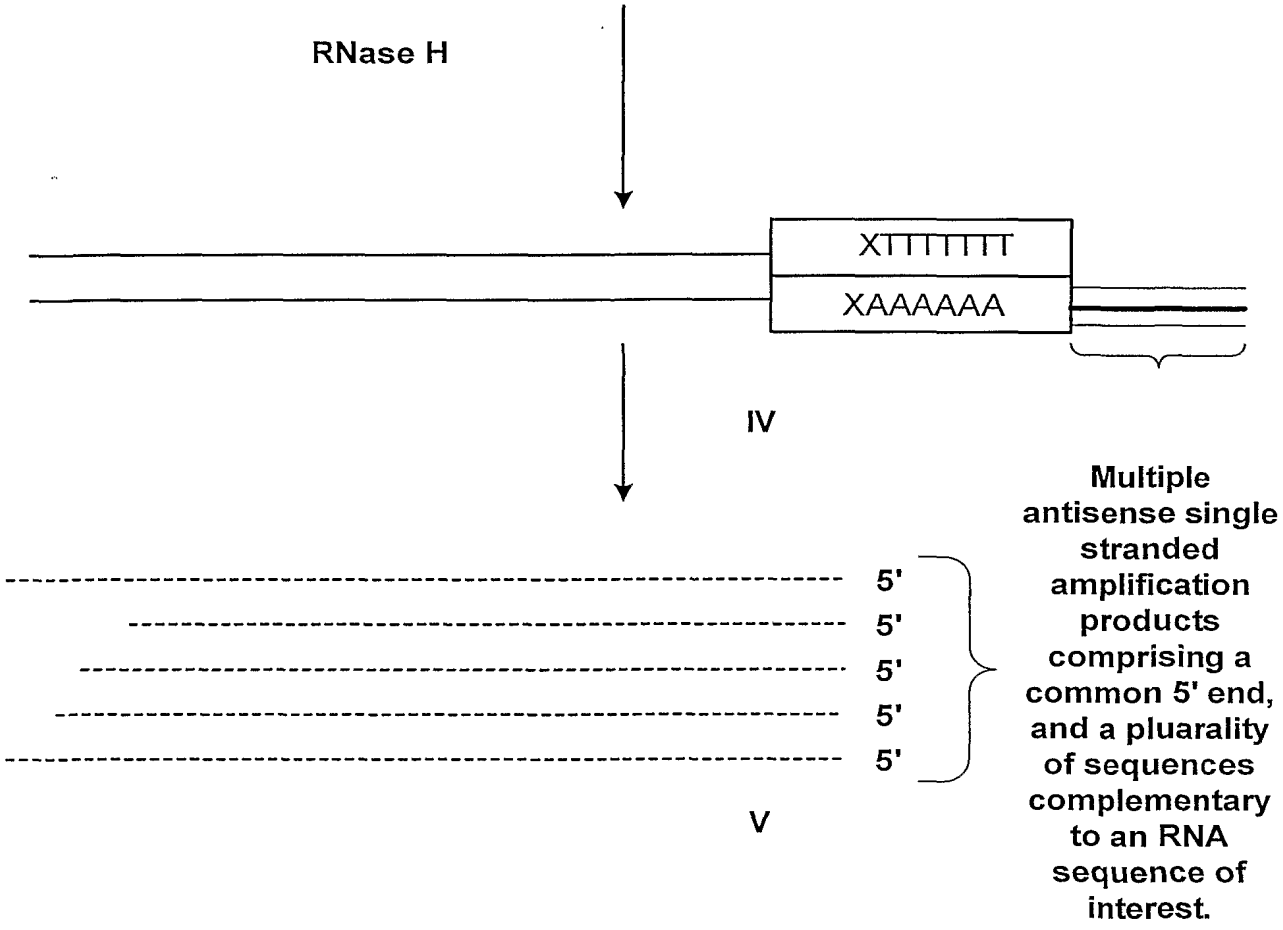


FIGURE 5B

IDT Exhibit 1020.9385

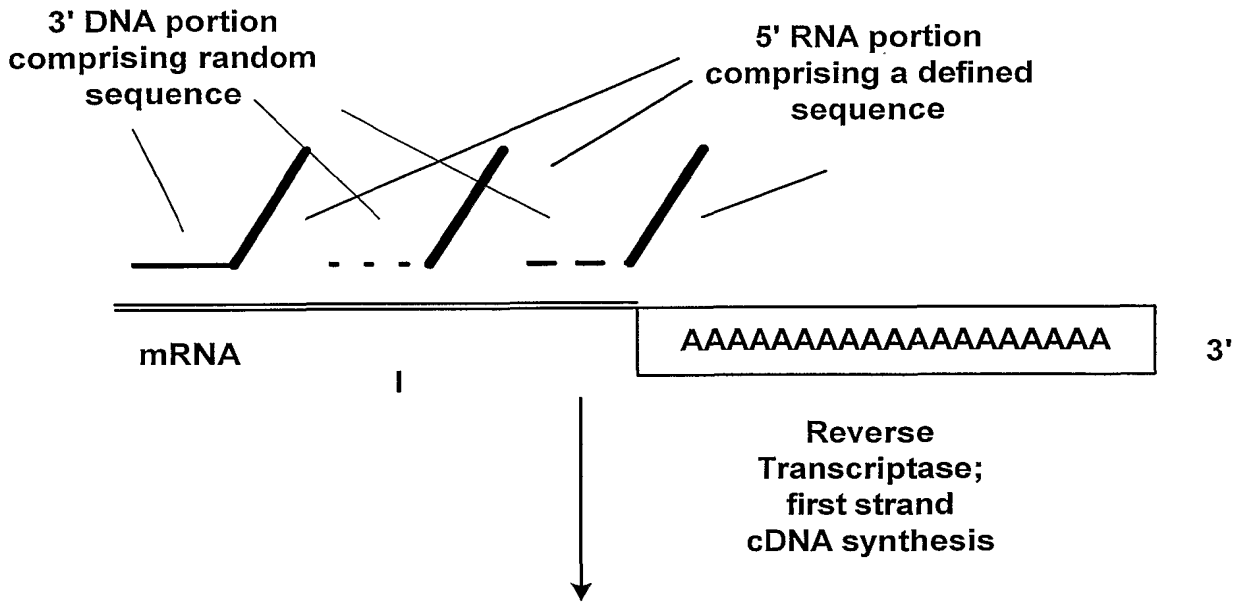


FIGURE 6A

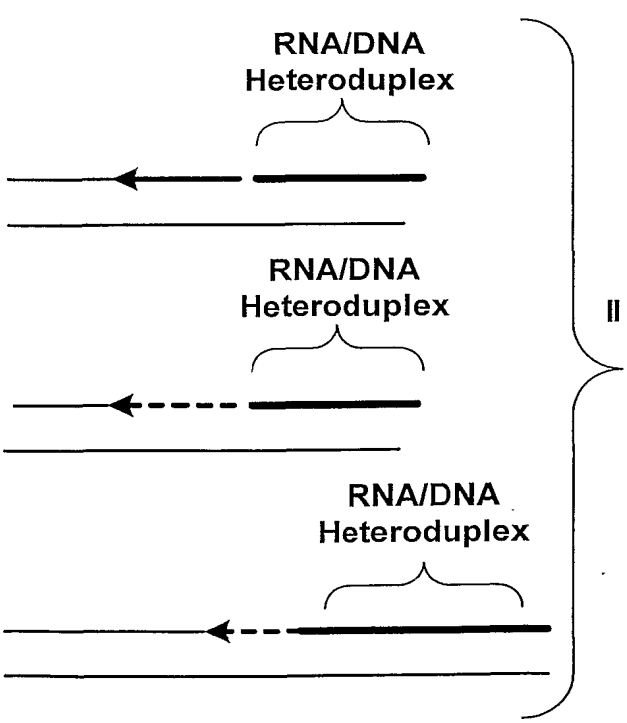


FIGURE 6B

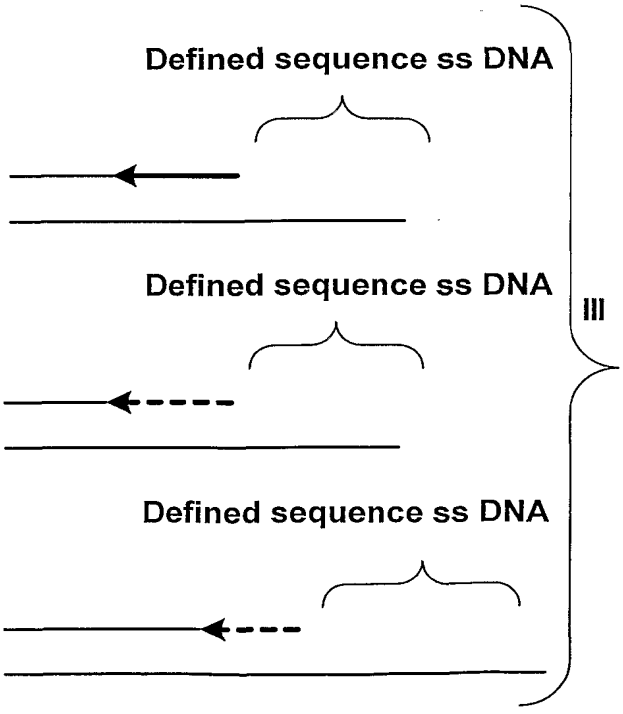


FIGURE 6C

IDT Exhibit 1020.9386

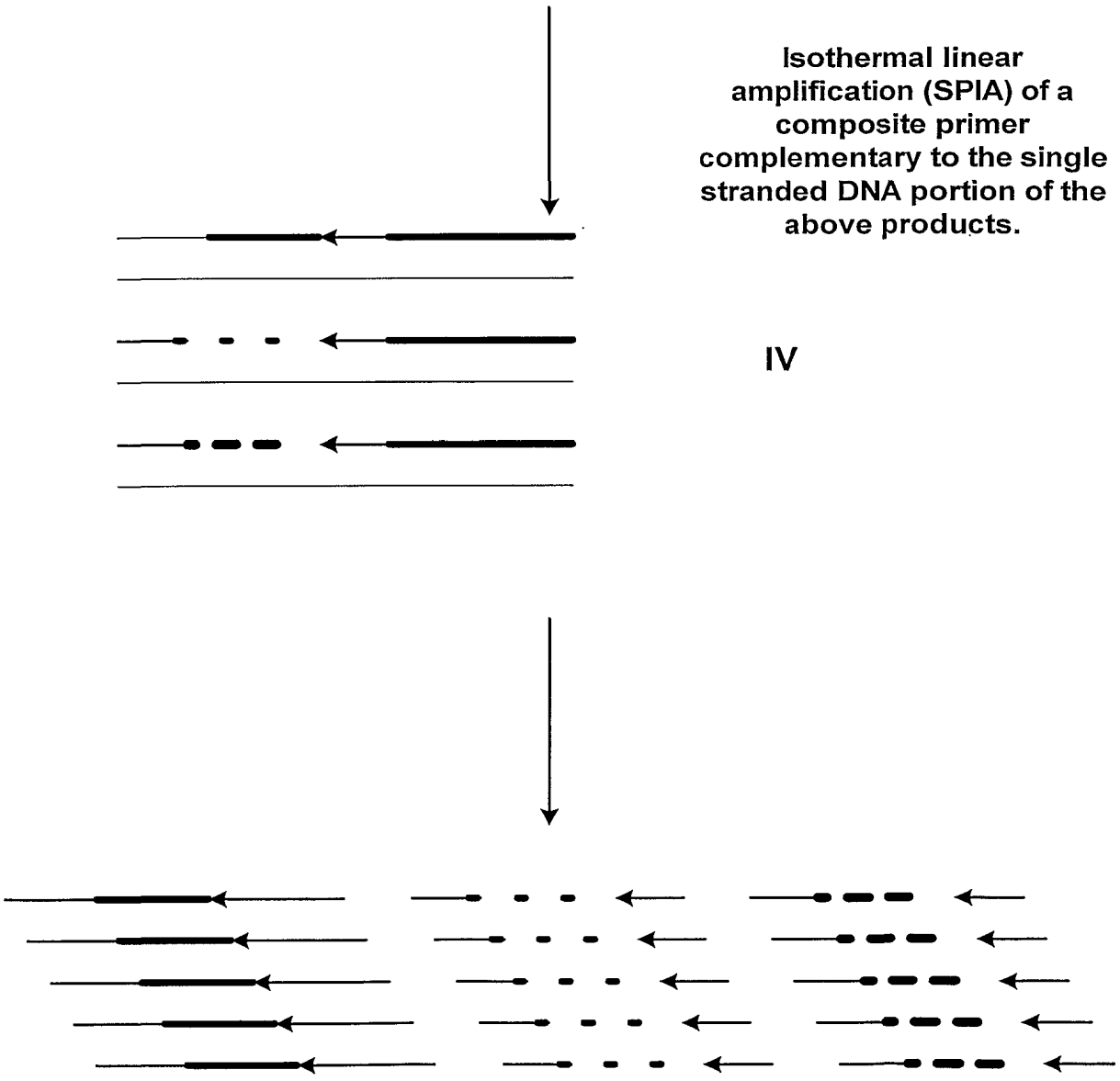


FIGURE 6D

IDT Exhibit 1020.9387

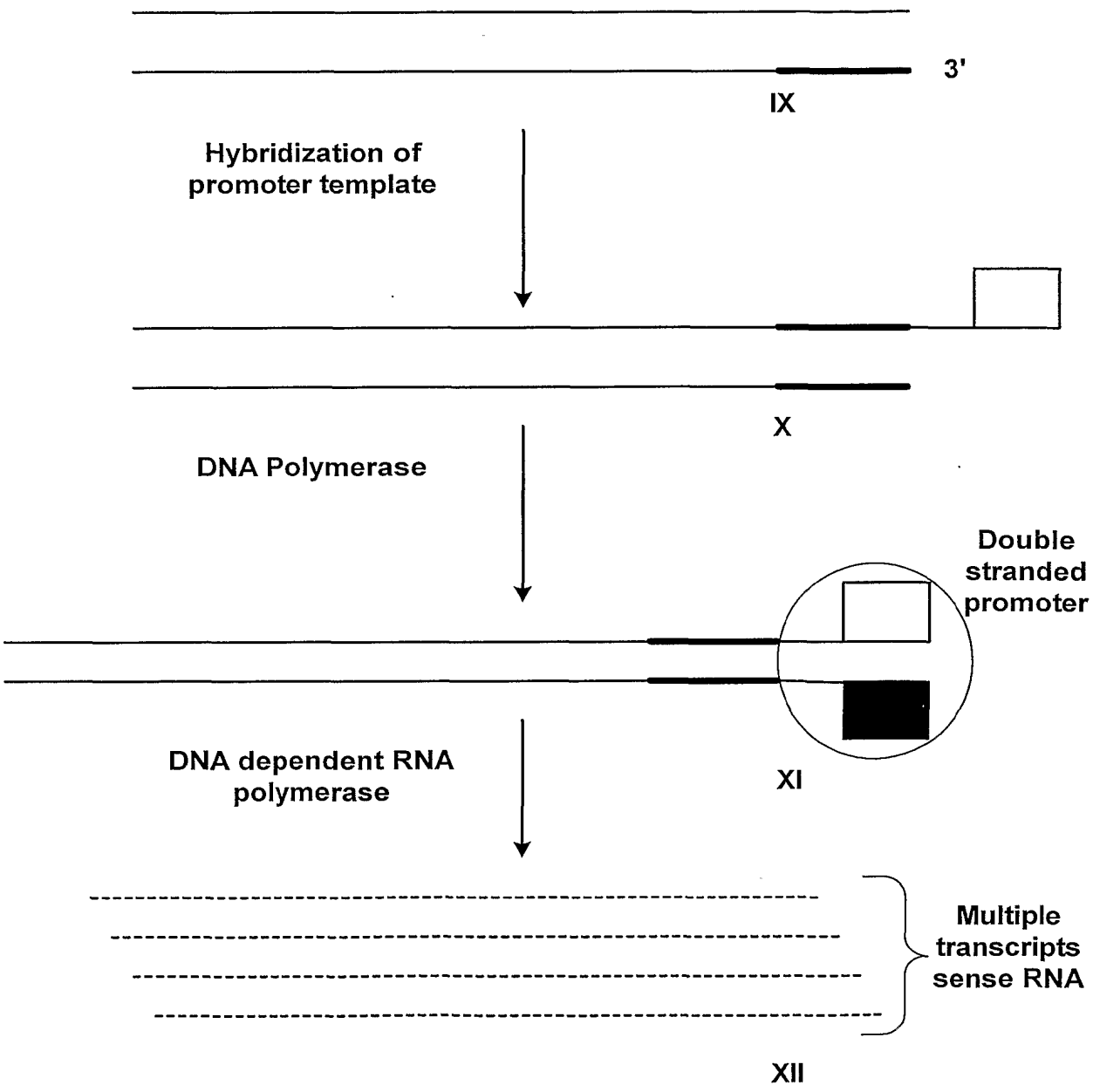


FIGURE 7

IDT Exhibit 1020.9388

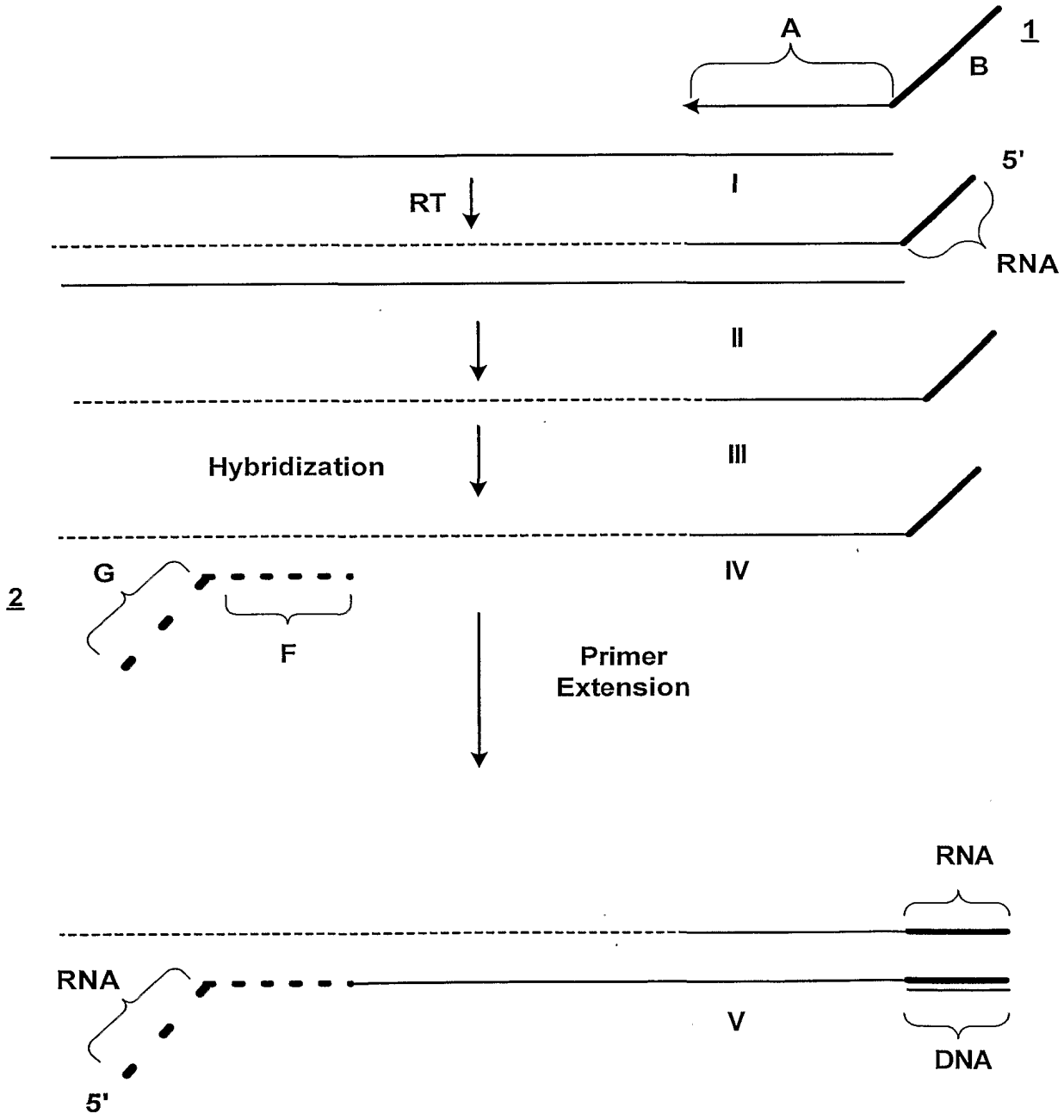


FIGURE 8A

IDT Exhibit 1020.9389

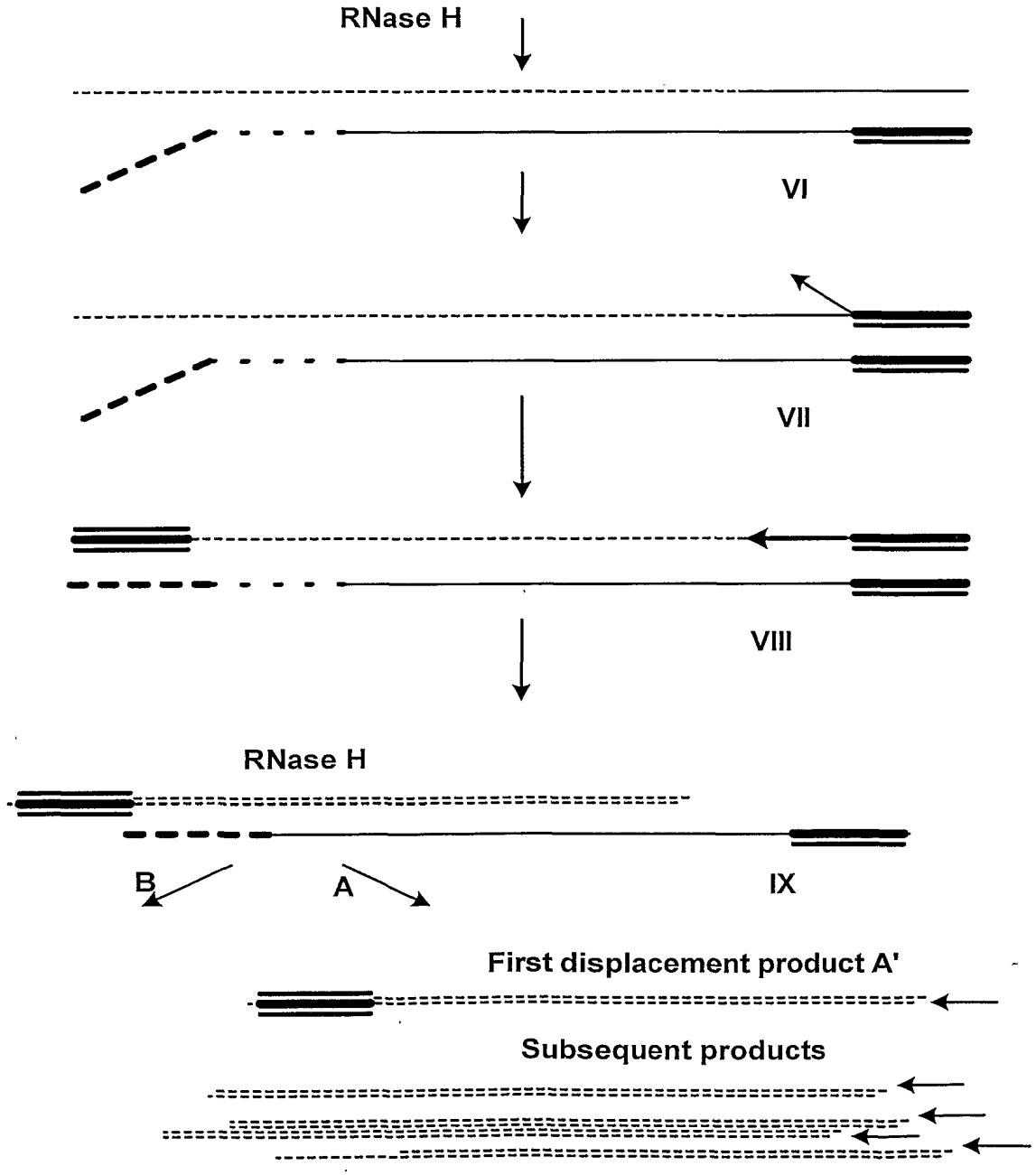


FIGURE 8B

IDT Exhibit 1020.9390

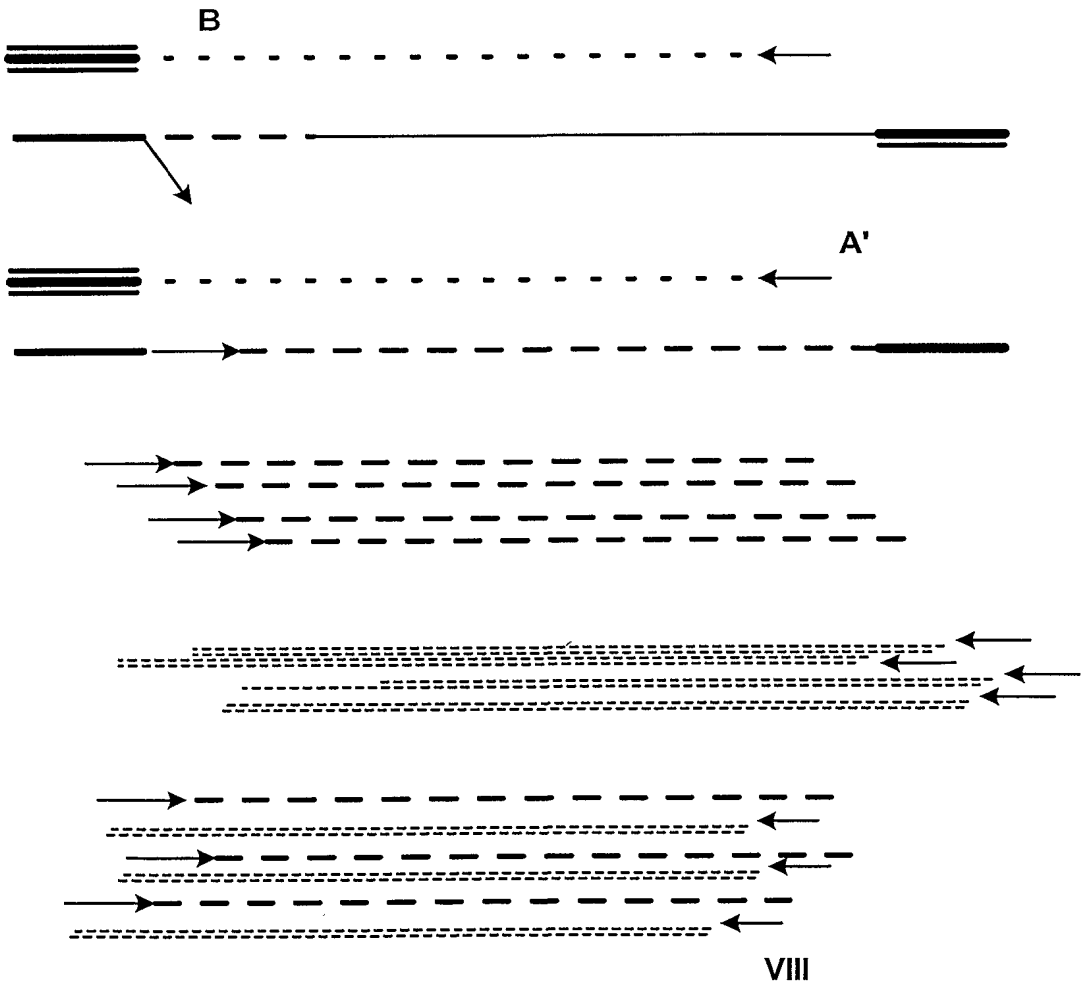
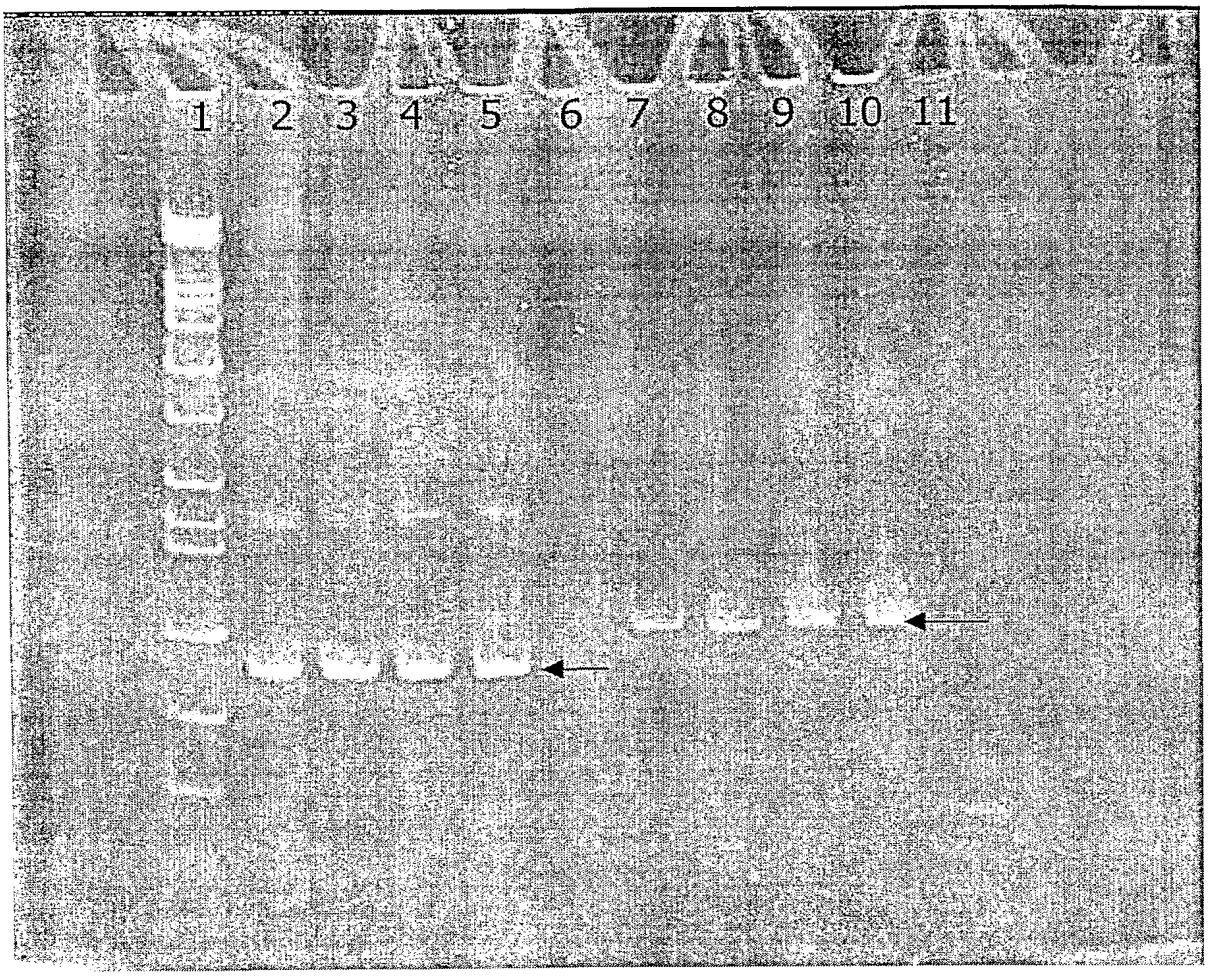


FIGURE 8C

IDT Exhibit 1020.9391

FIGURE 9



IDT Exhibit 1020.9392

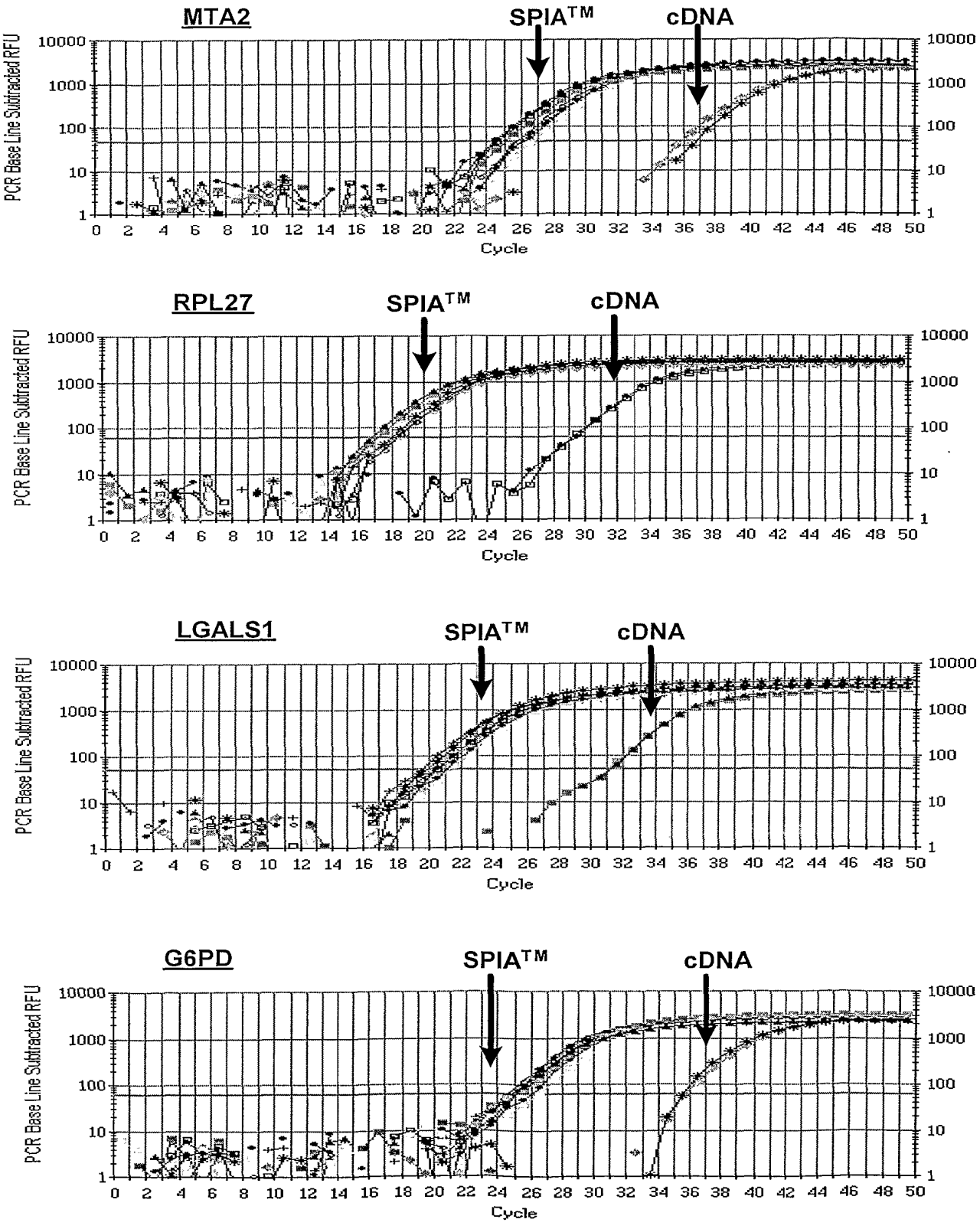


FIGURE 10

IDT Exhibit 1020.9393

Espacenet

Bibliographic data: WO0157248 (A2) — 2001-08-09

POLYNUCLEOTIDE ARRAYS AND THEIR USE IN SEQUENCING

Inventor(s): BALASUBRAMANISN SHANKAR; BENTLEY DAVID ± (BALASUBRAMANISN, SHANKAR, ; BENTLEY, DAVID, ; BALASUBRAMANIAN, SHANKAR)

Applicant(s): SOLEXA LTD [GB] ± (SOLEXA LTD)

Classification: **- international:** ***C12M1/00; C12N15/09; C12Q1/68; G01N33/53; G01N33/566; G01N37/00;*** (IPC1-7): C12Q1/68

- cooperative: **B82Y30/00**; **C12Q1/6837**; B01J2219/00497; B01J2219/00527; B01J2219/00576; B01J2219/00675; B01J2219/00707; B01J2219/00722; B01J2219/00729 more

Application number: WO2001GB00407 20010131 Global Dossier

Priority number (s): GB20000002310 20000201

Also published as: WO0157248 (A3) AU3038401 (A) CA2398168 (A1) EP1252332 (A2) IS6489 (A) more

Abstract of WO0157248 (A2)

A device for use in polynucleotide sequencing procedures comprises an array of polynucleotide molecules capable of interrogation and immobilised on a solid surface, wherein the array has a density which allows the molecules to be individually resolved. The molecules on the array comprise a target polynucleotide and a primer sequence, the primer being maintained in spatial relationship to the target, both being linked via a covalent bond. The device may be used in genotyping experiments, in which an array of genomic DNA is prepared to allow multiple sequencing reactions to be performed.

IDT Exhibit 1020.9394

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
9 August 2001 (09.08.2001)

PCT

(10) International Publication Number
WO 01/57248 A2

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

(21) International Application Number: PCT/GB01/00407

(22) International Filing Date: 31 January 2001 (31.01.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0002310.1 1 February 2000 (01.02.2000) GB

(71) Applicant: **SOLEXA LTD.** [GB/GB]; Chesterford Research Park, Little Chesterford, Saffron Walden, Essex CB10 1XL (GB).

(72) Inventors: **BALASUBRAMANISN, Shankar**; University of Cambridge, Department of Chemistry, Lensfield Road, Cambridge CB2 1EW (GB). **BENTLEY, David**; Sanger Centre, The Wellcome Trust Genome, Campus, Hinxton, Cambridge CB10 1SA (GB).

(74) Agent: **GILL JENNINGS & EVERY**; Broadgate House, 7 Eldon Street, London EC2M 7LH (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, 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, MZ, 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 01/57248 A2

(54) Title: POLYNUCLEOTIDE ARRAYS AND THEIR USE IN SEQUENCING

(57) Abstract: A device for use in polynucleotide sequencing procedures comprises an array of polynucleotide molecules capable of interrogation and immobilised on a solid surface, wherein the array has a density which allows the molecules to be individually resolved. The molecules on the array comprise a target polynucleotide and a primer sequence, the primer being maintained in spatial relationship to the target, both being linked via a covalent bond. The device may be used in genotyping experiments, in which an array of genomic DNA is prepared to allow multiple sequencing reactions to be performed.

IDT Exhibit 1020.9395

POLYNUCLEOTIDE ARRAYS AND THEIR USE IN SEQUENCING

Field of the Invention

This invention relates to devices having an array of polynucleotides, and to the sequencing of polynucleotides. In particular, this invention discloses methods for determining the sequence of arrayed polynucleotides.

Background to the Invention

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

An example of the technologies that have improved the study of nucleic acids, is the development of fabricated arrays of immobilised nucleic acids. These arrays typically consist of a high-density matrix of polynucleotides immobilised onto a solid support material. Fodor *et al*, Trends in Biotechnology (1994) 12:19-26, describes ways of assembling the nucleic acids using a chemically sensitized glass surface protected by a mask, but exposed at defined areas to allow attachment of suitably modified nucleotide phosphoramidites. Fabricated arrays may also be manufactured by the technique of "spotting" known polynucleotides onto a solid support at predetermined positions (e.g. Stimpson *et al* PNAS (1995) 92:6379-6383).

A further development in array technology is the attachment of the polynucleotides to the solid support material to form single molecule arrays. Arrays of this type are disclosed in WO-A-00/06770. The advantage of these arrays is that reactions can be monitored at the single molecule level and information on large numbers of single molecules can be collated from a single reaction.

Although these arrays offer particular advantages in sequencing experiments, the preparation of arrays at the single molecule level may be more difficult than at the multi-molecule level, where losses of target polynucleotide can be tolerated due to the multiplicity of the array. There is therefore a need

for improvements in the preparation of the single molecule arrays for sequencing procedures.

Furthermore, for DNA arrays to be useful their sequences must be determined. US-A- 5302509 discloses a method to sequence polynucleotides immobilised on a solid support. The method relies on the incorporation of fluorescently-labeled, 3'-blocked bases A, G, C and T to the immobilised polynucleotide, in the presence of DNA polymerase. The polymerase incorporates a base complementary to the target polynucleotide, but is prevented from further addition by the 3'-blocking group. The label of the incorporated base can then be determined and the blocking group removed by chemical cleavage to allow further polymerisation to occur.

Sequencing polynucleotides on a solid support can be difficult as the polynucleotide is typically bound to the solid support via a hybridisation reaction with a support-bound complement. Occasionally the conditions used in the sequencing protocol result in disruption to the bonds formed on hybridisation and the target polynucleotide is removed from the array.

Canard *et al.*, Gene (1994) 148:1-6, discloses a DNA sequencing procedure using a relatively harsh alkali denaturation step to remove fluorescent labels from nucleotides incorporated onto a template. The alkali treatment results in melting of the primer-template duplex, and Canard suggests providing a hairpin-like primer to the template to overcome this problem. The hairpin-like primer is suggested only in the context of the harsh alkali treatment with a homogeneous sample of template DNA.

WO-A-98/20019 discloses compositions and methods for the preparation of nucleic acid arrays. The general disclosure relates to the preparation of high density multi-molecule arrays, achieved by immobilising polynucleotides on microscopic beads attached to a solid support. Many different uses are proposed for the arrays.

WO-A-97/08183 relates to nucleic acid capture molecules. Hairpin loop structures are disclosed, but only in the context of capture molecules for use in hybridisation-based nucleic acid detection methods.

WO-A-97/04131 discloses primer amplification experiments carried out with polynucleotide hairpin structures. The general teaching is of methods for amplification of the hairpin structures, and it is clear that the structures contain a polynucleotide sequence to which a separate primer sequence is hybridised. The primer associates with the hairpin only through the hybridisation reaction.

Summary of the Invention

The present invention is based on the realisation that sample preparation and sequence analysis procedures can be improved if the target polynucleotide is maintained in spatial relationship to the primer, and the target and primer are linked via a bond stronger than that of hybridisation.

According to a first aspect of the invention, a device comprises an array of polynucleotide molecules immobilised on a solid support, wherein each molecule comprises a polynucleotide duplex linked via a covalent bond to form a hairpin loop structure, one end of which comprises a target polynucleotide, and the array has a surface density which allows the target polynucleotides to be individually resolved.

According to a further aspect of the invention, the device of the invention is used in a procedure to determine the sequence of the target polynucleotide.

According to a third aspect, a method for the preparation of the device comprises ligating a target polynucleotide to the 5'-end of a first molecule capable of forming a duplex as defined above, and immobilising the first molecule to the solid surface either before or after ligation.

The present invention provides a method for capturing target polynucleotides from solution and immobilising them on an array to form a stable primer-template complex. The target polynucleotide is maintained on the array together with the primer by bonds stronger than those formed in hybridisation, and so conditions which would otherwise result in the target or primer being dissociated from the array, may be used. This may be of particular use in genotyping experiments, where an array of genomic DNA is prepared, to allow multiple sequencing reactions to be performed.

Description of the Invention

The present invention is the formation of arrays of polynucleotides, the polynucleotides comprising a hairpin loop structure, one end of which comprises a target polynucleotide, the other end comprising a relatively short polynucleotide capable of acting as a primer in the polymerase reaction.

The arrays are typically high density arrays, and there are preferably at least 10^3 molecules/cm^2, more preferably at least 10^5 molecules/cm^2 and most preferably 10^6-10^9 molecules/cm^2. The molecules immobilised on the solid support surface are also preferably at a density so that they can be considered to be single molecules, i.e. each can be individually resolved. The term "individually resolved" is used herein to specify that, when visualised, it is possible to distinguish one molecule on the array from its neighbouring molecules. Separation between individual molecules on the array will be determined, in part, by the particular technique used to resolve the individual molecules. It will usually be the target polynucleotide portion that is individually resolved, as it is this which will be interrogated, e.g. by the incorporation of detectable bases.

Apparatus used to image arrays are known to those skilled in the art. For example, a confocal scanning microscope may be used to scan the surface of the array with a laser, to image directly a fluorophore incorporated on the individual arrayed molecule by fluorescence. Alternatively, a sensitive 2-D detector, such as a change-coupled detector, can be used to provide a 2-D image representing the individual molecules on the array. Resolving single molecules on the array with a 2-D detector is possible if adjacent molecules are separated by a distance of approximately at least 250nm, preferably at least 300nm and more preferably by at least 350nm.

The arrayed polynucleotide molecules each comprise a polynucleotide duplex which is used to retain a primer and a target polynucleotide in spatial relationship to each other. This ensures that the primer is able to perform its priming function during a polymerase-based sequencing procedure, and is not removed during any washing step in the procedure. The target polynucleotide

is capable of being interrogated. The term "capable of interrogation" refers to the ability of the target to act as a template during the polymerase reaction. The target polynucleotide is therefore interrogated by the incorporation of bases and the polymerase enzyme.

The duplex forms a hairpin loop structure, and the primer and target polynucleotides are comprised at the respective ends of the hairpin loop structure. The term "hairpin loop structure" refers to a molecular stem and loop formed from the hybridisation of complementary polynucleotides that are covalently linked at one end. The stem comprises the hybridised polynucleotides and the loop is the region that links the two complementary polynucleotides. Anything from a 10 to 20 (or more) base pair double-stranded (duplex) region may be used to form the stem. In one embodiment, the structure may be formed from a single-stranded polynucleotide having complementary regions. The loop in this embodiment may be anything from 2 or more non-hybridised nucleotides. In a second embodiment, the structure is formed from two separate polynucleotides with complementary regions, the two polynucleotides being linked (and the loop being formed) by a linker moiety. The linker moiety forms a covalent attachment between the ends of the two polynucleotides. Linker moieties suitable for use in this embodiment will be apparent to the skilled person. For example, the linker moiety may be polyethylene glycol (PEG).

The term 'polynucleotide' is intended to refer to nucleic acids in general, including DNA, RNA and synthetic analogs, e.g. PNA. DNA is preferred.

Preparation of the devices to form the target polynucleotide on the array may be carried out by methods known to those skilled in the art. The target polynucleotide will usually be one to which a sequence determination is to be made.

There are many different ways of forming the hairpin structure to incorporate the target polynucleotide. However, a preferred method is to form a first molecule capable of forming a hairpin structure, and ligate the target polynucleotide to this. Ligation may be carried out either prior to or after immobilisation to the solid support. The resulting structure comprises the single-

stranded target polynucleotide at one end of the hairpin and a primer polynucleotide at the other end.

In one embodiment, the target polynucleotide is genomic DNA purified using conventional methods. The genomic DNA may be PCR-amplified or used directly to generate fragments of DNA using either restriction endonucleases, other suitable enzymes, a mechanical form of fragmentation or a non-enzymatic chemical fragmentation method. In the case of fragments generated by restriction endonucleases, hairpin structures bearing a complementary restriction site at the end of the first hairpin may be used, and selective ligation of one strand of the DNA sample fragments may be achieved by one of two methods.

Method 1 uses a first hairpin whose restriction site contains a phosphorylated 5' end. Using this method, it may be necessary to first de-phosphorylate the restriction-cleaved genomic or other DNA fragments prior to ligation such that only one sample strand is covalently ligated to the hairpin.

Method 2: in the design of the hairpin, a single (or more) base gap can be incorporated at the 3' end (the receded strand) such that upon ligation of the DNA fragments only one strand is covalently joined to the hairpin. The base gap can be formed by hybridising a further separate polynucleotide to the 5'-end of the first hairpin structure. On ligation, the DNA fragment has one strand joined to the 5'-end of the first hairpin, and the other strand joined to the 3'-end of the further polynucleotide. The further polynucleotide (and the other strand of the fragment) may then be removed by disrupting hybridisation.

In either case, the net result should be covalent ligation of only one strand of a DNA fragment of genomic or other DNA, to the hairpin. Such ligation reactions may be carried out in solution at optimised concentrations based on conventional ligation chemistry, for example, carried out by DNA ligases or non-enzymatic chemical ligation. Should the fragmented DNA be generated by random shearing of genomic DNA or polymerase, then the ends can be filled in with Klenow fragment to generate blunt-ended fragments which may be blunt-end-ligated onto blunt-ended hairpins. Alternatively, the blunt-ended DNA fragments may be ligated to oligonucleotide adapters which are designed to

allow compatible ligation with the sticky-end hairpins, in the manner described previously.

The hairpin-ligated DNA constructs may then be covalently attached to the surface of a solid support to generate a single molecule array (SMA), or ligation may follow attachment to form the array.

The arrays may then be used in procedures to determine the sequence of the target polynucleotide. In the case that the target fragments are generated via restriction digest of genomic DNA, the recognition sequence of the restriction or other nuclease enzyme will provide 4, 6, 8 bases or more of known sequence (dependent on the enzyme). Further sequencing of between 10 and 20 bases on the SMA should provide sufficient overall sequence information to place that stretch of DNA into unique context with a total human genome sequence, thus enabling the sequence information to be used for genotyping and more specifically single nucleotide polymorphism (SNP) scoring.

Simple calculations have suggested the following based on sequencing a 10^7 molecule SMA prepared from hairpin ligation: For an 8 base pair recognition sequence, a single restriction enzyme will generate approximately 10^6 ends of DNA. If a stretch of 13 bases is sequenced on the SMA (i.e. 13×10^6 bases), approximately 13,000 SNPs will be detected. One application of such a sample preparation and sequencing format would in general be for SNP discovery in pharmaco-genetic analysis. The approach is therefore suitable for forensic analysis or any other system which requires unambiguous identification of individuals to a level as low 10^3 SNPs.

It is of course possible to sequence the complete target polynucleotide.

The sequencing method that is used to characterise the bound target may be any known in the art that measures the sequential incorporation of bases onto an extending strand. A suitable technique is disclosed in US-A-5302509 requiring the monitoring of sequential incorporation of fluorescently-labeled bases onto a complement using the polymerase reaction. Alternatives will be apparent to the skilled person. Suitable reagents, including fluorescently-labeled nucleotides will be apparent to the skilled person.

Immobilisation of the polynucleotides to the solid support may be carried out by any method known in the art, provided that covalent attachment is achieved. The solid support may be any of the conventional supports used in "DNA chips", e.g. glass slides, ceramic, silicon or plastics materials. The support will typically be a flat planar surface, usually of approximately 1cm^2.

The single molecule array can be prepared by contacting a suitably prepared solid support with a dilute solution containing the polynucleotides to be arrayed. Techniques for preparing the arrays are detailed in WO-A-00/06770.

Example

This Example illustrates the preparation of single molecule arrays by direct covalent attachment of hairpin loop structures to glass.

A solution of 1% glycidoxypropyltrimethoxy-silane in 95% ethanol/5% water with 2 drops H_2SO_4 per 500 ml was stirred for 5 minutes at room temperature. Clean, dry Spectrosil-2000 slides (TSL, UK) were placed in the solution and the stirring stopped. After 1 hour the slides were removed, rinsed with ethanol, dried under N_2 and oven-cured for 30 min. at 100°C. These 'epoxide' modified slides were then treated with 1 μM of labelled DNA (5'-Cy3-CTGCTGAAGCGTCGGCAGGT-heg-aminodT-heg-ACCTGCCGACGCT-3') (SEQ ID NOS. 1 and 2) in 50 mM potassium phosphate buffer, pH 7.4 for 18 hours at room temperature and, prior to analysis, flushed with 50 mM potassium phosphate, 1 mM EDTA, pH 7.4. The coupling reactions were performed in sealed teflon blocks under a pre-mounted coverslip to prevent evaporation of the sample and allow direct imaging.

The DNA structure was designed as a self-priming template system with an internal amino group attached as an amino deoxy-thymidine held by two 18 atom hexaethylene glycol (heg) spacers, and was synthesised by conventional DNA synthesis techniques using phosphoramidite monomers.

For imaging, one slide was inverted so that the chamber coverslip contacted the objective lens of an inverted microscope (Nikon TE200) via an immersion oil interface. A 60° fused silica dispersion prism was coupled optically to the back of the slide through a thin film of glycerol. Laser light was directed at the prism such that at the glass/sample interface it subtends an angle

of approximately 68° to the normal of the slide and subsequently undergoes Total Internal Reflection (TIR). The critical angle for glass/water interface is 66°.

Fluorescence from single molecules of DNA-Cy3, produced by excitation with the surface-specific evanescent wave following TIR, was collected by the objective lens of the microscope and imaged onto an Intensified Charge Coupled Device (ICCD) camera (Pentamax, Princeton Instruments, NJ). The image was recorded using a 532nm excitation (frequency-doubled solid-state Nd:YAG, Antares, Coherent) with a 580nm fluorescence (580DF30, Omega Optics, USA) filter for Cy3. Images were recorded with an exposure time of 500ms at the maximum gain of 10 on the ICCD. Laser powers incident at the prism were 50mW at 532nm.

Single molecules were identified by single points of fluorescence with average intensities greater than 3x that of the background. Fluorescence from a single molecule was confined to a few pixels, typically a 3x3 matrix at 100x magnification, and had a narrow Gaussian-like intensity profile. Single molecule fluorescence was also characterised by a one-step photobleaching process in the time course of the intensity and was used to distinguish single molecules from pixel regions containing two or more molecules, which exhibit multi-step processes.

To count molecules, a threshold for fluorescence intensities was first set to exclude background noise. For a control sample the background was essentially the thermal noise of the ICCD measured to be 76 counts with a standard deviation of only 6 counts. A threshold was arbitrarily chosen as a linear combination of the background, the average counts over an image and the standard deviation over an image. In general, the latter two quantities provide a measure of the number of pixels and range of intensities above background. This method gave rise to threshold levels which were at least 12 standard deviations above the background with a probability of less than 1 in 144 pixels contributing from noise. By defining a single molecule fluorescent point as being at least a 2x2 matrix of pixels and no larger than a 7x7, the probability of a single background pixel contributing to the counting was eliminated and clusters were ignored.

In this manner, the surface density of single molecules of DNA-Cy3 was measured at approximately 500 per 100 μm x 100 μm image or 5×10^6 cm^2.

CLAIMS

1. A device comprising an array of polynucleotide molecules immobilised on a solid surface, wherein each molecule comprises a polynucleotide duplex linked via a covalent bond to form a hairpin loop structure, one end of which comprises a target polynucleotide, and the array has a surface density which allows the target polynucleotides to be individually resolved.

2. A device according to claim 1, wherein immobilisation to the solid surface is via covalent attachment.

3. A device according to claim1 or claim 2, wherein the array has a density of at least 10^3 molecules/cm^2.

4. A device according to any preceding claim, wherein the array has a density of at least 10^6 molecules/cm^2.

5. A device according to any preceding claim, wherein the spacing between adjacent target polynucleotides on the array is at least 100nm.

6. A device according to any preceding claim, wherein the polynucleotide molecules are of DNA.

7. Use of a device according to any of claims 1 to 6, in an analysis procedure to determine the sequence of the target polynucleotide.

8. Use according to claim 7, wherein the procedure is genotyping.

9. Use according to claim 7 or claim 8, wherein the target polynucleotides are individually resolved by optical microscopy.

10. A method for the preparation of a device according to any of claims 1 to 6, comprising ligating a target polynucleotide to the 5' end of a first molecule capable of forming a duplex as defined in claim 1, and immobilising the first molecule to the solid surface either before or after ligation.

11. A method according to claim 10, wherein the immobilising is after the ligation of the target polynucleotide.

12. A method according to claim 10 or claim 11, wherein the target polynucleotide is in the form of double-stranded DNA, ligation is between one strand of the DNA and the first molecule, and the other strand is removed after ligation.

13. A method according to claim 12, wherein a further polynucleotide is hybridised to the first molecule with a one or more base gap between the further polynucleotide and the 3'-end of the first molecule, ligation is between the double-stranded DNA and the 5'-end of the first molecule and the further polynucleotide and hybridisation is subsequently disrupted to remove the further polynucleotide to form the target polynucleotide.

14. A method according to any of claims 10 to 13, wherein the 5'-end of the first duplex is phosphorylated and the target polynucleotide is dephosphorylated prior to ligation.

SEQUENCE LISTING

<110> Solexa Ltd.

<120> Polynucleotide Arrays and Their Use in Sequencing

<130> REP06208WO

<140> (not yet known)
<141> 2001-01-30

<150> 0002310.1
<151> 2000-02-01

<160> 2

<170> PatentIn Ver. 2.1

<210> 1
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<220>
<221> misc_structure
<222> (1)..(20)
<223> N = cytosine with a fluorescent Cy3 group
attached. M = thymine with hexaethylene glycol
attached.

<400> 1
nctgctgaag cgtcggcagg m 21

<210> 2
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<220>

IDT Exhibit 1020.9408

<221> misc_structure
<222> (1)..(13)
<223> N = adenine with hexaethylene glycol attached.

<400> 2
ncctgccgac gct 13

IDT Exhibit 1020.9409

Espacenet

Bibliographic data: EP2272976 (A1) — 2011-01-12

Method for differentiation of polynucleotide strands

Inventor(s): SOLDATOV ALEKSEY [DE]; BORODINA TATIANA [DE]; LEHRACH HANS [DE] ± (SOLDATOV, ALEKSEY, ; BORODINA, TATIANA, ; LEHRACH, HANS)

Applicant(s): MAX PLANCK GESELLSCHAFT [DE] ± (MAX-PLANCK-GESELLSCHAFT ZUR FOERDERUNG DER WISSENSCHAFTEN E.V)

Classification: **- international:** ***C12Q1/68***
- cooperative: **C12Q1/6848**; **C12Q1/6853**; **C12Q1/6855**; **C12Q1/6869**; **C12Q1/6876** more

Application number: EP20090008808 20090706 Global Dossier

Priority number (s): EP20090008808 20090706

Also published as: EP2451973 (A1) EP2451973 (B1) US2012237943 (A1) US2015105299 (A1) US2016053315 (A1) more

Abstract of EP2272976 (A1)

Objective of the present invention is to provide a method for keeping of directional information in double-stranded DNA. We suggest to convert polynucleotide into a hybrid double-stranded DNA. One particular strand of this hybrid double-stranded DNA should be synthesised using at least one modified nucleotide. Thus, this particular strand would contain modified nucleotides along the whole length. Density of directional markers would not depend on the length of polynucleotides. Any internal fragments of the hybrid double-stranded DNA would have directional information. When it is necessary the modified strand may be easily degraded or separated from the other strand. It was found that such hybrid double-stranded DNA may be easily generated in a number of molecular biology tasks and may be used for molecular cloning, library preparation and strand separation.

IDT Exhibit 1020.9410

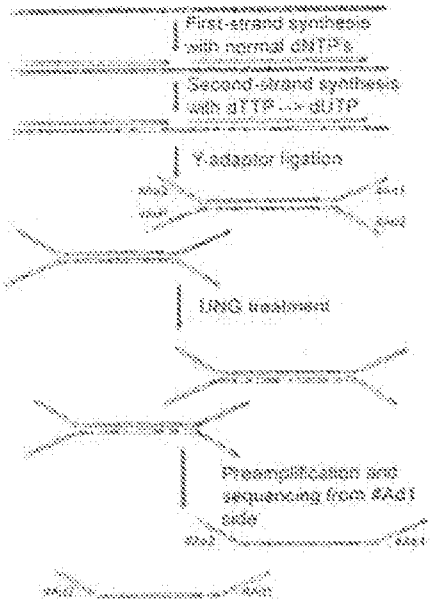


Figure 4

IDT Exhibit 1020.9411

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 2 272 976 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
12.01.2011 Bulletin 2011/02

(51) Int Cl.:
C12Q 1/68 *(2006.01)*

(21) Application number: **09008808.9**

(22) Date of filing: **06.07.2009**

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO SE SI SK SM TR
Designated Extension States:
AL BA RS

(71) Applicant: **Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V.**
80539 München (DE)

(72) Inventors:
- **Soldatov, Aleksey**
  14195 Berlin (DE)
- **Borodina, Tatiana**
  14195 Berlin (DE)
- **Lehrach, Hans**
  14129 Berlin (DE)

(74) Representative: **Arth, Hans-Lothar**
ABK Patent Attorneys
Jasminweg 9
14052 Berlin (DE)

(54) Method for differentiation of polynucleotide strands

(57) Objective of the present invention is to provide a method for keeping of directional information in double-stranded DNA. We suggest to convert polynucleotide into a hybrid double-stranded DNA. One particular strand of this hybrid double-stranded DNA should be synthesised using at least one modified nucleotide. Thus, this particular strand would contain modified nucleotides along the whole length. Density of directional markers would not depend on the length of polynucleotides. Any internal fragments of the hybrid double-stranded DNA would have directional information. When it is necessary the modified strand may be easily degraded or separated from the other strand. It was found that such hybrid double-stranded DNA may be easily generated in a number of molecular biology tasks and may be used for molecular cloning, library preparation and strand separation.

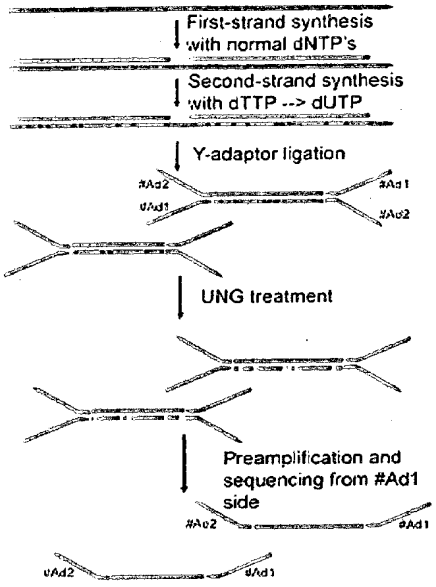


Figure 4

Printed by Jouve, 75001 PARIS (FR)

EP 2 272 976 A1

IDT Exhibit 1020.9412

Description

[0001] Cloning of single-stranded RNA and DNA molecules is much easier if they are transformed into double-stranded DNA. The invention relates to a novel method of preserving information about the direction of the original single-stranded molecules after being transformed into a double-stranded form.

[0002] Single-stranded RNA and DNA molecules play an important role in molecular biology. A number of viruses have single-stranded genomes. Most RNAs are present in single-stranded form. Recent studies have demonstrated an unexpected complexity of eukaryotic transcriptomes. A transcriptome is the set of all RNA transcripts produced in one or a population of cells. In addition to classical mRNAs (messenger ribonucleic acids) which cover approximately 1.5 percent of the genome in most higher eukaryotes, a large number of non-coding RNAs (ribonucleic acids) with widely varying expression levels have been identified. The biological functions of these novel transcripts are largely unknown and represent a new research area requiring high throughput transcriptome studies.

[0003] The polarity of the transcript is important for correct annotation of novel genes because it provides essential information about the possible function of a gene both on the RNA (structure and hybridization to other nucleic acid molecules) and the protein level. In addition, many genomic regions give rise to transcripts from both strands simultaneously. Antisense transcription is characteristic for eukaryotic genes and is thought to play an important regulatory role. Overlapping genes are common for compact genomes of prokaryotes and lower eukaryotes. Knowledge of transcript orientation helps to resolve the problem of colliding transcripts and to determine correctly gene expression levels in the presence of antisense transcription.

Background of the invention

[0004] From the art several methods are known that preserve directional information using distinguishable adapters for different ends of double-stranded DNA.

[0005] Adapters may be ligated directly to single-stranded RNA molecules.

[0006] Special composite primers may be used for first-strand cDNA synthesis: Herein the 5'- part of the primer provides a directional information, the 3'-part (oligo(dT)) or random nucleotides) is able to form a duplex with the RNA. Recently, first-strand cDNA synthesis was performed from a tagged random hexamer primer and second strand synthesis from a DNA-RNA template-switching primer.

[0007] Special homopolymeric adapters may be synthesised at the 3'-end of DNA molecules by Terminal Nucleotide Transferase.

[0008] DNA-RNA template-switching is used to attach a specific adapter to 3'-end of the DNA copy of the original single-stranded RNA molecule.

[0009] The main disadvantage of the adaptor-based approach it that only the ends of the clones preserve directional information. The problems are:

(i) strand-specific operations (directional immobilization on a solid phase; production of a single-stranded DNA for hybridization, etc.) should act on the adaptor sequence. For long clones the efficiency of such operations may be low. For example, strand separation where one of the strands contains a biotin molecule at the 5'-end is getting complicated if the length of the duplex is higher than 1 kb.
(ii) after fragmentation internal parts of the clones lose the directional information.

[0010] A special approach which is able to keep directional information in any part of the sequence was suggested recently: All cytidine residues in the RNA were converted to undines by bisulfite treatment prior to cDNA synthesis. The problem is that this approach is laborious and also leads to the loss of about 30% of uniquely matched sequencing reads because a part of the genome complexity is lost during the transformation from the four-bases into the three-bases code.

[0011] A lot of biologically important nucleic acids function in a single-stranded form where the polarity is extremely important. But for basic molecular cloning procedures single-stranded polynucleotides should be converted into double-stranded DNA molecules, because

* most of restriction enzymes work only on double-stranded DNA;
* ligation of double-stranded molecules is much simpler in comparison with the ligation of single-stranded molecules;
* the result of PCR is a double-stranded DNA.

[0012] At the moment most of the protocols keep directional information in the form of distinguishable adaptors on the ends of double-stranded molecules (Fig. 1B).

IDT Exhibit 1020.9413

Description of the invention

[0013] The objective of the present invention is to provide a method for keeping the directional information in double-stranded DNA. Surprisingly, it was found that this objective can be reached by the inventive method as disclosed by the description, the figures and the examples. A polynucleotide strand is converted into a hybrid double-stranded DNA. One particular strand of this hybrid double-stranded DNA should be synthesised using at least one modified nucleotide. Thus, this particular strand would contain modified nucleotides along the whole length of the strand. The density of directional markers would not depend on the length of hybrid double-stranded DNA. Any internal fragment of the hybrid would have directional information. If necessary, the modified strand may be easily degraded or separated from the other strand. It was found that such hybrid double-stranded DNA may be easily generated in a number of molecular biology tasks and may be used for molecular cloning, library preparation and strand separation.

[0014] The directional library construction method and the method of synthesis of single-stranded DNA according to the invention are easy to perform, inexpensive and highly reliable and thus overcome the shortcomings of the previously mentioned approaches.

[0015] The inventive method for preserving information about the direction of single-stranded nucleic acid molecules during molecular cloning operations with double-stranded DNA is provided. One of the strands of the double-stranded DNA is synthesized by using the other strand as a template and contains at least one modified dNTP. Thus, one strand of double-stranded DNA would contain unmodified nucleotides and the other strand at least one modified nucleotide along the whole length (Fig. 1C).

[0016] The new method of keeping of direction of polynucleotides is fully compatible with methods known from the art. Double-stranded DNA molecules may have simultaneously distinguishable primers and distinguishable strands (Fig. 1D). Fig. 2A shows how molecules with distinguishable strands may be converted into molecules with distinguishable primers. In an opposite operation the conversion of molecules with distinguishable primers into molecules with distinguishable strands is shown in Fig. 2B.

[0017] The main advantages compared to the adaptor-based method are:

(i) the density of modified nucleotides (directional markers) does not depend on the length of the double-stranded molecule. This is important for strand separation or strand-specific digestion of long DNA molecules.
(ii) any internal fragment of the clone carries directional information. After fragmentation internal parts of the clones do not lose the directional information.

[0018] Two examples demonstrate the convenience of the inventive method for differentiation of polynucleotide strands.

[0019] The method of synthesis of a single-stranded DNA is described in Fig. 3.

[0020] There is a number of methods of synthesis of single-stranded DNA relying on the difference between flanked adaptors:

* asymmetric PCR may be performed with the different amounts of PCR primers;
* the 5' end of one of the strands may contain thionucleotides which protect this strand from 5' to 3' exonucleases;
* protective thionucleotides may be introduced into the 3'-end part of only one strand to protect it from 3' to 5' exonucleases;
* the 5'-part of one of the strands may bear biotin for attachment of this strand to streptavidine-coated solid particles.

[0021] The problem is that all these methods proved to be unstable and show low efficiency for molecules larger than 1 kb.

[0022] Fig. 3 shows two possible ways for generating single-stranded DNA using the inventive method for differentiation of polynucleotide strands.

[0023] In a first approach (Fig. 3A) one strand of PCR product is substituted by a strand where dTTPs are substituted by dUTPs. After UNG treatment and APE 1 digestion (or alkalic hydrolisis) this strand is converted into a number of short single-stranded fragments which may be easily separated from the long unmodified strand by electrophoresis, anion-exchange chromatography, silica-based purification or gel-filtration.

[0024] In a second approach (Fig. 3B) one strand of the PCR product is substituted by a strand where a significant percentage of the dTTPs is substituted by biotin-associated dUTPs, respectively biotin-associated dTTPs. The concentration of biotin in a modified strand does not depend on the length of the double-stranded DNA, so the efficiency of the purification on streptavidine-coated particles should not depend on the length of the fragment.

[0025] The key steps of the method according to the invention include the generation of a cDNA strand including at least one modified nucleotide that in the course of the method is recognized by a selection factor which allows for a differentiation of the strand containing the at least one modified nucleotide from the strand containing the corresponding

IDT Exhibit 1020.9414

unmodified nucleotide or nucleotides.

[0026] The key steps of the inventive method are to be read in the following order:

I) Providing at least one polynucleotide template, primers, buffer solution, at least one enzyme with polymerase activity, sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or analogues thereof, wherein at least one of the four dNTPs is replaced by a corresponding modified dNTP used for strand-specific differentiation;
II) Performing specific strand synthesis on said at least one polynucleotide; and
III) Performing strand-specific differentiation by using a selection factor for differentiating strands containing the at least one modified nucleotide from strands containing the corresponding unmodified nucleotide.

[0027] If the polynucleotide template is a single-stranded RNA a second cDNA strand has to be synthesized. For this purpose the at least one modified dNTP has to be removed from the solution and has to be replaced by the corresponding unmodified dNTP or dNTPs or analogues thereof. By the use of all four unmodified dNTPs a second cDNA strand is generated.

[0028] In this case the key steps according to the method read as follows:

I) Providing at least one polynucleotide template, primers, buffer solution, at least one enzyme with polymerase activity, sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or analogues thereof, wherein at least one of the four dNTPs is replaced by a corresponding modified dNTP used for strand-specific differentiation;
II) Performing specific strand synthesis on said at least one polynucleotide;
III) Performing second-strand synthesis with all four unmodified dNTPs or analogues thereof after removing the at least one modified dNTP; and
IV) Performing strand-specific differentiation by using a selection factor for differentiating strands containing the at least one modified nucleotide from strands containing the corresponding unmodified nucleotide.

[0029] If the resulting cDNA shall undergo cloning, for example by an ensuing amplification step by PCR, it may be advantageous to carry out one or more preparative modifications on the strands to be cloned, such as shearing, restriction, digestion, end-polishing, A-tailing, G-tailing and adaptor ligation. These modifications may lead to a higher efficacy and unambiguous results in cloning.

[0030] With these additional modifications the key steps of the invention read as follows:

I) Providing at least one polynucleotide template, primers, buffer solution, at least one enzyme with polymerase activity, sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or analogues thereof, wherein at least one of the four dNTPs is replaced by a corresponding modified dNTP used for strand-specific differentiation;
II) Performing specific strand synthesis on said at least one polynucleotide;
III) Performing one or more molecular cloning operations on the double-stranded DNA, wherein the molecular cloning operations are selected from shearing, restriction, digestion, end-polishing, A-tailing, G-tailing and adaptor ligation; and
IV) Performing strand-specific differentiation by using a selection factor for differentiating strands containing the at least one modified nucleotide from strands containing the corresponding unmodified nucleotide.

[0031] As said before, if the polynucleotide template is a single-stranded RNA a second-strand cDNA has to be synthesized after the exchange of the at least one modified dNTP by the corresponding unmodified dNTP or dNTPs. If this double-stranded cDNA shall undergo cloning it will be advantageous to perform one or more of the modifications for the molecular cloning operation as described before. By combining these conditions the key steps of the invention read as follows:

I) Providing at least one polynucleotide template, primers, buffer solution, at least one enzyme with polymerase activity, sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or analogues thereof, wherein at least one of the four dNTPs is replaced by a corresponding modified dNTP used for strand-specific differentiation;
II) Performing specific strand synthesis on said at least one polynucleotide;
III) Performing second-strand synthesis with all four unmodified dNTPs or analogues thereof after removing the at least one modified dNTP;
IV) Performing one or more molecular cloning operations on the double-stranded DNA, wherein the molecular cloning operations are selected from shearing, restriction, digestion, end-polishing, A-tailing, G-tailing and adaptor ligation; and
V) Performing strand-specific differentiation by using a selection factor for differentiating strands containing the at least one modified nucleotide from strands containing the corresponding unmodified nucleotide.

IDT Exhibit 1020.9415

[0032] The initially provided at least one polynucleotide template may be a RNA such as mRNA, tRNA, rRNA, hnRNA, snRNA, snoRNA, scaRNA, asRNA, micro RNA, siRNA, short transcripts, synthetic polynucleotides and semisynthetic polynucleotides as well as physiological or externally induced modifications thereof, wherein mRNA is preferred.

[0033] mRNAs are protein-coding RNA. tRNAs are transfer RNAs responsible for providing correct amino acids in peptide (protein) synthesis in ribosomes. rRNAs are ribosomal RNAs, being a constitutive element of ribosomes. hnRNAs (heterogeneous nuclear RNA) is a precursor mRNA in the nucleus of eukaryotic cells. It is often named also pre-mRNA or precursor mRNA. snRNA (small nuclear RNA) is responsible in the eukaryotic nucleus for the processing of hnRNAs (spliceosomes). snoRNA (small nucleolar RNA) occurs in the nucleolus of eukaryotic cells. scaRNA (small Cajal body-specific RNA) is strongly related to snoRNA and occurs in Cajal bodies. asRNA (antisense RNA) plays an important role in regulating gene expression. microRNAs are involved in the regulation of cellular processes such as apoptosis and cell death. siRNA (small interfering RNA) is processed from cellular RNA into fragments with a length of ca. 22 base pairs by the enzyme Dicer and integrated into the enzyme complex RISC (RNA-induced silencing complex). RISC can thus regulate the expression of DNA segments or RNAs by switching off the respective coding region through analogous base pairing (silencing). This method by also used in gene therapeutic approaches. Retroviruses use RNA as the carrier of their genetic information.

[0034] Short transcripts refer to small RNA fragments with up to 20-40 base pairs.

[0035] Alternatively, the initially provided at least one polynucleotide template may be a DNA such as siDNA, DNA copies from retroviral or hepadnaviral RNA, single-stranded DNA, synthetic polynucleotides and semisynthetic polynucleotides as well as physiological or externally induced modifications thereof, wherein single-stranded DNA is preferred.

[0036] It shall be pointed out that in this invention the term *single-stranded DNA* can also refer to the first-strand cDNA derived from a RNA as template.

[0037] siDNA (small interfering DNA) is the corresponding DNA of siRNAs and mainly fulfils the same tasks in the cell. Single-stranded DNA is a single-stranded type of satellite DNA, occurring among others in certain soil-dwelling bacteria species.

[0038] According to the invention also double-stranded DNA can be the polynucleotide template. Specific primers and optionally adjuvants are needed to allow the generation of a cDNA strand against a particular strand of the double-stranded DNA (see below and Fig. 2B). Herein, the double-stranded DNA is destabilized in the region of hybridization of said specific primer for allowing annealing of the specific primer. Such a destabilization can be achieved for example by Uracil-N-Glycosylase digestion of deoxyuridines.

[0039] In a specific embodiment the double-stranded DNA template is a PCR product.

[0040] Sources for RNAs and DNAs can be eukaryotes (animals including humans and other mammals, plants, fungi), prokaryotes (bacteria, archaea), viruses, synthetic or semisynthetic constructs wherein these constructs or semisynthetic constructs may mirror at least partially a naturally occurring nucleic acids or may have a widely artificial sequence.

[0041] Said RNAs and DNAs can be isolated by well-known methods (Molecular Cloning: A Laboratory Manual, Third Edition. J. Sambrook, D. Russell).

[0042] The term *classic dNTPs,* respectively *unmodified dNTPS* refers to dATP (deoxyadenosine triphosphate), dCTP (deoxycytidine triphosphate), dGTP (deoxyguanosine triphosphate) and dTTP (deoxythymidine triphosphate). These four dNTPs (deoxynucleoside triphosphates) are usually used in nature for the generation of DNA strands.

[0043] It shall be pointed out that the plural form *classic dNTPs* refers to the total of dATP, dCTP, dGTP and dTTP. A referral in a quantifying sense to one or more of these classic dNTPs is not intended.

[0044] The term *analogues of dNTPs,* respectively *dNTPs and analogues thereof* refers to any molecule suitable for replacing a dNTP in presence and behavioural characteristics in a respective polynucleotide-containing solution and particularly in its matching characteristics to be able to undergo a base pair matching identical or similar to the classic dNTP it replaces. In particular, such analogues of dNTPs should be able to build the same or similar selective hydrogen bonds that allow for a selective pairing of base pairs (2 hydrogen bonds between A/T, T/A, A/U or U/A, and 3 hydrogen bonds between G/C and C/G), thus allowing for a correct translation of a polynucleotide sequence into its antisense sequence. Furthermore, it should be satirically similar to its corresponding classic dNTP so that it allows integration into the base pair sequence. Not only the nucleotide can be varied in such an analogue, but also the saccharide (usually a ribose or deoxyribose) or the phosphate backbone. For example, the phosphate backbone may be partially or entirely substituted by a sulphonated backbone. Also combinations of such variations are subsumed under this term.

[0045] The quantitative term *sufficient amount* herein refers to any amount, respectively concentration of an agent which allows for a qualitative and quantitative implementation of the method according to the invention in respect of the expected amounts which shall be needed for this purpose. Usually, the respective agents, in particular dNTPs, are added in exceeding amounts in order to ensure that all possible copies can be effected qualitatively and quantitatively.

[0046] The term *template* refers in molecular biology to a DNA or RNA strand which sets the base pair sequence to which a corresponding strand with an antisense sequence shall by polymerized. In a broader sense it may also refer to a RNA encoding a sequence for translation of this sequence into an amino acid sequence such as a peptide or a protein in ribosomes.

IDT Exhibit 1020.9416

[0047] The term *cDNA* stands for complementary deoxyribonucleic acid. It refers to a DNA which is generated as a copy of a RNA (usually by means of Reverse Transcriptase) or from a single DNA strand (usually by means of a DNA polymerase). Subsequently, a second cDNA strand may be generated by using the first cDNA strand as a template (usually by means of a DNA polymerase).

[0048] Enzymes with reverse transcriptase (RNA-dependent DNA polymerase) activity are able to transcribe a RNA sequence into a corresponding cDNA sequence. This group of enzymes includes among others HIV-1 reverse transcriptase from human immunodeficiency virus type 1 (PDB 1HMV), M-MLV reverse transcriptase from the Moloney murine leukemia virus, AMV reverse transcriptase from avian myeloblastosis virus and telomerase reverse transcriptase fulfilling a role in the maintenance of telomeres of eukaryotic chromosomes as well as commercially available forms such as "M-MLV reverse transcriptase RNase H Minus", "SuperScript® II", "SuperScript® III" or Tth polymerase. Among this group M-MLV reverse transcriptase RNase H Minus, SuperScript® II and SuperScript® III polymerases are preferred.

[0049] The terms *primers* and *random primers* refer to oligonucleotides which shall serve as starting point for DNA-replicating enzymes such as DNA polymerases. They can consist of a few base pairs, but in selected cases can reach up to 2.000 base pairs. Preferred are 5 to 50 base pairs. DNA polymerases need a hydroxy group for starting transcription. This hydroxy group is provided by the primer with its 3' hydroxy end. Often, the DNA sequence to be investigated is known. Hence suitable primers can be synthesized precisely according to the antisense sequence of aforesaid polynucleotide sequence. If this polynucleotide sequence is not or only incompletely known so-called random primers can be used. In such a method a large number of primers with a random base pair sequence are provided. One or more of these random primers statistically anneal to the polynucleotide sequence that shall be transcribed. Thus corresponding transcripts of different length can be obtained, depending where the random primer anneals to the polynucleotide sequence of interest. Alternatively, also composite primers with a random 3' part can be used. A further variation corresponding to the inventive method is the use of primers which are complementary to a homo-oligomeric tail synthesized by terminal transferase on the polynucleotide which serves as a template.

[0050] If the polynucleotide template is a RNA with a polyA tail it is preferred that said primers are oligo(dT), oligo(dU) or oligo(U) primers or composite primers with oligo(dT), oligo(dU) or oligo(U) region on 3' part. Alternatively, specific primers complementary to said single-stranded RNA template can be used.

[0051] If a double-stranded DNA template is used it is likewise possible to mark one particular strand of the double-stranded DNA with at least one modified nucleotide. Herein specific primers complementary to a particular strand of said double-stranded DNA can be used. In this case RecA protein may facilitate the hybridization of said primers to the double-stranded DNA template. Specific primers may appear as a result of introducing a nick in a particular strand of said double-stranded DNA template. Such nicks can for example result by the use of nicking endonuclease. Alternatively, such nicks may result from the digestion of the at least one modified nucleotide in a particular strand of said double-stranded DNA template.

[0052] In a specific embodiment the double-stranded DNA template is a PCR product. In this case, the recognition site of a nicking exonuclease or at least one modified nucleotide is located in one of the PCR primers. Herein it is preferred that said modified nucleotide is deoxyuridine.

[0053] If a double-stranded DNA template is used it is preferred that said polymerase is a DNA-dependent DNA polymerase with 5' to 3' exonuclease activity and said strand-specific synthesis on such a double-stranded DNA template is effected by means of a nick-translation.

[0054] In analogy, the polynucleotide template may be also a DNA::RNA hybrid. Hybrid herein means that a DNA strand is matching with a RNA strand having a corresponding antisense sequence. For example, but not limiting, it is possible that such a DNA::RNA hybrid is the product of a first-strand synthesis with unmodified nucleotides generated on a single-strand RNA as a template. For generating a cDNA of the DNA strand specific primers complementary to a particular DNA region of said double-stranded DNA::RNA hybrid have to be used. It is preferred that said primers are complementary to a homo-oligomeric tail synthesized by terminal transferase. In another preferred embodiment this specific primer is a template-switch primer. Such template-switch primers bind to the extra random nucleotides attached to the 3' end of the newly synthesized cDNA strand by using SuperScript® II enzyme (Invitrogen). This allows for full-length reverse transcription of the RNA strand. If the cDNA strand shall be generated against the RNA strand of the DNA::RNA hybrid said polymerase is DNA-dependent DNA polymerase with 5' to 3' exonuclease activity.

[0055] If the cDNA strand shall be generated against the DNA strand of the DNA::RNA hybrid the specific primers may appear as a result of introducing a nick into the RNA strand of said DNA::RNA hybrid. Such a nick may be set by using the RNA-specific enzyme RNase H. In a preferred embodiment RNase H, DNA polymerase I and E.coli ligase are used simultaneously. DNA ligases are enzymes able to ligate DNA strands by building an ester bond between the phosphate residue and the saccharide deoxyribose in the backbone of the DNA strands.

[0056] For second-strand cDNA synthesis using the first strand cDNA as a template an enzyme with DNA polymerase activity is needed. Also if the original polynucleotide is a single-stranded DNA and only one cDNA strand is generated such an enzyme with DNA polymerase activity is needed. DNA-dependent DNA polymerases are usually enzymes which generate a DNA copy strand from a DNA template. This group includes DNA polymerase I, II, III holoenzyme and

IDT Exhibit 1020.9417

IV, Taq polymerase, Tth polymerase, Pfu polymerase, wherein DNA polymerase I and Taq polymerase are preferred. This enzyme with polymerase activity does not have to be the same as used optionally for amplifying the remaining intact cDNA.

[0057] If the cDNA strand shall be generated against the RNA strand of the DNA::RNA hybrid the procedure is similar to that for generating a cDNA strand against one strand of a single-stranded RNA.

[0058] The term *modified dNTP* refers to the any molecule suitable for substituting exactly one corresponding classic dNTP in presence and behavioural characteristics in a respective polynucleotide-containing solution and particularly in its matching characteristics to be able to undergo a base pair matching identical or similar to the classic dNTP it replaces. Additionally, this modified dNTP must be suitable either for a specific degradation in which selectively this modified dNTP is degraded by a suitable degrading agent, thus rendering the DNA strand containing at least one of this degraded dNTP essentially unfit for amplification and/or hybridization, or it must render the DNA strand containing this modified dNTP eligible for specific removal of this DNA strand from the solution, or at least allowing a local and thus functional separation of polynucleotide strands. Such a removal respectively separation can be achieved by molecules or particles binding selectively to such exchange dNTPs, thus selectively only to one polynucleotide strand. Removal can occur by purification, adsorption to suitable structures or surfaces, binding to suitable selective structures such as antibodies, by selective degradation through certain chemicals, heat or radiation, or by centrifugation, magnetic or electrophoretic methods, isoelectric focusing. Also any kind of immobilization can be perceived.

[0059] Different purification methods may be used for removal of modified and unmodified dNTPs from the solution: electrophoretic separation, salt-alcohol precipitation, *anion*-exchange chromatography, silica purification, gel filtration. Purification on silica and gel-filtration are preferred.

[0060] Degradation respectively degradable herein means that a deoxynucleotide is specifically modified by addition of at least one enzymatic, chemical, heat, radiation (X-ray, UV, microwave) or optical methods such as lasers or combinations thereof in such a way that the resulting deoxynucleotide does not serve anymore for hybridization, cloning or as a suitable template in an amplification system such as PCR.

[0061] Analogously, the plural form *modified dNTPs* refers to the total of the respective modified dNTPs. A referral in a quantifying sense to one or more of these modified dNTPs is not intended.

[0062] The term *at least one modified dNTP* refers to the option that not only one dNTP is exchanged. According to the invention either one dNTP such as dATP, dCTP, dGTP or dTTP, or two dNTPs can be exchanged in the combinations dATP + dCTP, dATP + dGTP, dATP + dTTP, dCTP + dGTP, dCTP + dTTP or dGTP + dTTP, or three dNTPs can be exchanged in the combinations dATP + dCTP + dGTP, dATP + dCTP + dTTP, dATP + dGTP + dTTP or dCTP + dGTP + dTTP, or all four dNTPs are exchanged in the combination dATP + dCTP + dGTP + dTTP. It is understood that also analogues of these dNTPs as defined before shall be subsumed under the respective dNTPs. Also possible quantitative combinations thereof shall be subsumed under this combinational scheme.

[0063] There are several principles by which a selection factor can differentiate between a strand marked with at least one modified nucleotide and a strand with the corresponding unmodified nucleotide or nucleotides. For example, the at least one modified nucleotide has a chemically active group that can be modified after strand-specific synthesis. Another possibility is that the at least one nucleotide carries an affinity group and the selection factor binds selectively to this affinity group. Such am affinity group may for example be recognized by an antibody. A further possibility is that the at least one modified nucleotide carries a group which allows for selective recognition and binding or attachment by another agent which then enables selective removal of the such recognized strand. A further selective factor may be a different sensitivity to specific nucleases. Also the incorporation of ribonucleotides (NTPs) instead of deoxyribonucleotides (dNTPs) is suitable for enabling a differentiation between two strands.

[0064] The at least one modified nucleotide may display a different sensitivity to a nuclease, in comparison with the corresponding unmodified nucleotide or nucleotides. Thus such nuclease degrades specifically the at least one modified nucleotide while the unmodified nucleotides are not or only insignificantly affected by such nuclease action. Alternatively, such nuclease degrades specifically the unmodified nucleotides. In both cases the strand containing the such degraded nucleotide or nucleotides has become unfit for amplification or hybridization. Thus both strands can be differentiated.

[0065] In a preferred embodiment the differentiation of strands of polynucleotides is carried out with the modified dNTP being dUTP (deoxyuridine triphosphate). Herein the modified dNTP is dUTP (deoxyuridine triphosphate) as a substitute for dTTP. dUTP displays the same matching characteristics as dTTP, i.e. it is able to match with an opposed adenosine, but not with opposed guanosines, cytosines, thymidines or uridines. dUTP can be specifically degraded by the nuclease Uracil-N-Glycosylase (UNG). UNG is neutral towards dTTP and the other dNTPs or analogues thereof. This enzyme degrades uracil specifically by cleaving the N-glycosylic bond resulting in an abasic site. The such degraded DNA strand is unfit for undergoing appropriate amplification. At temperatures of over 50°C in slightly alkali conditions nicks occur in the DNA strand at the abasic positions. Additionally, Phusion DNA polymerase preferentially used for amplification (see below) is strongly suppressed by deoxyuridine-containing templates (Hogrefe, H.H., Hansen, C.J., Scott, B.R. and Nielsen K.B. 2002. Archaeal dUTPase enhances PCR amplifications with archaeal DNA polymerases by preventing dUTP incorporation. Proc Natl Acad Sci USA 99: 596-601). Consequently, only the DNA strand that does not contain

IDT Exhibit 1020.9418

dUTP can be amplified properly.

[0066] The combination of dUTP as exchange dNTP and UNG as the degrading agent of the exchange dNTP is a preferred embodiment. It is also preferred to replace only dTTP.

[0067] In another preferred embodiment the at least one modified nucleotide is thio_modified and said nuclease is an exonuclease that degrades thio_bonds with low efficiency. Such exonucleases sensitive to thio_bonds can be lambda-exonuclease, Exo III nuclease, T4 polymerase and T7 exonuclease.

[0068] Other preferred dNTP-specific degradation systems include the use of specific biotinylated dNTPs. Herein NHS esters of biotin are reacted with free amino groups of the respective nucleotide. These biotinylated nucleotides can be recognized by polymerase enzymes as eligible for integration into the actually formed strand. The matching properties to their counter nucleotide on the other strand are comparable. The biotinylated strand might be removed using streptavidin- or avidin-coated magnetic particles. Such specifically attached magnetic particles can be removed by applying a suitable magnetic field.

[0069] In another preferred embodiment the at least one modified polynucleotide is an Aminoallyl nucleotide (5-(3-aminoallyl)-nucleotide, aa-dNTP). Herein the nucleotide has an aminoallyl group attached to carbon 5 of the pyrimidine ring of uracil and cytosine. This aminoallyl group reacts with dyes from the cyanine series such as HiLyte Fluor® or Alexa Fluor® dyes that also been modified to be amine-reactive. As these dyes use to react also with standard buffer mediums a carbonate buffer is typically used. Herein it is preferred that the modified nucleotide is Aminoallyl-dUTP.

[0070] In a further preferred embodiment the modified nucleotide can be selectively recognized by an antibody. The antibody binds specifically to this modified nucleotide and the strands are differentiated by binding or immobilizing said antibody. It is further preferred that such antibody recognizes specifically the modified nucleotide 5-bromodeoxyuridine. It is also further preferred that such antibody or antibodies recognize specifically at least one modified nucleotide that has been modified by biotinylation.

[0071] If the strand carrying at least one modified nucleotide which carries a group which interacts which an affinity factor it is preferred that such strand is immobilized on a solid phase.

[0072] In a further preferred embodiment the at least one modified nucleotide is a ribonucleotide. Then it is preferred that said polymerase is a DNA polymerase displaying enhanced ability to incorporate modified substrates such as ribonucleotides (for example, Therminator™ II DNA Polymerase). The such incorporated at least one ribonucleotide is then specifically degraded by a specific nuclease, in this case a RNase.

[0073] If at least one classic dNTP is replaced by the corresponding modified dNTP(s) to generate a modified DNA strand the ratio of the at least one modified dNTP shall amount to at least 5% of the corresponding classic dNTP in the solution. Preferred are 20% ratio, more preferred 30%, more preferred 40%, more preferred 50%, more preferred 60%, more preferred 70%, more preferred 80% and particularly preferred more then 90%. Theoretically, only one degraded nucleotide is enough to render the degraded strand unfit for appropriate amplification. This requires that the DNA is long enough that at least one modified dNTP is integrated into the DNA strand. This is a matter of probability (when an even distribution of all four nucleotides is ideally assumed for the original polynucleotides) according to the distribution of the corresponding classic dNTPs and exchange dNTPs. With a length of at least 60 base pairs the probability is close to nil that not at least one modified dNTP has to be integrated into the second-strand cDNA. In order to ensure that also cDNAs with a length less than 60 base pairs always include at least one degradable nucleotide it is preferred that a sufficiently large percentage of the respective nucleoside triphosphate is modified.

[0074] If an amplification step is carried out in any of the embodiments of this invention an enzyme with polymerase activity is needed which is able to use the remaining intact cDNA strand as a template. In theory, any such enzyme with polymerase activity is suitable for this step. However, those enzymes that are used in conventional PCR protocols are preferred. They should grant a high reliability and a high throughput under conventional PCR conditions.

[0075] The term PCR refers to polymerase chain reaction. This is a general term comprising all protocols, conditions, automatized and half-automatized devices in which at least one DNA template is amplified by default by means of a enzyme with polymerase activity. In general, such molecules with polymerase activity are enzymes such as Taq DNA polymerase, Pfu DNA polymerase, Phusion DNA polymerase, wherein Taq and Phusion DNA Polymerases are_preferred.

[0076] It is particularly preferred that the at least one modified nucleotide is dUTP and said amplification is a Pfu DNA polymerase-dependent amplification.

[0077] If an amplification step is carried out in any of the embodiments of this invention a useful PCR (polymerase chain reaction) protocol shall be used by means of which the remaining intact DNA strand shall be amplified. Suitable PCR protocols of the include the known protocols for Standard PCR, Allele-Specific PCR, Alu PCR, Assembly PCR, Asymmetric PCR, Helicase Dependent Amplification HDA, Hot Start PCR, Inverse PCR, Intersequence-Specific PCR (ISSR PCR), Late-PCR, Ligation-mediated PCR, Long PCR, Methylation-Specific PCR, Miniprimer PCR, Multiplex PCR, Nested PCR, Overlap-Extension PCR, Quantitative PCR, Real Time PCR, Solid Phase PCR, Thermal Assymetric Interlaced PCR (TAIL-PCR), Touchdown PCR or combinations thereof. Preferred is the use of Standard PCR. Particularly preferred is the use of dUTP as the at least one modified nucleotide for amplification of the remaining intact strand with

IDT Exhibit 1020.9419

Standard PCR. Equally particularly preferred is the use of ribonucleotides as the least one modified nucleotide for amplification of the remaining intact strand with Standard PCR.

[0078] Herein the term "at least one intact DNA strand" means that the number of the intact DNA strands to be amplified can vary from one to many billions of different polynucleotides.

[0079] In a further optional step the resulting sequences can be aligned with already known sequence patterns of the corresponding DNA, if applicable. Such results can be plotted in a way such as shown in Fig. 6.

[0080] Optionally, the method according to the invention comprises the step of contig construction from the obtained sequences. Herein the sequences of different overlapping (contiguous) DNA fragments derived from the same source can be used for (re-)constructing the original respectively complete DNA sequence of this source.

[0081] Alternatively to the amplification step the at least one intact DNA strand can be hybridized to a nucleic acid structure displaying an antisense sequence of the at least one intact DNA strand. Suitable hybridization methods include *in situ* hybridization (hybridization of a DNA to a complementary sequence in isolated cells and/or tissues), Southern Blot (a method for verifying the presence of a DNA sequence in a DNA sample by separating the DNAs in the sample by size using gel electrophoresis and bringing said sample to a filter membrane for probe hybridization), Northern Blot (the equivalent of Southern Blot with RNA samples) and Southwestern Blot (identifying the presence of a DNA sequence in a DNA sample prepared according to the Southern blot method by hybridizing the sample on a nitrocellulose gel with characterizing DNA-binding proteins) microarray hybridization and correspondent "reverse" hybridizations were intact DNA's are loaded on a membrane and hybridised with particular probes.

[0082] When an amplification is intended the key steps of the inventive method are to be read in the following order:

I) Providing at least one polynucleotide template, primers, buffer solution, at least one enzyme with polymerase activity, sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or analogues thereof, wherein at least one of the four dNTPs is replaced by a corresponding modified dNTP used for strand-specific differentiation;
II) Performing specific strand synthesis on said at least one polynucleotide;
III) Performing one or more molecular cloning operations on the double-stranded DNA, wherein the molecular cloning operations are selected from shearing, restriction, digestion, end-polishing, A-tailing, G-tailing and adaptor ligation; and
IV) Performing strand-specific differentiation by using a selection factor for differentiating strands containing the at least one modified nucleotide from strands containing the corresponding unmodified nucleotide.
V) Amplifying the at least one remaining intact cDNA strand by performing PCR;

[0083] In the particularly preferred embodiment 1 of the method according to the invention a strand specific cloning of RNAs such as a whole transcriptome of a cell or a tissue shall be carried out. Therefore, at least one single strand RNA is provided, for example by isolation from the respective cell or tissue. The four dNTPs dATP, dCTP, dGTP and dTTP are added in sufficient amounts to the solution containing the at least one RNA strand. Then, a first cDNA strand is generated by using random primers (or oligo(dT), oligo (dU), oligo (U) primers) for annealing to said at least one RNA strand and an enzyme with Reverse Transcriptase activity. In the following at least one of the aforesaid dNTPs is replaced by at least one corresponding modified dNTP. For example, the modified dNTP may be a dNTP which can be specifically degraded by adding a degrading agent in order that the strand containing the such degraded at least one modified dNTP is not suitable anymore for amplification, or the modified dNTP may expose the such marked cDNA strand for being removed by interaction with specific agents interacting with the additional functional group or groups on the at least one modified dNTP. With the four dNTPs including the at least one modified dNTP second strand cDNA synthesis is performed by using the first cDNA strand as a template and a suitable enzyme with polymerase activity. In the next step polarity specific adaptors are ligated to the ends of cDNA strands, allowing a polymerase to dock to the adaptor structure for PCR amplification. After the ligation the second cDNA strand containing the at least one modified dNTP is for example degraded by degrading the at least one modified dNTP with a suitable degrading agent, or the at least one second cDNA strand containing the at least one modified dNTP is separated from the at least one first cDNA strand by special molecules or particles attaching to characteristic structures on the modified dNTP. Finally, the remaining intact first cDNA strand is amplified by PCR using a suitable enzyme with polymerase activity.

[0084] Thus the steps of this embodiment of the inventive method can be summarized in the following schematic order:

1) Providing at least one single-stranded RNA molecule;
2) Adding sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof;
3) Generating a first cDNA strand;
4) Replacing at least one dNTP with at least one corresponding modified dNTP;
5) Performing second-strand synthesis on the first cDNA strand;
6) Ligating polarity-specific adaptors to one or both ends of the cDNA double strand;
7) Specifically modifying or removing the second cDNA strand containing the at least one modified dNTP; and

IDT Exhibit 1020.9420

8) Amplifying the at least one remaining intact cDNA strand by performing PCR.

[0085] In embodiment 2 strand specific cloning of RNA can be achieved also by varying the steps of the inventive method. In this embodiment the first cDNA strand shall contain the at least one modified dNTP. In the replacement step this at least one modified dNTP is substituted by the corresponding classic dNTP. Thus in this embodiment the at least one first cDNA strand will be either degraded or removed. The amplification step is then performed on the remaining intact second cDNA strand. All other steps are adjusted correspondingly.

[0086] The steps of this embodiment of the inventive method can be summarized in the following schematic order:

1) Providing at least one single-stranded RNA molecule;
2) Adding sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof, wherein at least one of the four dNTPs is a modified dNTP;
3) Generating a first cDNA strand;
4) Replacing the at least one modified dNTP with the corresponding dNTPs or analogues of the corresponding dNTPs;
5) Performing second-strand synthesis on the first cDNA strand;
6) Ligating polarity-specific adaptors to one or both ends of the cDNA double strand;
7) Specifically modifying or removing the first cDNA strand containing the at least one modified dNTP; and
8) Amplifying the at least one remaining intact cDNA strand by performing PCR.

[0087] In embodiment 3 strand specific differentiation according to the invention can also be achieved using single-stranded DNA as a template. Here at least one DNA single strand is provided. A cDNA strand is synthesized wherein at least one modified dNTP is integrated. In this embodiment only one cDNA strand is synthesized since a second cDNA strand would be identical to the original DNA single strand. Finally, the original single-stranded DNA shall be amplified. Thus all transcripts resulting from the PCR will bear the same reading direction of the sequence as the original single-stranded DNA. The other steps are adjusted accordingly.

[0088] This leads to the following schematic order:

1) Providing at least one single-stranded DNA molecule;
2) Adding sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof, wherein at least one of the four dNTPs is a modified dNTP used for the strand specific differentiation;
3) Generating a cDNA strand;
4) Ligating polarity-specific adaptors to one or both ends of the resulting DNA double strand;
5) Specifically modifying or removing the cDNA strand containing the at least one modified dNTP; and
6) Amplifying the at least one intact DNA strand by performing PCR.

[0089] In embodiment 4 the purpose of the inventive method is not an amplification but a hybridization. This hybridization can occur against any structure carrying a corresponding antisense sequence to the at least one intact first cDNA strand (see below). For a hybridization without amplification no adaptors have to be ligated. Consequently, this step is omitted in this embodiment and the amplification step is replaced by a hybridization step. Thus the order of the steps of the inventive method reads as follows:

1) Providing at least one single-stranded RNA molecule;
2) Adding sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof;
3) Generating a first cDNA strand;
4) Replacing at least one dNTP with at least one corresponding modified dNTP used for the strand specific differentiation;
5) Performing second-strand synthesis on the first cDNA strand;
6) Specifically modifying or removing the second cDNA strand containing the at least one modified dNTP; and
7) Hybridizing the at least one remaining intact cDNA strand to a suitable antisense DNA sequence.

[0090] Embodiment 5 varies embodiment 4 in so far that herein the first cDNA strand contains the at least one modified dNTP. Thus the remaining intact second cDNA strand is hybridized. Consequently, the sequence against which the remaining intact second cDNA strand is hybridized must display an antisense sequence of the second strand. The steps of the inventive method are adjusted according to embodiments 2 and 4 and read as follows:

1) Providing at least one single-stranded RNA molecule;
2) Adding sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof, wherein at least one of the four dNTPs is a modified dNTP used for the strand specific differentiation;

IDT Exhibit 1020.9421

3) Generating a first cDNA strand;
4) Replacing the at least one modified dNTP with the corresponding dNTPs or analogues of the corresponding dNTPs;
5) Performing second-strand synthesis on the first cDNA strand;
6) Specifically modifying or removing the first cDNA strand containing the at least one modified dNTP; and
7) Hybridizing the at least one remaining intact cDNA strand to a suitable antisense DNA sequence.

[0091] In embodiment 6 the hybridization step follows after an amplification step. In various hybridization modes it may be advantageous to hybridize not only one or a few copies of the respective intact cDNA strand, but a large number as resulting from an amplification. Thus the steps of ligating a polarity-specific adaptor and of an amplification have to be integrated again. The order of the steps reads like embodiment 1 and a subsequent hybridization step:

1) Providing at least one single-stranded RNA molecule;
2) Adding sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof;
3) Generating a first cDNA strand;
4) Replacing at least one dNTP with at least one corresponding modified dNTP;
5) Performing second-strand synthesis on the first cDNA strand;
6) Ligating polarity-specific adaptors to one or both ends of the cDNA double strand;
7) Specifically modifying or removing the second cDNA strand containing the at least one modified dNTP; and
8) Amplifying the at least one remaining intact cDNA strand by performing PCR.
9) Hybridizing the amplified cDNA to a suitable antisense DNA sequence.

[0092] Correspondingly, in embodiment 7 the hybridization step can ensue after performing an amplification of the intact second-strand cDNA, as in embodiment 2:

1) Providing at least one single-stranded RNA molecule;
2) Adding sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof, wherein at least one of the four dNTPs is a modified dNTP;
3) Generating a first cDNA strand;
4) Replacing the at least one modified dNTP with the corresponding dNTPs or analogues of the corresponding dNTPs;
5) Performing second-strand synthesis on the first cDNA strand;
6) Ligating polarity-specific adaptors to one or both ends of the cDNA double strand;
7) Specifically modifying or removing the first cDNA strand containing the at least one modified dNTP; and
8) Amplifying the at least one remaining intact cDNA strand by performing PCR.
9) Hybridizing the amplified cDNA to a suitable antisense DNA sequence.

[0093] Also in the case of single-stranded DNA it may be advantageous to combine an amplification and a hybridization step. Thus in embodiment 8 the order of the steps reads as follows:

1) Providing at least one single-stranded DNA molecule;
2) Adding sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof, wherein at least one of the four dNTPs is a modified dNTP used for the strand specific differentiation;
3) Generating a cDNA strand;
4) Ligating polarity-specific adaptors to one or both ends of the resulting DNA double strand;
5) Specifically modifying or removing the cDNA strand containing the at least one modified dNTP; and
6) Amplifying the at least one intact DNA strand by performing PCR.
7) Hybridizing the amplified DNA to a suitable antisense DNA sequence.

[0094] Embodiment 9: If the sequence of the original RNA is not known it may be desirable to perform sequencing. Thus such a step is added to the step order of embodiment 1:

1) Providing at least one single-stranded RNA molecule;
2) Adding sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof;
3) Generating a first cDNA strand;
4) Replacing at least one dNTP with at least one corresponding modified dNTP;
5) Performing second-strand synthesis on the first cDNA strand;
6) Ligating polarity-specific adaptors to one or both ends of the cDNA double strand;
7) Specifically modifying or removing the second cDNA strand containing the at least one modified dNTP; and
8) Optionally, amplifying the at least one remaining intact cDNA strand by performing PCR.

IDT Exhibit 1020.9422