DNA – containing B.U.S.-aliquots and processed via flow cytometry. The optical signature of those beads that present positive signals from probe 6 define the genes that contain SNPs and mutations in the population.

In order to detect SNPs and mutations, the control DNA which is attached to the beads has to be relatively short (e.g. in the region 10-100 base pairs) so that, following capture of a corresponding fragment that contains an SNP/mutation, the genomic region containing the genetic alteration is automatically defined with a good resolution (~100 base pairs). In addition, for each gene that will be examined, enough regions of the gene have to be immobilized (each region on a separate B.U.S.), so that the whole gene (from 3' to 5' -end) is adequately represented at regular intervals. In this manner, every target DNA fragment which is isolated and contains a genetic alteration will be assured of finding a complementary sequence to hybridize in the B.U.S.-aliquot.

B.U.S. can be constructed with optimized properties and sold in a kit, for example, en masse manufacture by a commercial supplier. The kit can have reagents to attach multiple fluorophores and specific DNA sequences on a bead coated with streptavidin for use with biotin-end-labeled oligonucleotides for attachment to streptavidin – binding sites, as well as cocktails of amine-reactive fluorescent probes (e.g. succinimidyl esters of fluorescent probes) for the simultaneous attachment to the free primary amines of streptavidin. Alternatively, on carboxylic acid-coated beads, both the fluorescent compounds and amine-end-labeled nucleic acids can be attached.

The user can decide which and how many genes and which controls are included in each individual experiment, and can change the genes included 'at a moment's notice', in the next experiment. This is currently impossible with chip microarray technology. Furthermore the technique should be significantly cheaper than current chip microarray technology, as the manufacturing or reading of the beads does not entail sophisticated procedures, while the demand is anticipated to be very high.

In still another embodiment, multiple subsets of non-fluorescent microbeads are engineered, each subset tagged with numerous copies of a distinct single stranded DNA ('Control DNA, e.g. a gene; or a specific gene fragment; or an oligonucleotide representing a gene fragment). Numerous single strands of control DNA from each specific gene can be attached to the functionalized surface of a microbead as described above. Alternatively, oligonucleotides of a specific sequence can be attached on the microbead, or grown directly on the microbead

using phosphoramidite chemistry which is established in the field of DNA synthesis. The size of this 'control cDNA' on the microbead surface can be of any length, but preferably such that it ensures maximum hybridization with a corresponding 'target cDNA' that has a complementary sequence. Large amounts of the same microbead carrying a single DNA sequence can then be manufactured. Next, a second microbead is selected and the process repeated for a second gene/fragment of interest. By repeating the procedure several times (e.g. 1,000 times), stocks of microbeads each carrying a unique DNA sequence are manufactured. The process can readily be automated.

As explained above, the gene immobilized on the microbead hybridizes to a labeled DNA, e.g., a 'fluorescently labeled target DNA', the hybridization produces a unique fluorescent signal which may be detected by the flow cytometer as the microbead passes through, and then the microbead is appropriately sorted. Therefore several millions microbeads emitting fluorescent signals, representing hundreds or thousands of diverse genes in the target DNA, can be rapidly detected and separated from non-fluorescent microbeads by this procedure.

A collection of individuals with a specific pathology can be screened for a common genetic trait (e.g. common set of uknown mutations/polymorphisms in an unknown set of genes) because the present invention allows the isolation and identification of only those genes that appear in common in the patient population. For example, for a set of 5 patients having early-age lung cancer, and for whom an inherited set of unknown mutated genes can be hypothesized.

DNA from a tissue sample (e.g. lymphoblasts) from each patient is extracted. The DNA is then enzymatically digested into small fragments (e.g.50-200 base pairs). Third, those fragments that contain polymorphisms (mutations) that appear as heterozygosities among the two alleles, are selected and isolated from the population of non-mutated DNA fragments. One preferred procedure to accomplish this task is the one described above using aldehyde-linker based methods, and will not be further addressed here. Another less preferred possibility is to utilize any other described technology that can 'capture' and isolate the heterozygous sequences, while discarding the non-mutant sequences. A combination of the above technologies is also possible. The 'target' DNA selected from each affected individual for polymorphism screening may be only from one gene; or preferably from several genes or from the whole cDNA library, or from the whole genome.

Following isolation of mutant fragments from each individual, the fragments are PCR-amplified using primers which are labeled with a fluorescent probe, or

combination of probes. Thus, each of the 5 individual's mutant DNA is labeled with a different fluorescent probe from that of another member of the group, or combination of their probes, appropriate for flow cytometry (7-hydroxycoumarin; fluorescein; rhodamine; Texas Red; Bodipy; etc).

The fluorescently-labeled DNA from the 5 individuals is then mixed together and hybridized with microbeads engineered as described above. As described above, each subset of microbeads carries a specific gene (or gene fragment) as a 'control' DNA. Let us assume for simplicity that only one subset of microbeads is included, representing only one gene fragment. Each such microbead will allow target DNA from all 5 individuals to bind to it, if the corresponding gene fragment is mutated in all 5 individuals simultaneously. This microbead will emit all 5 fluorescent wavelengths when screened via flow cytometry, and therefore can be sorted and separated from microbeads with 4, 3, 2, 1 or none fluorescent emissions.

If instead of one, several subsets of microbeads representing numerous genes are utilized, the procedure remains unchanged. The flow cytometer will sort in the same container all microbeads (i.e. all genes, irrespective of their identity) that present fluorescent signals from a set percentage, e.g., all 5 fluorescent probes (i.e. genes from all 5 individuals). These are the genes of interest, i.e. the mutated genes that are likely to be the cause of the common disease in the patient population.

To discover the identity of these common genes, following flow cytometry the sorted microbeads are used in a PCR reaction to amplify the DNA fragments immobilized on them. By the design of the procedure followed for isolation of the mutant fragments, each fragment is flanked by the same, known PCR primers (see DFCI patent previously submitted). Finally, following PCR, the amplified fragments can be identified by a single application on a DNA microarray such as those that are currently commercially available, or preferably those described above.

Thus, the present procedure allows the isolation and identification of those mutated/polymorphic genes that appear in common in the affected patient population, and therefore have an increased probability to be related (or to be the cause) of the specific disease. Although the mutated/polymorphic genes in any single individual are likely to be many (e.g. 100,000 polymorphisms across the whole cDNA), the common genes are likely to be much fewer in number, since they must appear in common in all 5 individuals. The more individuals are

simultaneously screened with the present method, the fewer the 'common' mutated genes are going to be, therefore the faster the 'disease genes' will be identified.

If more than 5 individuals are to be screened for common mutated genes with the present method, the flow-cytometric method again can be applied as follows: First, DNA from 5 individuals is fluorescently labeled and screened as described above. The sorted microbeads corresponding to genes mutated in all 5 individuals is then briefly treated to remove the fluorescence (e.g. via heating for 2 min, 96°C, to remove the fluorescent strand; or via an enzyme that removes the DNA segment containing the fluorescent probe; etc). These microspheres are then again processed via the same procedure, to screen 5 more individuals: they are mixed with the labeled DNA from these individuals, to hybridize to mutated genes; then they are sorted via flow cytometry to select those microbeads that emit all 5 fluorescent signals. These represent microbeads that have captured mutant genes both in the first and in the second set of 5 individuals, and they now represent genes that are mutated in all 10 (5+5) individuals. By repeating the procedure more individuals can be screened. Ultimately the number of sorted microbeads will decrease significantly, and should correspond to the few 'disease - specific genes' that are sought.

It is possible however that, although a mutated gene may be involved in the causes of cancer in most individuals, it is not required for cancer formation. Therefore it may only appear as a mutated gene in 80% of cancers. A significant feature of the present invention is that, apart from identifying genes that are mutated in all individuals, it also allows for mutated genes that appear in most but not all individuals to be identified. For example, flow-cytometric sorting of the microspheres can be adjusted so that if 2 out of signals, more preferably, 3 out of 5 signals are present in a single microsphere this is also selected in a separate container. One can readily select a particular cut-off percent. Similarly, if the mutated gene is present in 4 out of 5 individuals it will also be sorted, etc. In this manner, not only genes mutated in 100% of the patients, but also genes mutated in preferably 50%, or more preferably 60%, etc, will also be identified. Still more preferably the mutation is present in 80% of the members. Higher per centages will be seen when the group is composed of related individuals.

A practical advantage of utilizing flow cytometry and microbeads for the present approach is that, because the DNA is not immobilized on 'inflexible' DNA chips but on individual microbeads, the user has full control over which genes and controls will be included in his study, thereby the method can be adjusted to the

needs of a particular application, on an experiment–by–experiment basis.

In an alternative application, instead of searching for common mututatios/ polymorphisms in a patient population, common expression of specific sets of genes is sought (e.g. common genes up –or down – regulated in cancer tissues, or in a set of patients with susceptibility to lung cancer, etc).

To utilize the present invention for this purpose, a similar flow-cytometric analysis can be applied. Briefly, the cDNA generated from each individual is enzymatically fragmented, denatured and ligated with PCR primers tagged with a unique fluorescent probe, or combination of probes. The labeled, single stranded cDNA molecules from several individuals are then mixed together and incubated with microbeads designed and described above. The described flow-cytometric process is then applied to sort microspheres that present signals from all individuals simultaneously, signifying combined up-regulation of a specific gene in all individuals. Certain signal intensities can be selected as a 'threshold' above which a gene is considered as up-regulated (i.e. many copies bound to microbead) or down - regulated. Both commonly up–regulated and down-regulated genes in the population are bunched together and sorted in separate containers in this procedure. Therefore by following the protocol described above for deriving common mutations in the population, the genes that are up-regulated or down-regulated in common can also be identified.

Although flow cytometry is a preferred and very convenient embodiment of this invention, it has limitations as to how many fluorescent signals can be simultaneously detected each time a microsphere is measured, and how many different fluorophores can be used. With typical flow cytometers, 5-7 signals can be simultaneously measured; therefore each flow-cytometric measurement can readily detect common genes in 5-7 individuals, and the procedure can be repeated to screen additional individuals. For screening a population of 50 patients, the procedure typically needs to be repeated 7-10 times. Specialized flow-cytometers that can screen more than 5-7 signals simultaneously also exist, and these can increase the efficiency of this procedure.

Alternatively, the combined signals from microspheres can be detected with a different detection system, such as an ICCD camera, a microscope, or a photomultiplier that detects several wavelengths simultaneously.

Several control genes and microbeads can preferably be included in the flow-cytometric determination, in order to ensure each time that the method performs optimally.

In one embodiment, kits for carrying out the identification of these DNA mismatches with these beads can be sold. The kits would include the repair glycosylase, an X-Z-Y compound and preferably instructions. These materials can be in any vial. The materials can be in lyophilized form.

In a preferred embodiment, PCR primers would also be included.

In one preferred embodiment the following kit materials and instructions can be included:

Kit Formulation:

1. Isolate target and control cDNA.
2. Add PCR primers that contain a cleavable biotin at the end.
3. Mix target with control, cross-hybridize.
4. Bind sample to streptravidin - coated magnetic bead.
   (alternatively, streptavidin - coated microplates can be used).
5. With the sample immobilized on solid support, perform: hydroxylamine treatment/washing; Mut Y/TDG treatment(s)/washing; FARP/BARB/AED labeling/washing. Antibody labeling/washing; Chemiluminescence detection of mismatches. All these steps are very easy and convenient to perform with the DNA immobilized.
6. To recover sample and isolate the mutation - containing DNA, add DTT (see below) to break the S-S bond on the cleavable biotin.
7. Now apply the preparation on an appropriate solid support for the ligand compound chosen: (antifluorescein, streptavidin, succinimidyl - ester -coated plates for FARP, BARP and AED respectively). Remove unbound DNA, capture only mutated DNA.
8. Now collect mutated DNA from microplates. This can be done by several methods; e.g. adding 1 M of hydroxylamine to break the bond between the ligand and the DNA; or raising the temperature to denature captured DNA and collect the unmodified strand; or, in the case of cleavable –S-S- containing probes, simply add DTT to break the bond to the microplate.
9. Apply PCR using the primers inserted in step 2.
10. Detect mutated genes using the mutation detecting arrays described above using hybridization techniques.

In an alternative embodiment, the following protocol can be used:

1. Mixing the target and the control DNA (e.g. In a 1:1 ratio). The control DNA sequence is the wild type DNA corresponding to the target DNA sequence. The DNA can be whole or fraction of genomic DNA; or PCR-amplified portions of

genomic DNA; or a cDNA library corresponding to the coding portions of the DNA. In the case that snps among the two alleles in cells of a single individual are examined, DNA from one allele is taken as ' the target', while DNA from the second allele is the 'wild type'.

2. Digesting the mixture into small double stranded fragments (e.g. 50-200 base pairs). Digestion can be done using one or more restriction endonucleases (e.g saui, alui, hpa-2, etc) at standard conditions (1 h, 37°C). Digestion with enzymes that generate overhangs ('sticky ends') is preferred.
3. Hybridizing the target DNA with a control DNA sequence to create a duplex. The result of this hybridization is that mismatches occur at positions of SNPs and mutations. Self-hybridization of DNA from a single individual will automatically generate mismatches among fragments from the 2 alleles.
4. Ligating double – stranded linkers to both ends of the DNA fragments. Ligation is done via ligase, under standard conditions, using the sticky ends of the DNA fragments, or via established 5' end –ligation methods.
5. Crosslinking biotinylated mismatch repair enzyme (Mut Y) at the positions of mismatches on DNA, by following the procedure described above.
6. Isolating the biotinylated DNA fragments on avidin-coated magetic beads.
7. Pcr-amplifying the isolated DNA fragments, directly from beads, using the known primers ligated in step 4. The primers may be biotinylated if a chip DNA array is to be used in the last step.
8. The resulting end-biotinylated DNA sample can be hybridized directly on a chip mutation scanning array. The recognizable moiety can then be used to reveal gene regions that contain SNPs/mutations (e.g. Biotin – labeled DNA can be followed by fluoresceinated streptavidin, etc.).

The following examples are illustrative of the invention and are not limitations thereof.

EXAMPLE 1 METHOD FOR LARGE-SCALE DETECTION OF BASE-SUBSTITUTION MUTATIONS IN CANCEROUS SAMPLES, USING ONE OF THE X-Z-Y COMPOUNDS, THE FARP MARKER MOLECULE (See Figure 1)

Isolated mRNA from a cancerous tissue is transcribed into cDNA. Primers can be added to DNA at this stage for PCR amplification at a later stage (see Figure 1). The sample is then hybridized with a corresponding wild-type sample of DNA to generate mismatch pairing at the positions of mutations. The hybridized DNA is treated with hydroxylamine to remove any aldehydes that may have formed

spontaneously. The hybridized DNA sample is then treated with the Mut Y enzyme. Enzyme treatment recognizes A/G mismatches and upon recognition, depurinates the DNA and simultaneously generates an aldehyde at the site of mismatch. The DNA is then treated with the labeling compound AED or FARP or BARP to generate a covalent oxime bond at the position of the mismatch. Upon labeling, the DNA is immobilized on microplates appropriate for the specific labeling compound and excess, unlabeled DNA is washed away. The DNA labeled at mismatch sites can now be analyzed by a variety of methods including detection of total mutations by chemiluminescence or identification of labeled genes via DNA arrays.

MATERIALS AND METHODS

1) DNA, oligomers and chemicals: FARP [5-(((2-(carbohydrazino)-methyl)thio)acetyl)-aminofluorescein, aminoxyacetyl hydrazide, Fluorescent Aldehyde Reactive Probe] was synthesized as described (Makrigiorgos GM, Chakrabarti S and Mahmood S. Int J Radiat Biol, 74: 99-109, 1998). High purity genomic calf thymus DNA and double stranded ladder (pUC18 Msp I digest, 27-500 base pairs) was purchased from Sigma Chemical and used without further purification. Single stranded (+strand) M13 DNA was purchased from Pharmacia Biotech and pGXIs14 plasmid DNA, a gift from Professor MacLeod, MD, Anderson Cancer Center, was isolated from the host bacteria as described earlier (Makrigiorgos GM, Chakrabarti S and Mahmood S. *Int J Radiat Biol,* 74: 99-109, 1998). Both agarose gel electrophoresis and the absorbance ratio at 260 nm to 280 nm were performed to determine the purity of the plasmid. Gel-purified 49-mer oligonucleotides representing the TFIIIA transcription factor - binding sequence of the *Xenopus* rRNA gene (enumerated in Table 1, at the end of this Example) were supplied by Oligos Etc Inc. Enzyme Mut Y (*E. coli*) was purchased from Trevigen Inc. and stored as recommended by the manufacturers. Hydroxylamine purchased from Sigma Chemical was already freshly made prior to the experiments. GTG agarose was obtained from FMC Bioproducts, polyacrylamide gel electrophoresis reagents were from National Diagnostics while SYBR GOLD nucleic acid gel stain and Picogreen® DNA quantitation dye was supplied by Molecular Probes. For chemiluminescence studies, Reacti-Bind NeutrAvidin coated polystyrene plates (pre-blocked with Bovine Serum Albumin) were supplied by Pierce. Anti-fluorescein-Fab fragments (Sheep) - alkaline phosphatase conjugate (antiF-AP) was purchased from Boehringer Mannheim. CDP-Star, a 1, 2 dioxetane chemiluminescent enzyme substrate and Emerald-II® enhancer used with CDP-star

was purchased from TROPIX. Micro Bio-Spin G25 chromatography columns were obtained from Bio-Rad laboratories. Label IT9 Nucleic Acid biotinylation kit was purchased from PanVera Inc. All reagents and buffers were of analytical grade and made with ultrapure water (1800 Mohm m^{-1} resistivity) delivered by an Alpha-Q system (Millipore).

2) Acidic or physiological depurination of calf thymus DNA. Treatment with hydroxylamine.

Aldehyde containing apurinic/apyrimidinic (AP) sites were chemically induced in calf thymus or plasmid DNA by a short exposure (0-60 seconds) to acidic conditions (pH=3.5) over a set time period at a temperature of 38°C, as described (Makrigiorgos GM, Chakrabarti S and Mahmood S. Int J Radiat Biol, 74: 99-109, 1998). The reaction was halted by placing the sample quickly on ice and adding a neutralization solution (10% of 3M sodium acetate and 1M potassium phosphate buffer at pH 7 and 7.5 respectively), to final volume of 50 :l. AP sites were also slowly generated in calf thymus DNA via spontaneous depurination at 37°C, pH=7.0, over a period of days, and these were monitored with the present assay. Prior to incubation at 37°C, the DNA was treated with 5mM hydroxylamine for 1 hour at room temperature to remove traces of existing aldehydes from the pool of potential FARP-binding sites. The hydroxylamine was then removed via G25 ultracentrifugation and the sample was resuspended in sodium phosphate buffer, pH 7.

3) FARP-trapping of aldehydes and subsequent DNA biotinylation.

To covalently trap open-chain aldehydes generated in DNA at the position of AP sites, 500 :M FARP was reacted with 0.05-2.5:g of DNA in 40 mM sodium citrate pH 7.0 at 15-22°C, for 30 minutes. Non-covalently bound FARP was removed by G25 ultracentrifugation. FARP-labeled DNA was either used on the same day or stored at 4°C or -20°C for a few days, prior to further experiments. To immobilize FARP-labeled DNA on neutravidin microplates, the DNA was exposed for one hour to a commercially available biotinylation reagent (Biotin Label IT™ reagent, 1:l; reagent per :g DNA, in MOPS buffer , pH 7.5 at 37°). Excess reagent was them removed by G25 ultracentrifugation. The samples were either used immediately or stored at 4°C for a few days, prior to chemiluminescent studies.

4) Chemiluminescence measurement of FARP-trapped aldehydes in calf thymus or plasmid DNA.

Double stranded DNA, doubly labeled with FARP and biotin, was immobilized on neutravidin - coated microplate strips in the presence of 5 nM

antiF-AP. 30-50 ng of doubly labeled DNA plus 5 nM antiF-AP in a total of 50 :l was incubated at room temperature for one hour in TE pH 7.5. Unbound sample and antiF-AP were removed by pipeting and washing with TE at least four times. The microplate strips were then transferred in to 50 ml polypropylene tubes and washed four times in 30ml - 50 ml of TE buffer with constant agitation for 10 minutes. The chemiluminescent substrates(CDP-Star plus Emerald II enhancer) were then added in 0.1 M diethanolamine, pH 8.5 and the anti-F-AP-catalyzed reaction was carried out at room temperature for 1 hour, after which maximum light generation was achieved. In separate experiments, to quantitate the fraction of biotinylated DNA captured on microplates Picogreen9dye was used to measure double stranded DNA just prior and after its removal from neutravidin-coated plates.

5) Chemiluminescence Instrumentation

The low light from the chemiluminescence reaction was detected using an intensified charged coupled device (ICCD) system (Princeton Instruments). This ICCD camera utilizes a proximity focused microchannel plate (MCP) image intensifier, fiber-optically coupled to the CCD array. The entire area of the ICCD is capable of light detection, giving a total of 576 x 384 pixels on a Pentium ® PC computer screen. Both the intensifier and CCD are cooled to -35°C thermoelectrically and the dark current is less than 50 counts per minute. The ICCD was used to detect total light generation from each cell of the microplate strip. Cells were individually placed in a reproducible geometry at ~2mm distance from the ICCD and the total light output per second measured. The background chemiluminescence (signal measured when FARP was omitted from the procedure) was subtracted from all samples. All measurements were repeated at least three times.

6) Formation of homoduplex and heteroduplex oligonucleotides.

49-mer oligonucleotides and their complementary strands with or without a centrally located T-to-G base substitution were synthesized. In another synthesis of the same oligomers, 5' biotinlyated 49-mers and their complementary unbiotinylated strands were synthesized (Table 1). For hybridization, equimolar amounts (~0.5 :g) of each oligonucleotide were annealed in 40 mM Tris-HCl (pH 7.5), 20 mM $MgCl_2$ and 50 mM NaCl to form duplex oligonucleotides. The mixture was first heated to 95°C for 2 minutes, then allowed to hybridize at 65° C for 3 hours and cooled slowly to room temperature. Following hybridization, the double stranded 49-mers were treated with hydroxylamine (5mM in citrate pH 7.0, for 30

minutes, 25°C) to remove traces of spontaneously or heat--generated aldehydes from the pool of FARP-reactive sites.

7) Treatment of M13 DNA, ladder DNA and duplex oligonucleotides with Mut Y and TDG and gel electrophoresis:

50 ng of the test DNA (single stranded M13, ladder DNA, or duplex oligonucleotide were incubated for 1 hour, 37°C with 1.0 unit Mut Y in 40 mM Na-citrate buffer (pH 7.0) and then alkali treated to concert positions of missing adenine to strand breaks. Analysis of cleavage products for single stranded M13 DNA was done by agarose gel electrophoresis 0.9% agarose, run overnight at 20 V in 1X TBE buffer and stained with 1:g/ml ethidium bromide). Fragment analysis for ladder DNA and oligonucleotides was done by 16% denaturing polyacrylamide gel electrophoresis in the presence of 7.5M urea at 20 V/cm. The DNA fragments were detected by SYBR Gold dye or by ethidium staining and photographs taken by Eagle Eye™ Still Video (Stratagene).

8) Chemiluminescence measurement of FARP-trapped mismatches in oligonucleotides, ladder and M13 DNA.

M13 DNA, ladder DNA, or 5'-biotinylated oligonucleotide duplexes, hydroxylamine-treated, were exposed to Mut Y, FARP-labeled biotinylated with the protocols described above. The biotinylation step was omitted for the oligonucleotides since these were pre-biotinylated In some experiments, samples were kept at 70°C for 8 minutes to inactivate the enzyme at this stage. Typically 50 ng from the doubly (biotin plus FARP) labeled nucleic acids were applied on neutravidin - coated microplates and their chemiluminescence measured.

RESULTS

1) Dual labeling of DNA and chemiluminescence detection using the present protocol

Figure 2 shows chemiluminescence obtained with the present setup when serial dilutions of free alkaline phosphatase were added to CDP-Star® substrate and Emerald II enhancer and measured using the cooled ICCD. The chemiluminescence detection limit of this set up is less than 0.01 attomoles alkaline phosphatase. Examination of the buildup of alkaline phosphatase chemiluminescent signal in solution following mixing with substrate plus enhancer at room temperature, demonstrates that after 60 minutes a relatively constant value is achieved (Figure 2, inset). Therefore all measurements reported were conducted 60 - 80 minutes following addition of the substrate. To estimate the

fraction of biotinylated DNA captured on the neutravidin-coated microplates, biotinylated DNA was quantitated using the fluorescence of Picogreen ™ dye prior to its application and immediately following removal of unbound DNA from microplates (not shown). 49-mer oligonucleotides resulted in approximately 10% capturing on the plates while of the 50-100 ng high molecular weight calf thymus DNA less than 2% was immobilized on the plates, possibly due to secondary structures and associated steric hindrances.

2) Ultrasensitive detection of aldehydes in DNA

Chemiluminescence detection of aldehyde-containing AP sites generated in 100 ng plasmid DNA following depurination in sodium citrate, pH 3.5 at 38°C for up to 60 seconds and trapping of AP sites by FARP is depicted in Figure 3. The induction of luminescence is linear with respect to depurination exposure. The inset, from an earlier work (Makrigiorgos GM, Chakrabarti S and Mahmood S. Int J Radiat Biol, 74: 99-109, 1998), demonstrated detection of fluorescence following FARP-labeling of this same plasmid exposed under identical conditions to higher depurination times (0-60 minutes). The fluorescence-based approach is less sensitive than the present method, however, it allows direct quantitation of the number of FARP molecules per DNA base pair. Five minutes depurination under the same protocol yields approximately 1 AP site per 34,000 bases (Makrigiorgos GM, Chakrabarti S and Mahmood S. Int J Radiat Biol, 74: 99-109, 1998). Assuming a linear decrease of AP sites for lower depurination exposures, the 15 second exposure in Figure 3 corresponds to approximately 1 AP site per 7×10^5 bases. The amount of microplate-captured DNA generating this signal is approximately 1-2 ng. Therefore the absolute number of AP sites recorded following 15 seconds depurination is approximately 5 attomole (see right axis in Figure 3).

To estimate the lowest number of AP sites detectable, hydroxylamine treatment of genomic calf thymus DNA was first employed in order to remove traces of spontaneously-generated AP sites (e.g. AP sites expected to be present in genomic DNA from mammalian cells prior to DNA extraction plus AP sites generated during handling). Hydroxylamine is a small molecule and is expected to react rapidly with aldehydes, as previously demonstrated for methoxyamine (Talpaert-Borle M, and Liuzzi M. *Biochimica Biophysica Acta*, 740: 410-416, 1983), thereby prohibiting subsequently added FARP to react at the same positions. Figure 4A depicts the decrease in the chemiluminescence signal obtained following hydroxylamine treatment of genomic calf thymus DNA depurinated for 15 seconds. Following hydroxylamine removal and reaction with FARP, the chemiluminescence

was reduced to almost background levels. When hydroxylamine-treated calf thymus DNA was kept at 37°C, phosphate buffer pH=7, and assayed for AP sites via FARP as a function of time, a linear increase in spontaneously-generated aldehydic AP sites was detected (Figure 4B). DNA kept at 4°C under similar conditions did not display any luminescence signal (Figure 4B). According to Figure 4B, the limit of detection by the present microplate-based method is ~0.2 attomole AP sites, or 1 AP site per $2x10^7$ bases, using a starting DNA material of about 100 ng.

3) Gel electrophoresis of Mut Y-treated oligonucleotides and single stranded M13 DNA.

49-mer oligomers engineered to form a double stranded structure, with or without a centrally located A/G mismatch upon hybridization, were exposed to Mut Y, alkali treated and examined upon denaturing gel electrophoresis. Generation of the two expected fragments was observed for the heteroduplex oligomers, while no cutting as present in the homoduplexes (Figure 5A). Under the conditions applied, the fragmented DNA appears to be less than 50% of the total DNA per lane, which would result if all A/G mismatches were reacted upon by Mut Y. The homoduplex-containing double stranded DNA ladder (27-500 base pair fragments) did not demonstrate additional fragmentation following enzymatic treatment (Figure 5B). In contrast, Mut Y treatment of the 7249 base-long M13 single stranded DNA resulted in the generation of approximately 6 fragments, the largest of which is about 1000 bases long, as demonstrated in lane 5, Figure 4C. Generation of Mut Y-recognized sites in the single stranded high molecular weight DNA is attributed to sequence self-complementation generating transient mismatches. It can be inferred that, to generate 6 discrete fragments, and assuming a less than 100% efficiency of Mut Y in cutting each site, an average of 3 Mut Y-recognized cutting sites are generated per each 7249 base-long M13 molecule.

4) FARP-based chemiluminescence detection of mismatches in high and low molecular weight DNA.

Starting with 100 ng of biotinylated 49-mer homoduplexes or heteroduplexes,
the nucleic acid was treated successively with hydroxylamine, Mut Y, then FARP and applied on neutravidin microplates for chemiluminescence detection of mismatches. A strong signal was obtained for A/G mismatch-containing oligonucleotides (Figure 6), while no signal was obtained when Mut Y was omitted, or when oligonucleotides without mismatch were Mut Y-treated. A mixture of

double stranded homoduplexes (DNA ladder) treated in the same way also demonstrated absence of chemiluminescence signals (Figure 7). In contrast, single-stranded M13 demonstrated a chemiluminescence signal of about 100 times the signal obtained without Mut Y indicating the generation of FARP-reactive sites following Mut Y treatment (Figure 7). The chemiluminescence results agree with the fragmentation results obtained by gel electrophoresis (Figure 5).

Table-1: Sequences of the synthesized oligonucleotides

1. B-5'-GTC TCC CAT CCA AGT ACT AAC CAG GCC CGA CCC TGC TTG GCT TCC GAT T-3' (SEQ ID NO:1)

2. B-5'-AAT CGG AAG CCA AGC AGG GTA GGG CCT GGT TAG TAC TTG GAT GGG AGA C-3' (SEQ ID NO:2)

3. B-5'-AAT CGG AAG CCA AGC AGG GTA GGG CCT GGG TAG TAC TTG GAT GGG AGA C-3' (SEQ ID NO:3)

1 and 2 are complementary and form a homoduplex. 1 and 3 form a heteroduplex with an A/G mismatch at position 20. On a separate set of oligonucleotides, a biotin molecule (B) was incorporated at 5' end during synthesis.

EXAMPLE 2

BARP - BASED DETECTION OF MISMATCHES FORMED VIA SELF-COMPLEMENTATION OF SINGLE - STRANDED M13 DNA.

Samples of M13 single stranded DNA that contain approximately 1 Mut Y-recognizable mismatch per 2,500 bases were treated with Mut Y to generate aldehyde -containing reactive sites appropriate for reaction with BARP. Nominal gel electrophoretic studies as well as BARP-based chemiluminescent studies were then preformed. Control samples used were: Single stranded M13 without enzymatic treatment; Double stranded M13 DNA without any mismatches and no enzyme treatment; and double stranded M13 DNA without mismatches and enzyme. Figure 8 (A and B) shows the results of both methods of detection. Figure 8A (luminescence studies) show that only when mismatches are present (single stranded M13) and Mut Y is used is there a chemiluminescence signal. In agreement, gel electrophoresis (Figure 8B) shows cuts in M13 are only generated under the same conditions. It can be seen that there is good agreement among the two methods. As described, the method is highly specific for mismatch - containing DNA, i.e. DNA without mismatches, or DNA with mismatches but no Mut Y generate no signals.

EXAMPLE 3

DETECTION AND ISOLATION OF DNA CONTAINING BASE - SUBSTITUTION MUTATIONS: DETECTION OF A *SINGLE* A-TO-C TRANSVERSION ENGINEERED IN A P53 GENE WITHIN A 7091 - LONG PLASMID.

The ability of the present technology (A.L.B.U.M.S) to detect base mismatches (demonstrated in previous examples) is directly applicable to detection of base substitution mutations. For example, a standard procedure to generate mismatches at the positions of mutations in DNA, is to mix mutation - containing DNA with wild - type DNA. Upon heating and re-hybridization of the mixture, heteroduplexes with mismatches are generated at the positions of mutations (Figure 1), which can then be detected with high sensitivity and specificity as demonstrated in example 1.

To isolate mutation - containing DNA from normal DNA, following BARP - labeling of the generated aldehydes at positions of mismatches (Figure 1) the DNA is immobilized on neutravidin-coated microplates, followed by *exhaustive* washing to remove the homoduplex DNA. As a result, only BARP-containing DNA is retained on the plates, thereby isolating mutant DNA.

To recover the purified mutation - containing DNA from the microplate, the samples can be either heated 2 min at 96°C or treated 1 min with NaOH to denature the DNA and recover the non-covalently modified strand, which is then used for amplification via PCR. The following section detail the procedure.

A 7,091 bp long plasmid that incorporates the full-length human cDNA p53 sequence (1,691 bp) was engineered to contain base subsitutions, via site-specific mutagenesis. The present technology was used to detect a known A-to-C base substitution mutation engineered in codon 378 within the plasmid-incorporated p53. Circular plamids (1 μg) containing mutant p53 genes were treated with a 5'-CG/CG-3' cutting enzyme (BstU I, Sigma, 1 unit, 1 h, 37°C) to generate linear fragments (~400 to 2,500 bp), followed by a 10 minute, 70°C treatment to inactivate the enzyme. The mutant-containing sample (1 μg) was mixed (1:1) with a similarly treated normal p53-containing sample, heated (96°C, 2 minutes) and hybridized overnight, at 65°C to generate A/G (25 %), and T/C (25 %) mismatches at p53 codon 378, as well as homoduplex p53 and plasmid fragments.

To detect the presence of the mutation via ALBUMS, 100 ng of the mismatch-containing DNA mixture (p53 plus plasmid fragments) was treated exactly as described for the M13 treatment in example 2: (a) hydroxylamine treatment and removal, (b) Mut Y treatment and BARP-binding, (c) fluoresceination

and (d) binding to neutravidin plates and chemiluminescence detection. Figure 9A demonstrates that strong signals are observed when the mutation is present, while background signals are obtained from normal p53-containing plasmid (i.e. complete lack of false positives). Figure 9B shows variation of signals versus DNA amount applied on microplates. These data represent an average of 4 independent experiments.

In conclusion, the present technology (A.L.B.U.M.S) allows a sensitive and specific detection of 1 base substitution mutation within a 7,091 bp-long, p53-containing plasmid with a virtual absence of false positives (defined as signal when no mismatch is present, Figure 9A). Unequivocal detection of a single base substitution within a 7,091 -long plasmid cannot easily be conducted with any of the existing methodologies (Nollau P and Wagener C. Clinical Chemistry 43: 1114-1128, 1997). ALBUMS on the other hand can detect the mutation on a microplate with minimal sample (<100 ng) and effort involved. Following formation of heteroduplexes, the procedure is currently completed in 6 hours, requires no special equipment or laborious handling and can be automated on microplates so that 96 samples can be examined at once. To achieve a similar result using conventional sequencing would not be possible (Primrose SB, Principles of Genome Analysis, Chapter 5, Sequencing Methods and Strategies, p125, Second Edition, Blackwell Science Ltd., Oxford, UK).

EXAMPLE 4

COMPARISON OF SMALL VERSUS LARGE LIGAND COMPOUNDS IN BINDING TO Mut Y – OR TDG- GENERATED REACTIVE SITES IN DNA: SYNTHESIS AND ADVANTAGE OF AED VERSUS BARP AND FARP. CHEMILUMINESCENCE SIGNALS BY AED.

(a) To synthesize AED, O-(Carboxymethyl)hydroxylamine hydrochloride was conjugated to ethylenediamine (Aldrich) in distilled water using 1-Ethyl-3-[3-(dimethylamino)propyl] carbodiimide (EDAC) as the coupling reagent. An 100-fold excess of ethylenediamine over O-(Carboxymethyl)hydroxylamine hydrochloride was utilized during the reaction to allow preferential coupling of ethylenediamine to the carboxyl groups. The conditions for the catalysis of this reaction by EDAC is well known to those skilled in the art. TLC analysis and purification on silica gel with $CHCl_3$:CH_3OH:CH_3COOH in a 70:20:5 ratio indicated the product at an R_f of 0.2-0.25. The certificate of analysis provided 1H NMR data consistent with the AED structure provided earlier.

(b) The ability of hydroxylamine - based compounds (e.g. FARP, AED, BARP, or methoxyamine) to bind reactive sites in DNA can be tested with a simple experiment. It is well known that, if hydroxylamine - compounds (such as methoxyamine) are covalently bound to aldehyde - containing abasic sites in DNA, then treatment with alkali (NaOH) cannot generate a strand break at the position of base loss (-otherwise a cut is generated). This simple observation allows direct testing of ligand binding to DNA following Mut Y - treatment of the nucleic acid (Figure 10A) or TDG – treatment of nucleic acid (Figure 10B). Mismatch - containing single - stranded M13 DNA was subjected to Mut Y to generate aldehyde containing abasic sites, and then alkali - treated to generate fragments at the positions of mismatches. Lane 2, in Figure 10A (agarose gel stained with ethidium bromide and photographed under UV light) demonstrates the generated fragments. In lanes 3, 4, 5, and 6, during Mut Y incubation the following ligand compounds were also included: 5 mM methoxyamine, 5 mM AED, 10 mM AED or 5 mM BARP respectively. As expected, the very low molecular weight compound methoxyamine prevents formation of any fragments, indicating a 100% binding to all reactive sites formed. Also, AED (bands D and E) demonstrates an almost complete binding to the reactive sites, especially when 10 mM is used (Lane E). In contrast, BARP can only prevent to a very small degree the formation of bands, indicating a very low (<5%) binding affinity to the reactive sites.

Similarly, in Figure 10B, the TDG enzyme was used (TDG recognizes mismatched thymine and generates an aldehyde at that position following excision of thymine). Oligonucleotides with a G/T mismatch were synthesized (lanes 1, 2, oligos alone) and exposed to TDG in the absence (lanes 3) or in the presence of 5 mM methoxyamine (lane 4), 5 mM BARP (lane 5), 5 mM AED (lane 6) or 0.5 mM FARP (lane 7). It can be seen that the cuts generated by TDG (lane 3 lower band) are not present when methoxyamine (lane 4) or AED (lane 6) are included in the reaction, demonstrating the binding of these compounds to the mismatches. BARP and FARP on the other hand (lanes 5 and 7) demonstrate significantly lower binding, since the lower band is present.

In conclusion: (a) AED is almost as efficient as methoxyamine (100%) in binding the Mut Y - generated reactive sites. (Methoxyamine itself however cannot be used in the present application because, unlike AED, following binding it allows no further derivatization as it has no secondary binding site available for antibody binding). (b) BARP only shows little (<5%) binding; despite that, and because the present method is extremely sensitive, high chemiluminescence signals are still

generated with BARP when mismatches are present, as shown in the previous example. The same is valid for FARP.

(c) The ability of DNA-bound AED to be recognized by a secondary ligand and then by an antibody, as described in the Detailed Description section of this invention was demonstrated by the following. The free primary amine (-NH2 group) of AED was covalently bound to biotin by addition of 1 mM biotin-LC-succinimidyl ester (Pierce) in 0.1 M sodium bicarbonate, pH=8.5 for 2h. The conjugate was purified by ultracentrifugation through 2 G25 filters (Pharmacia), fluoresceinated by using the Mirus fluoresceination reagent (Panvera Inc, see example 1) and then applied on neutravidin microplates. Addition of antifluorescein-AP antibody generated a strong chemiluminescence signal (Figure 12) in the sample treated with Mut Y enzyme (i.e. aldehydes were generated), but not in the sample not-treated with Mut Y (aldehydes not generated).

EXAMPLE 5

LABELING OF MISMATCHES WITH FARP, BARP or AED: INACTIVATION OF ENZYMATIC ACTION DURING LABELING.

A DNA sample containing mismatches is dissolved in a buffered solution and treated with a repair glycosylase, either Mut Y or TDG (1 unit enzyme per μg DNA). The reaction is incubated at 37 C for 1 hour. Upon completion of the reaction with Mut Y or TDG, the solution is cooled to 15° C, to arrest enzymatic activity. FARP is added to the sample and allowed to react for 30 minutes at 15° C. At the end of the 30 minute incubation with FARP, the reaction solution is suddenly heated to 70° C for two minutes to inactivate the enzyme. The sample of DNA is now ready for purification and detection as previously described. Alternatively, instead of heating to 70oC the enzyme can be solubilized and removed via a standard phenol-chloroform extraction, or via addition of Proteinase K (0.1 mg/ml, 2h, 37oC).

EXAMPLE 6

STRATEGY TO UTILIZE DNA CHIPS FOR DETECTION OF BOTH INHERITED POLYMORPHISMS AND MUTATIONS, AS WELL AS ACQUIRED MUTATIONS FROM CANCER SAMPLES.

The ability to derive both inherited and acquired genetic alterations in a single step over 6800 genes with the present procedure, using the Affymetrix array as an example, is described below.

Inherited single nucleotide polymorphisms (SNPs) are estimated to be present in the two alleles of each gene with a frequency of ~1:1000 bases. When an SNP in the coding sequence causes a debilitating change in the protein, heterozygous mutations arise which could result to early onset of cancer (e.g. the Li-Fraumeni syndrome). When cDNA from normal cells is melted and self-hybridized, mismatches will occur at positions of heterozygocities and SNPs, whenever both alleles are expressed, which will be detectable by the present technology (A.L.B.U.M.S) and would display positive on the DNA arrays. Because SNPs among alleles occur at a high frequency (~1:1000 bp) it is possible that within every single gene (average~2,000 bp) there is one or more SNPs. Therefore, if both paternal and maternal alleles are transcribed, self-hybridizing cDNA from whole genes would be expected to result in one or more mismatches per gene, as a result of allelic cross-hybridization. All array elements would then display positive, resulting to trivial information. By digesting the cDNA to ~100-200 bp pieces prior to ALBUMS genotypic selection (as described in example 3) the problem is avoided: Most fragments are likely to contain none, or occasionally one *inherited* SNP. ALBUMS will select mismatch - containing fragments, and array elements that score positive will be only those capturing a 100-200-mer gene fragment with an SNP.

Acquired mutations can be detected by following the same strategy, and by using cancer samples from the same individual as the normal sample. Again, by self-hybridizing cDNA from cancer samples and fragmenting to 100-200-mers, it is likely that most fragments will contain none, or occasionally one *inherited* SNP, or very occasionally one acquired mutation. Array elements that score positive will be those corresponding to genes that contain either inherited or acquired mutations, but rarely both.

An example of using the high resolution Affymetrix array (described earlier) to detect genetic alterations in parallel normal and cancer samples is displayed in Figure 12. cDNA from normal tissue is melted and self-hybridized to generate mismatches (Figure 12), then digested with appropriate enzymes to generate 100-200-mers and add primers; then the present technology, (ALBUMS), utilizing one of the probes (FARP, AED or BARP) selects the mismatches, PCR amplifies them and these are applied on the Affymetrix array: The mutation-containing 200-mers isolated via ALBUMS will cause certain 25-mer array elements to display positive, thereby identifying both the gene and the approximate (± 100-200 bp) location of an inherited polymorphism among the two alleles.

Next, cDNA from the cancer sample is melted, self-hybridized and processed similarly. Acquired mutations will show up as positive array elements that are negative on the normal tissue array. Acquired mutations scored on the same gene as an inherited mutation provide candidate genes to be examined for loss of heterozygocity, using existing methodologies. Finally, cDNA from cancerous cells will be cross-hybridized to cDNA from normal cells and the procedure will be repeated (not illustrated in Figure 12). This will detect acquired mutations in those genes that express a single allele in their mRNA, which would not be detected by self-hybridization alone.

The use of the Clontech array will provide similar information to the Affymetrix array. However, this array would be used with fewer genes and with smaller 'resolution', since the array elements contain 500 bases-long cDNA and it is possible that certain elements will capture both inherited SNPs and acquired mutations, thereby providing unclear information. On the other hand these arrays are simpler to use and do not require the fluorescent laser scanner, hence they are currently more accessible to users.

EXAMPLE 7
USE OF MICROBEAD MUTATION SCANNING ARRAY

ALBUMS was utilized as described above to isolate mutation - containing DNA fragments out of a mixture of diverse DNA fragments. In this system, we inserted a mutation in the p53 gene and we then used ALBUMS to select the mutation – containing fragment. Microbeads coated with oligonucleotides having a sequence complementary to the known, mutation-containing sequence, were then constructed and used to demonstrate detection. The methods and results are presented below:

a. Conjugation of oligonucleotides onto caboxylated Microspheres

1. Vortex carboxylated microspheres (from Molecular Probes) stock.
2. 5,000,000 microspheres are dispensed into a 1.5 ml microcentrifuge tube.
3. Disperse the microspheres by sonication for 30 seconds.
4. Centrifuge the microspheres at 8000 g for 1 minute.
5. Remove the supernatant.
6. Add 50 ul of 0.1 M MES (2-(N-Morpholino) ethanesulfonic acid) (pH 4.5).

Vortex.

7. Add 1 nmole of oligonucleotide (5'-amino-unilinker-25 mer-3') to microspheres. Vortex briefly.

8. Immediately before use, add 1 ml sterile water to 10 mg of EDC (1-ethyl-3-(3-dimethyaminopropyl) carbodiimide-HCl). Vortex until dissolved.

9. Add 2.5 ul of the fresh EDC solution to the microspheres. Vortex immediately.

10. Incubate for 30 minutes at room temperature.

11. Repeat steps 8-10 with fresh EDC.

12. Add 1.0 ml of 0.02% Tween-20. Vortex.

13. Centrifuge microspheres at 8000 g for 1 minute.

14. Remove the supernatant.

15. Add 1.0 ml of 0.1% SDS. Vortex.

16. Centrifuge microspheres at 8000 g for 1 minute.

17. Remove the supernatant.

18. Resuspend the microspheres in 100 ul of 0.1 M MES (pH 4.5).

19. Store the preparation at 4 °C.

b. Hybridization and capturing of ALBUMS – derived PCR fragments onto microspheres.

DNA Hybridization

1. Add DNA 17 µl of diluent (TE, etc.) to the control tube.
2. Add DNA (1, 5, 10, 25 ng) in 17 µl of volume to the sample tubes.
3. Incubate all tubes at the 96 °C heatblock for 10 minutes.
4. Vortex and sonicate oligo LGA-conjugated microspheres.
5. Dilute microspheres in 1.5x TMAC (Tetramethylammonium chloride) to a concentration of 10,000 microspheres per 33 ul of 1.5x TMAC hybridization buffer.
6. Place microspheres / 1.5x TMAC mixture at hybridization temperature (50 °C)
7. With the tubes still in the 96 °C heatblock, add 33 µl of microspheres / 1.5x TMAC mixture with hybridization temperature to the first tube.
8. Immediaterly close, remove and vortex the first tube.
9. Place the first tube at hybridization temperature.
10. Repeat steps 7-9 for all remaining tubes.
11. Incubate all tubes at hybridization temperature for 10 minutes.

12. Wash microspheres with PBS (pH 7.4).
13. Centrifuge tubes at 8000 g for 5 minutes. Remove supernatant.
14. Add 50 µl of PBS (pH 7.4) to the tubes.
15. Dilute streptavidin-alexa-488 to 10 µg/ml in PBS (pH 7.4)
16. Add 12 ul of streptavidin-alexa-488 to the tubes.
17. Incubate tubes at room temperature for 15 minute.
18. Analyze samples with flow cytometer.

c. Measurement of microsphere – bound DNA sequences on a flow cytometer.

Following hybridization of fluorescently – tagged ALBUMS – derived DNA sequences on microbeads, the sample was processed for flow cytometry, to identify the presence of sequences complementary to the sequences conjugated to the microbeads. Figure 14 demonstrates the results of 4 different hybridization experiments, as a function of the amount of DNA applied for hybridization to the microspheres: Curves 1-3 represent ALBUMS-derived DNA fragments that are 236, 86 and 57 base pair long, and all contain a sequence complementary to the microsphere – sequence. Curve 4 represents the hybridization of irrelevant (non-complementary) DNA sequences to the microspheres. A clear signal is observed only when the hybridized DNA sequences are complementary to the microbead-bound sequence.

Therefore we have demonstrated the ability of the system to specifically detect the presence or absence of certain sequences in the sample. In the simplified example shown above, the microspheres were not 'optically encoded', and only one set of microspheres was used. To construct a full Mutation Scanning Array, several sets of optically encoded microspheres can be used, each set containing a different sequence. By hybridizing the ALBUMS-derived DNA fragments to a mixture of different microspheres, readings from different micospheres can be simultaneously obtained over hundreds or thousands of sequences, via flow-cytometry. An illustration of how a microsphere – based Mutation Scanning Array is constructed, is presented in Figure 15.

All references discussed herein are incorporated herein by reference.

WHAT IS CLAIMED:

1. A method of using a mutation scanning array, wherein said mutation scanning array comprises a plurality of elements, wherein the elements contain immobilized oligonucleotides 8 - 50 bases long, that collectively span at least 10 different whole genes, to identify mutations in a target DNA sequence which comprises:

(a) hybridizing the target DNA with a control DNA sequence to create a duplex, wherein the control DNA sequence is the wild-type DNA corresponding to the target DNA sequence,

(b) tagging any mismatch in said duplex with a detectable moiety,

(c) cleaving the duplex into segments of 50 - 300 bases,

(d) removing the segments tagged with the detectable moiety,

(e) contacting the segments tagged with the detectable moiety with the mutation scanning array, and

(f) identifying in which gene and gene segment the selected mismatch belongs to.

2. The method of claim 10, wherein the segments tagged with the detectable moiety are amplified before being used on the mutation scanning array.

3. The method of claim 1 or 2, wherein the whole gene is represented by array elements; each element containing immobilized oligonucleotides that sample in 25-300 bases for the whole 3' to 5' mRNA sequence of each represented gene.

4. The method of claim 1 or 2, wherein each of the whole genes is represented by the coding genomic portion of the gene.

5. The method of claim 1 or 2, wherein each of the whole genes is represented by both the coding and non-coding genomic portions of a gene.

6. The method of claim 1 or 2, wherein at least 10 different genes are selected from the genome, collectively known to predispose an individual to a particular disease.

7. The method of claim 6, where the disease is a particular kind of cancer.

8. The method of claim 6, where the disease is a cardiovascular abnormality, or a neurodegenerative disorder, or diabetes.

9. The method of claim 1 or 2, where the genes selected are all known tumor suppressor genes or oncogenes.

10. The method of claim 1 or 2, where the genes selected are genes known to be overexpressed in a malignant cell, wherein overexpression is determined by comparison to the gene's expression in a corresponding non-malignant cell.

11. The method of claim 1, wherein the array is a chip or a microsphere.

12. A method of using a mutation scanning array to identify mutation in a large DNA sequence, wherein said mutation scanning array comprises a plurality of elements, wherein the elements contain immobilized oligonucleotides 8 - 50 bases long, that collectively span at least 5 different genes, wherein said method comprises:

(a) hybridizing the target DNA sequence with a control DNA sequence wherein said control DNA sequence is the wild-type DNA sequence corresponding to the target DNA sequence to create a duplex;

(b) digesting the duplex to fragments of 50-300 base pairs, with restriction enzymes that allow generic addition of PCR primers;

(c) adding PCR primers to the duplex

(d) treating the duplex to remove any spontaneous aldehydes;

(e) reacting the duplex with a repair glycosylase to convert any mismatched sites in the duplex to reactive sites containing an aldehyde-containing abasic site;

(f) reacting the duplex with a compound of the formula X-Z-Y, wherein X is a detectable moiety, Y is NHNH2, O-NH2 or NH2, and Z is a hydrocarbon, alkyhydroxy, alkylethoxy, alkylester, alkylether, alkylamide or alkylamine, wherein Z may be substituted or unsubstituted; or where Z may contain a cleavable group; for a sufficient time and under conditions to covalently bind to the reactive sites;

(g) detecting the bound compound to identify sites of mismatches;

(h) isolating the DNA that contains mismatches from DNA without mismatches;

(i) PCR-amplifying the mismatch-containing DNA

(j) applying the mismatch-containing DNA on the Mutation Scanning Array, to determine the genomic position(s) where mismatches occur; and

k) determining whether the mismatch is a mutation or polymorphism.

13. The method of claim 12, where the detectable moiety is selected from the group consisting of NH_2, SH, $NHNH_2$, a fluorescein derivative, a hydroxycoumarin derivative, a rhodamine derivative, a BODIPY derivative, a digoxigenin derivative and a biotin derivative.

14. A method of using a mutation scanning array to identify a common mutation in a group of at least 5 individuals having a disease comprising:

a) obtaining DNA or mRNA from said group of individuals;

b) digesting the DNA or mRNA into fragments of 50-200 base pairs;

c) identifying and tagging those fragments, creating tagged fragments, where a mismatch is present when said fragment is compared with a control wild-type fragment; and specifically labelling DNA or mRNA from each member of the group;

d) isolating said tagged fragments;

e) PCR-amplifying said tagged fragments using primers labeled with a probe, creating labeled DNA;

f) mixing the labeled DNA with a microsphere, wherein said microsphere contains single strand DNA of 50-300 base pairs obtained from wild-type genes of interest, under conditions permitting hybridization under moderate stringency;

g) subjecting the hybridized microsphere of step (f) to flow cytometry to sort in the same container those microbeads containing identical signals for a mutation from at least two members of said group ; and

h) identifying in which gene and gene segment the common mutation occurs.

15. The method of claim 14, wherein the flow cytometry is used to select identical signals from at least 50% of said group of individuals.

FIGURE 1

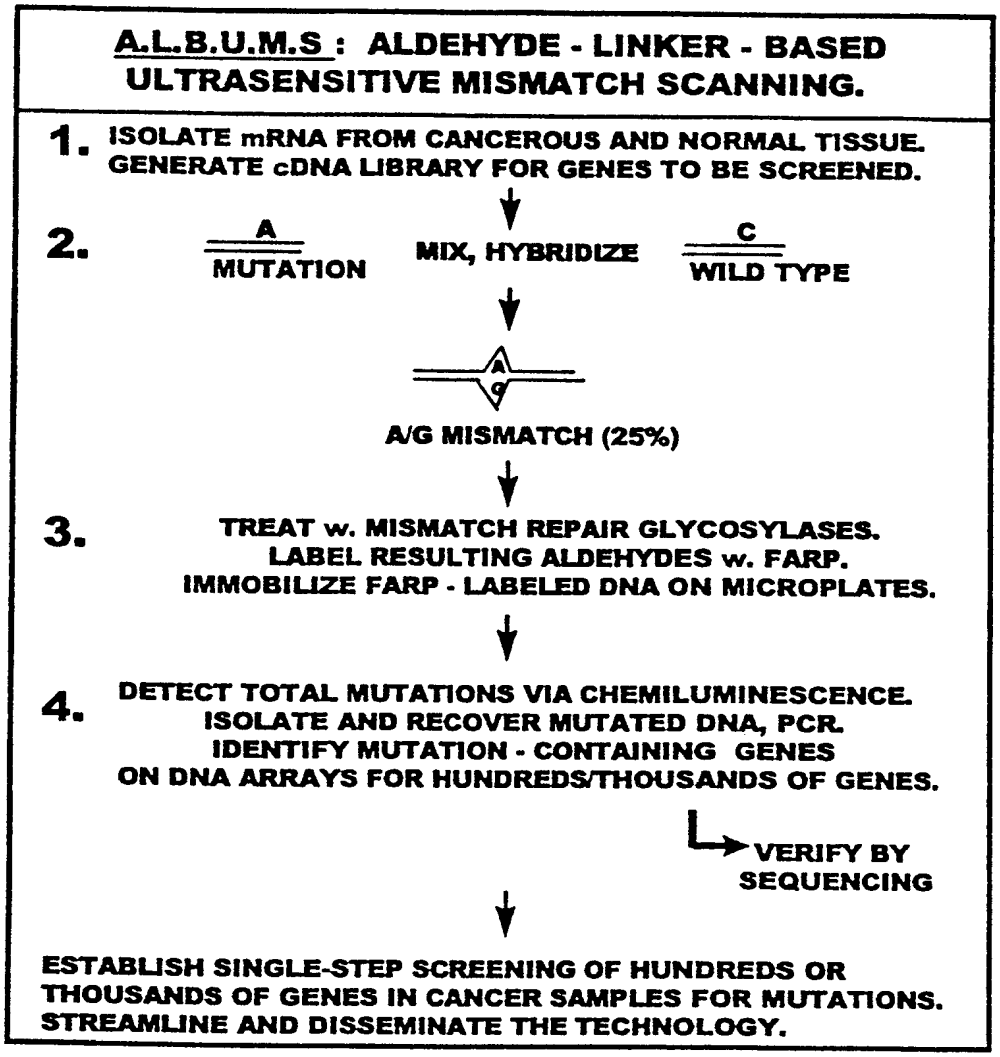

A.L.B.U.M.S : ALDEHYDE - LINKER - BASED ULTRASENSITIVE MISMATCH SCANNING.
1. ISOLATE mRNA FROM CANCEROUS AND NORMAL TISSUE. GENERATE cDNA LIBRARY FOR GENES TO BE SCREENED.
2. A MUTATION
MIX, HYBRIDIZE
C WILD TYPE
A
G
A/G MISMATCH (25%)
3. TREAT w. MISMATCH REPAIR GLYCOSYLASES. LABEL RESULTING ALDEHYDES w. FARP. IMMOBILIZE FARP - LABELED DNA ON MICROPLATES.
4. DETECT TOTAL MUTATIONS VIA CHEMILUMINESCENCE. ISOLATE AND RECOVER MUTATED DNA, PCR. IDENTIFY MUTATION - CONTAINING GENES ON DNA ARRAYS FOR HUNDREDS/THOUSANDS OF GENES.
VERIFY BY SEQUENCING
ESTABLISH SINGLE-STEP SCREENING OF HUNDREDS OR THOUSANDS OF GENES IN CANCER SAMPLES FOR MUTATIONS. STREAMLINE AND DISSEMINATE THE TECHNOLOGY.

FIGURE 2

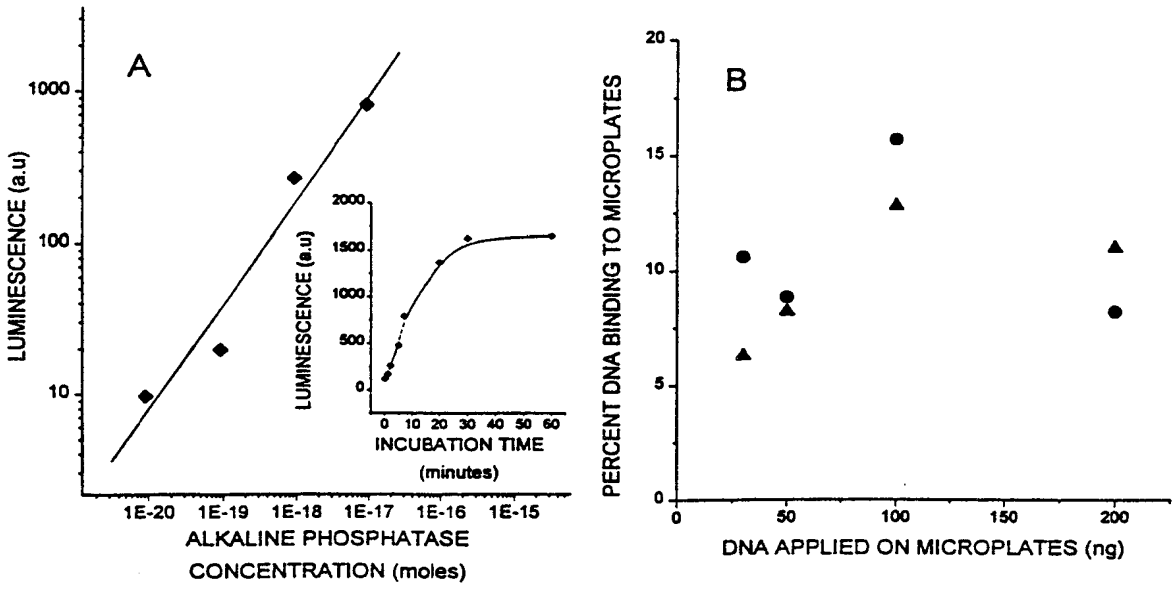
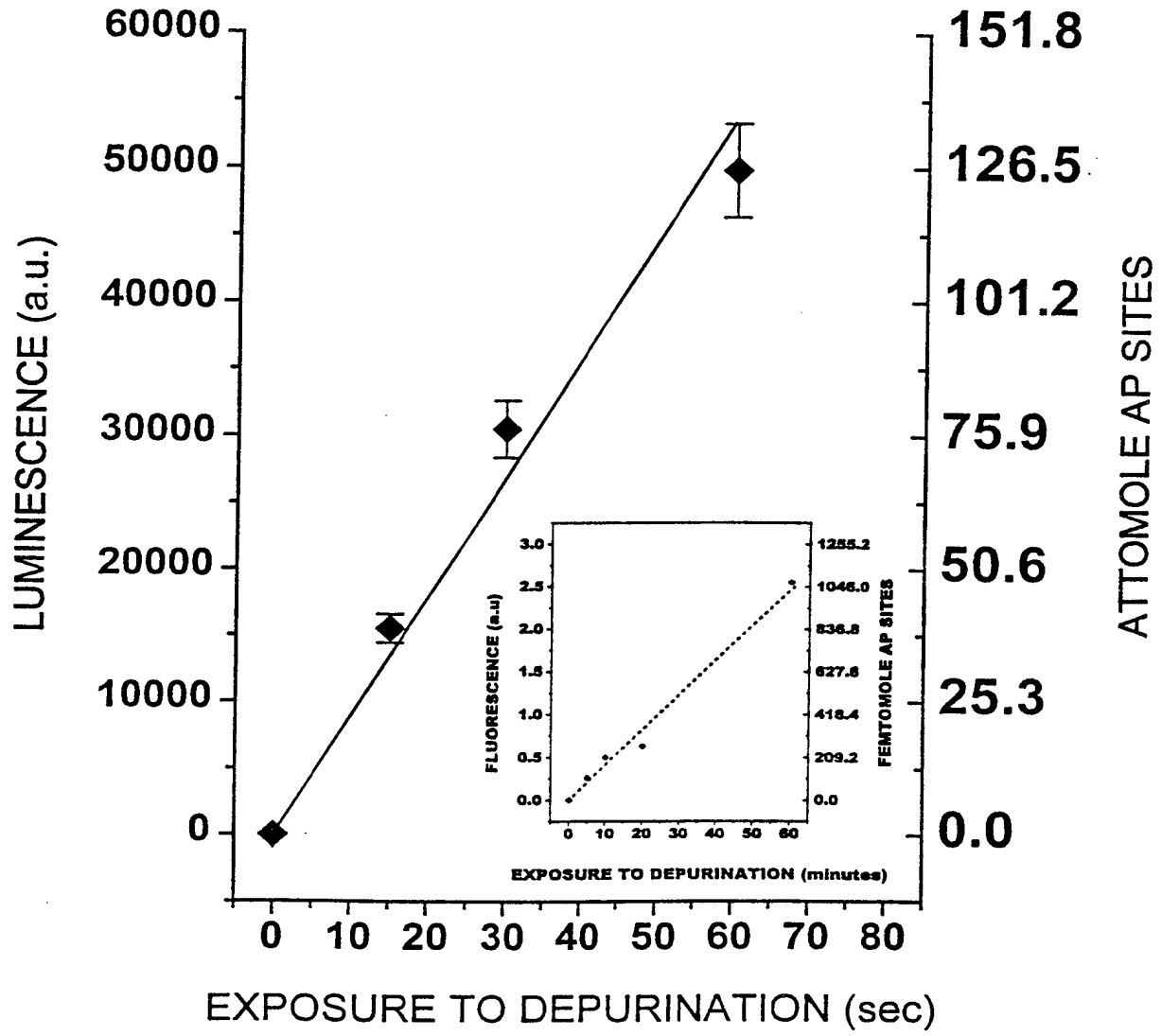

A
LUMINESCENCE (a.u)
1000
100
10
1E-20 1E-19 1E-18 1E-17 1E-16 1E-15
ALKALINE PHOSPHATASE
CONCENTRATION (moles)
LUMINESCENCE (a.u)
2000
1500
1000
500
0
0 10 20 30 40 50 60
INCUBATION TIME
(minutes)
B
PERCENT DNA BINDING TO MICROPLATES
20
15
10
5
0
0 50 100 150 200
DNA APPLIED ON MICROPLATES (ng)

FIGURE 3


LUMINESCENCE (a.u.)
60000
50000
40000
30000
20000
10000
0
0 10 20 30 40 50 60 70 80
EXPOSURE TO DEPURINATION (sec)
ATTOMOLE AP SITES
151.8
126.5
101.2
75.9
50.6
25.3
0.0
FLUORESCENCE (a.u)
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0 10 20 30 40 50 60
EXPOSURE TO DEPURINATION (minutes)
FEMTOMOLE AP SITES
1255.2
1046.0
836.8
627.6
418.4
209.2
0.0

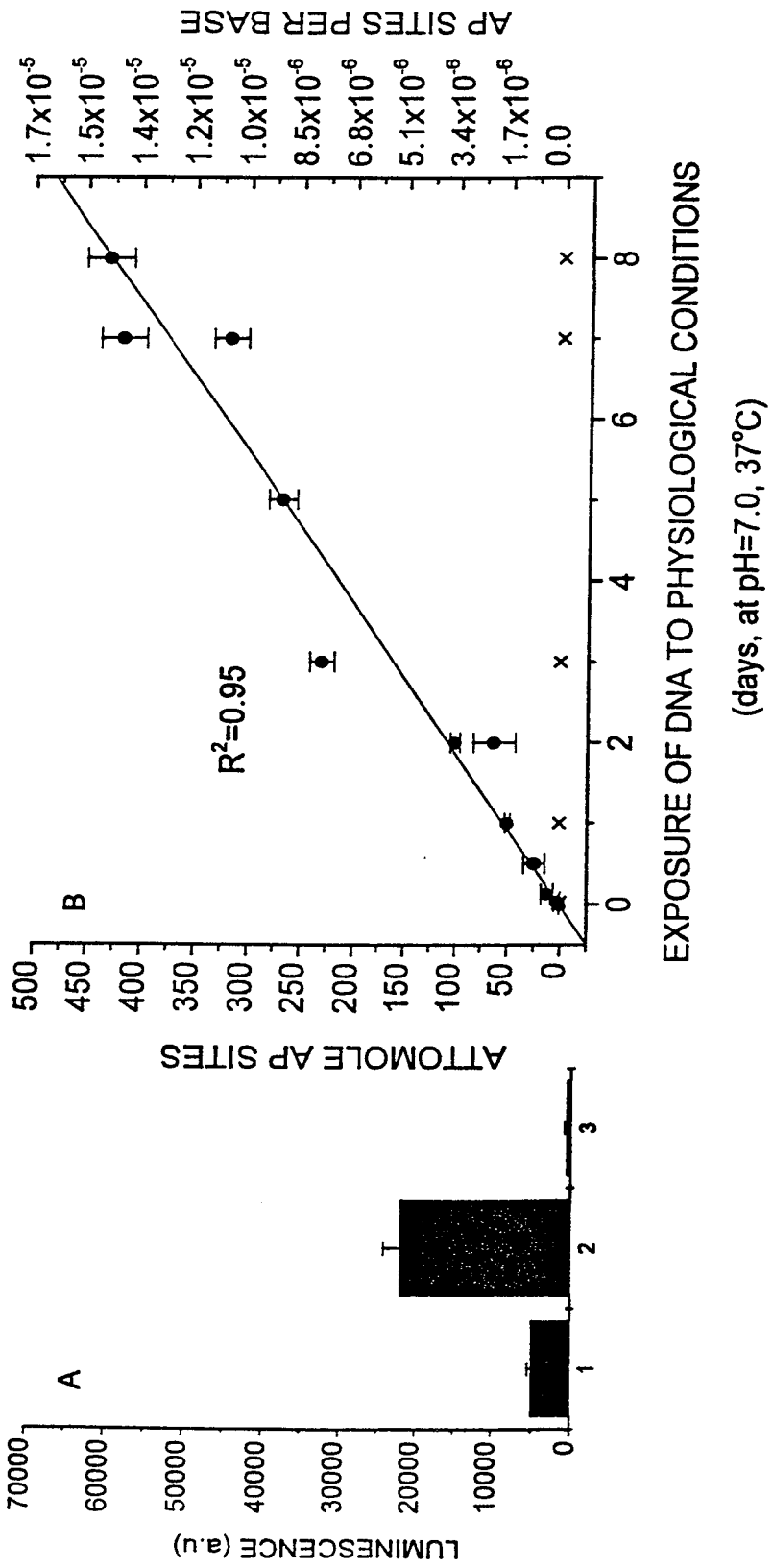

A
B
LUMINESCENCE (a.u)
ATTOMOLE AP SITES
AP SITES PER BASE
R^2=0.95
EXPOSURE OF DNA TO PHYSIOLOGICAL CONDITIONS
(days, at pH=7.0, 37°C)

FIGURE 4

FIGURES 5A, 5B, 5C

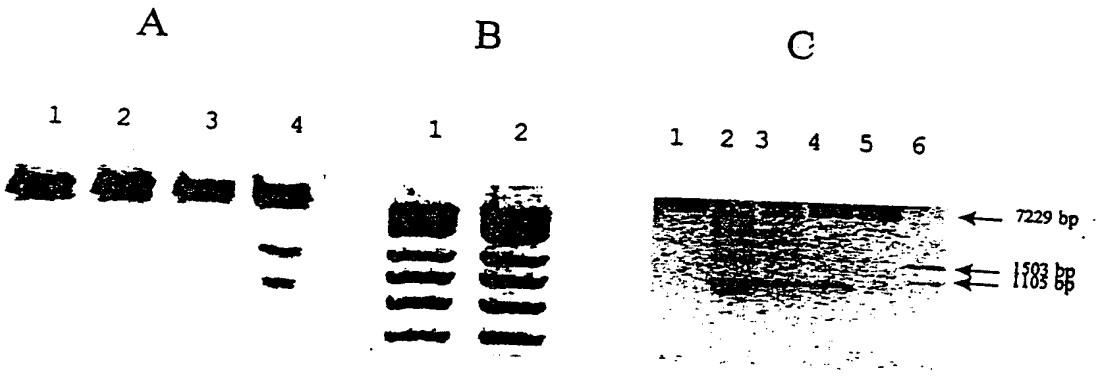


FIGURE 6

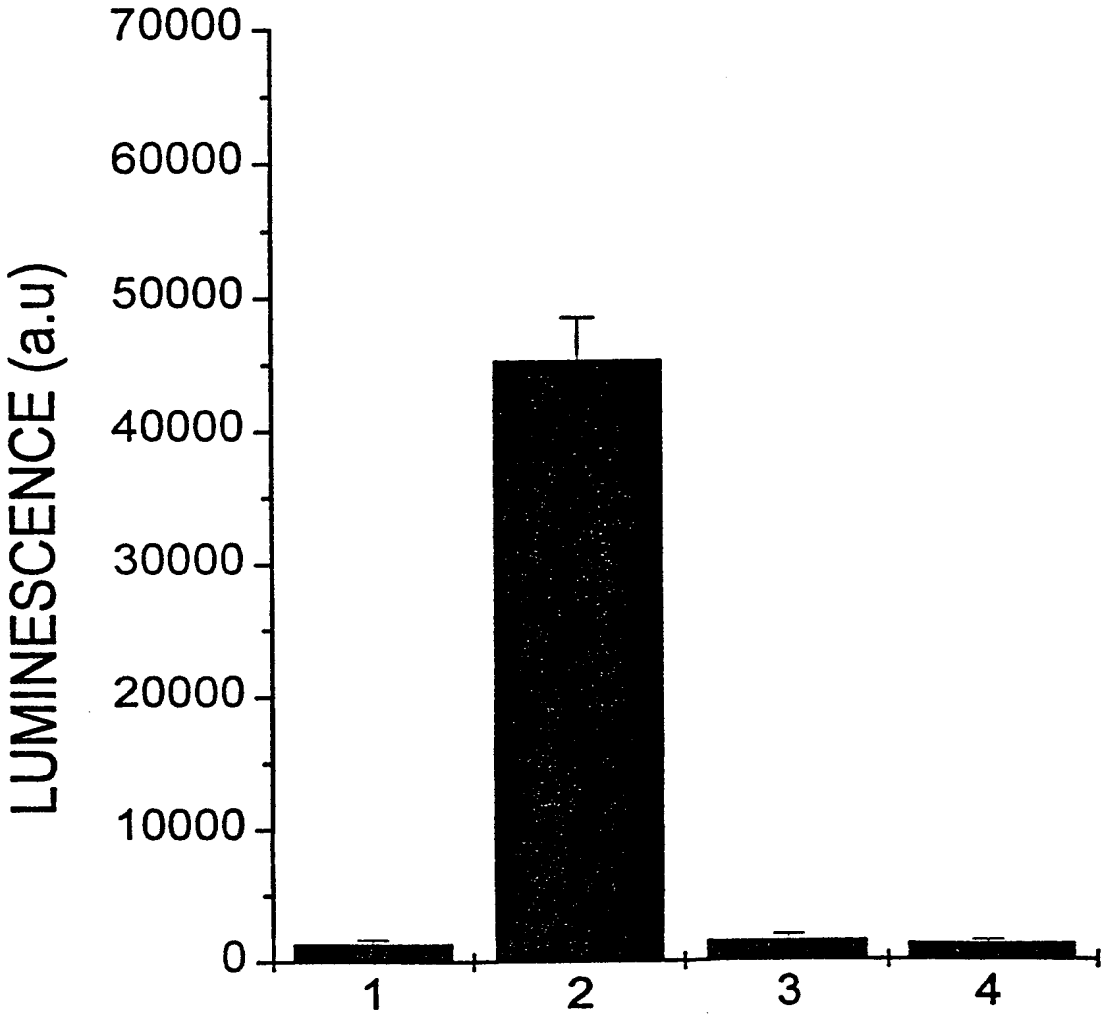
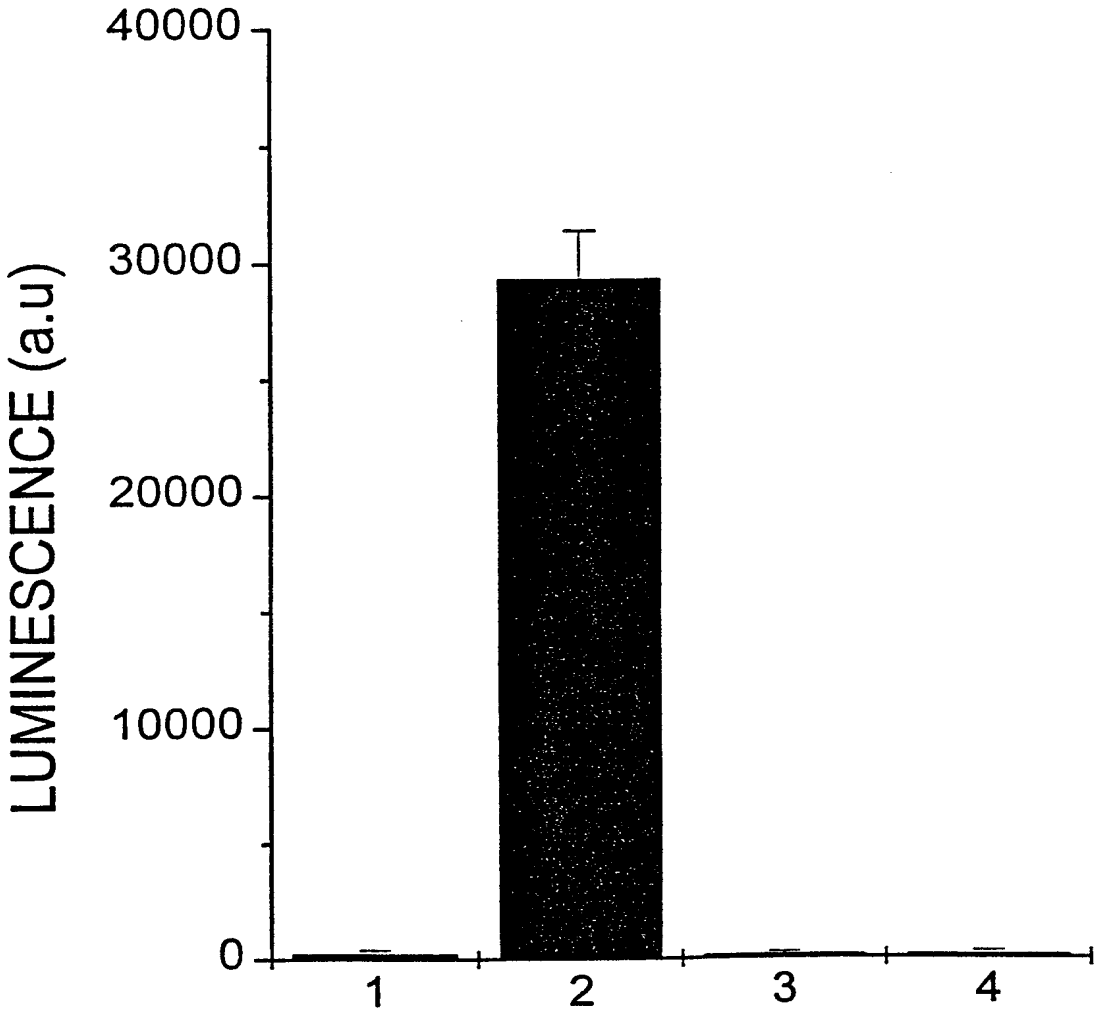

LUMINESCENCE (a.u)
70000
60000
50000
40000
30000
20000
10000
0
1
2
3
4

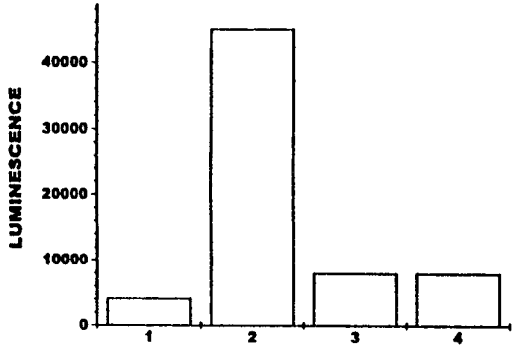
FIGURE 7



FIGURES 8A, 8B

A

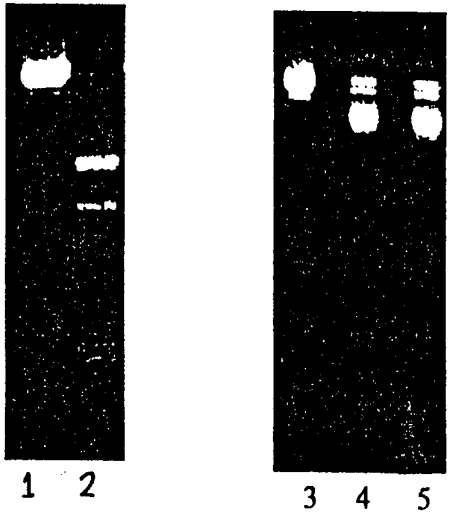
Detection of mismatches formed by hybridizing oligonucleotides with M13 DNA and exposed to MutY enzyme



1 = s.s. M13 DNA with ~ 1 A/G mismatch per 2500 bases, no enzyme
2 = s.s. M13 DNA with ~1 A/G mismatch per 2500 bases, plus enzyme
3 = d.s. M13 DNA without mismatch, no enzyme
4 = d.s. M13 DNA without mismatch, plus enzyme

4/13/98

B

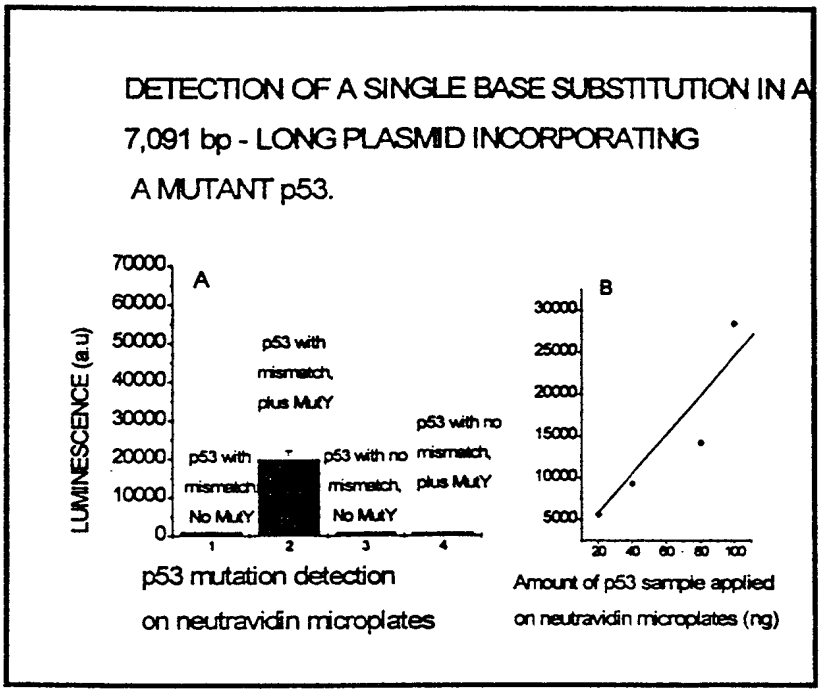


1 = single stranded (s.s) M13 with 1 A/G mismatch per 2500 bases, no enzyme
2 = s.s. M13 with 1 A/G mismatch per 2500 bases, plus enzyme

3 = s.s. M13 with 1 A/G mismatch per 2500 bases, no enzyme
4 = double stranded M13 without mismatch, no enzyme
5 = d.s. M13 without mismatch, plus enzyme

IDT Exhibit 1020.7943

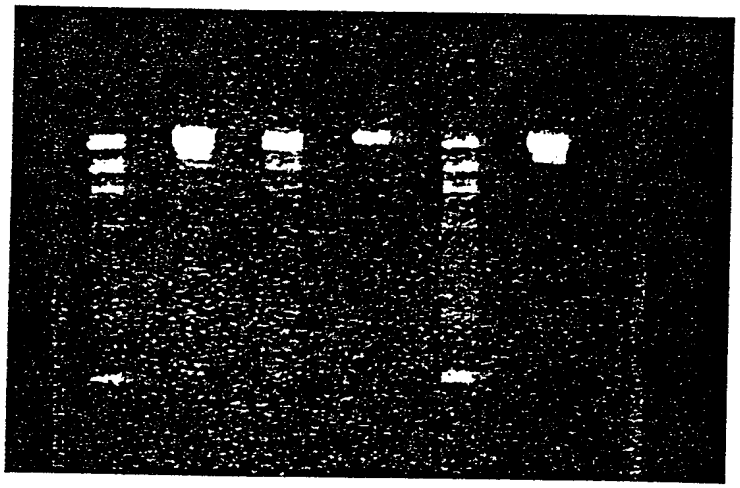
FIGURE 9


DETECTION OF A SINGLE BASE SUBSTITUTION IN A
7,091 bp - LONG PLASMID INCORPORATING
A MUTANT p53.
A
LUMINESCENCE (a.u)
70000
60000
50000
40000
30000
20000
10000
0
p53 with mismatch, plus MutY
p53 with mismatch, No MutY
p53 with no mismatch, No MutY
p53 with no mismatch, plus MutY
1
2
3
4
p53 mutation detection
on neutravidin microplates
B
30000
25000
20000
15000
10000
5000
20
40
60
80
100
Amount of p53 sample applied
on neutravidin microplates (ng)

FIGURES 10A, 10B

M13mp19 DNA TESTED.
A=DNA ALONE
B=DNA PLUS MUTY
C=DNA PLUS 5mM METHOXYAMINE PLUS MUTY
D=DNA PLUS 5 mM AED PLUS MUTY
E=DNA PLUS 10 mM AED PLUS MUTY
F=DNA PLUS 5 mM BARP PLUS MUTY

A



B

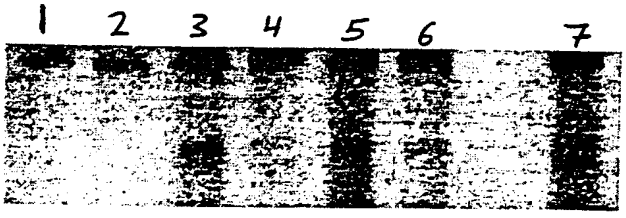
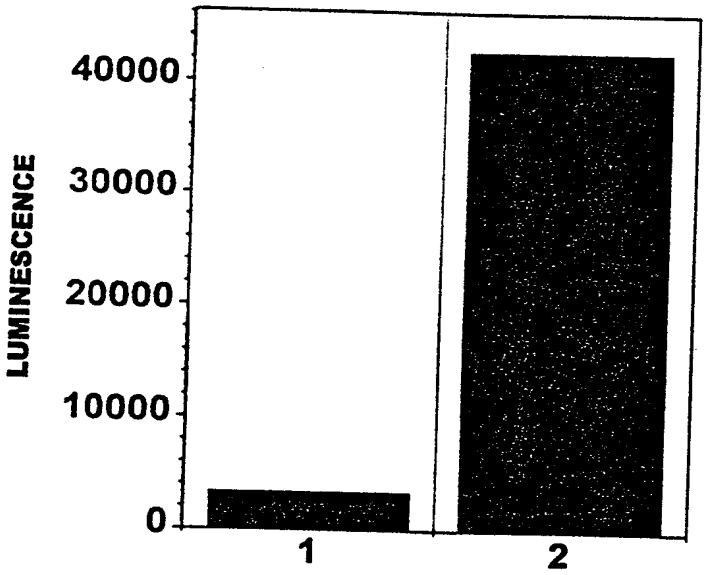


FIGURE 11



1:AED w/o enzyme MutY
2: AED with MutY

FIGURE 12

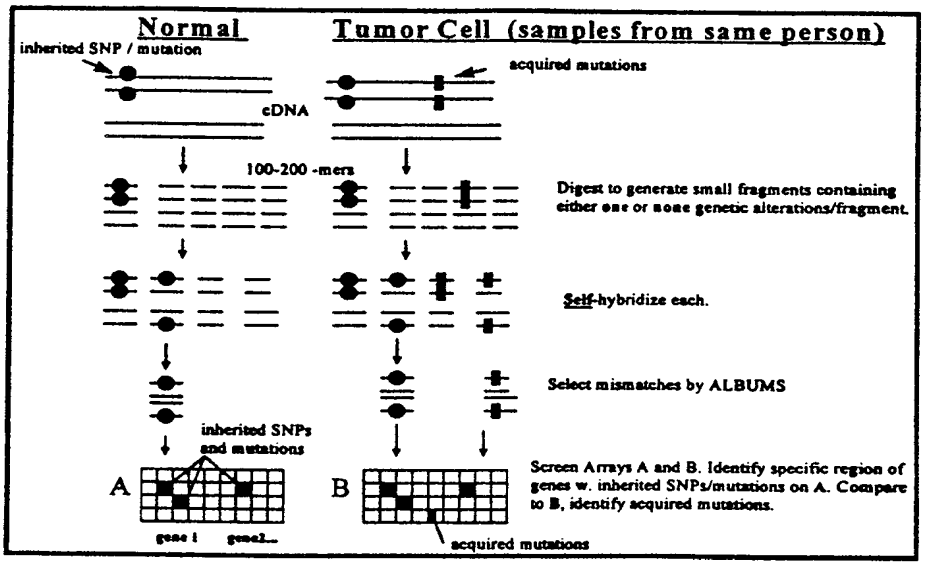
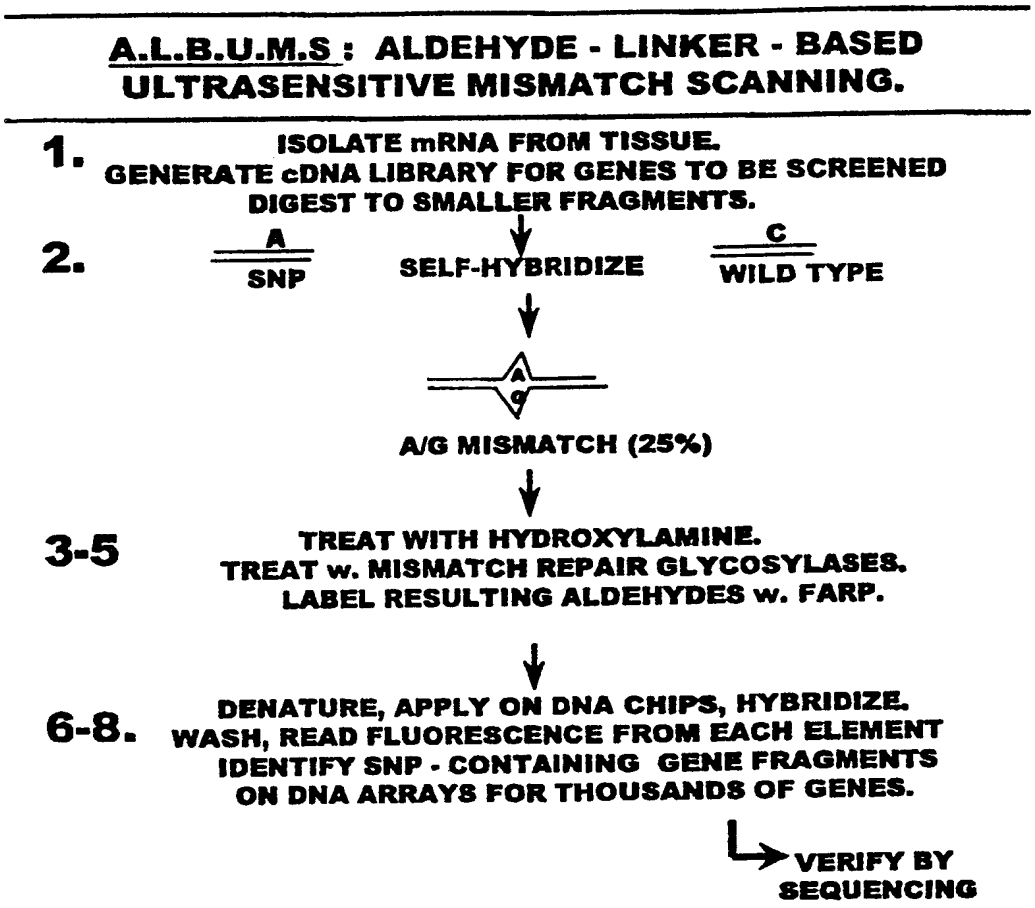

Normal
Tumor Cell (samples from same person)
inherited SNP / mutation
acquired mutations
cDNA
100-200 -mers
Digest to generate small fragments containing either one or none genetic alterations/fragment.
Self-hybridize each.
Select mismatches by ALBUMS
inherited SNPs and mutations
A
B
Screen Arrays A and B. Identify specific region of genes w. inherited SNPs/mutations on A. Compare to B, identify acquired mutations.
gene 1
gene2...
acquired mutations

FIGURE 13



SIMPLIFIED PROTOCOL TO DETECT SNPs AND MUTATIONS ON DNA CHIPS

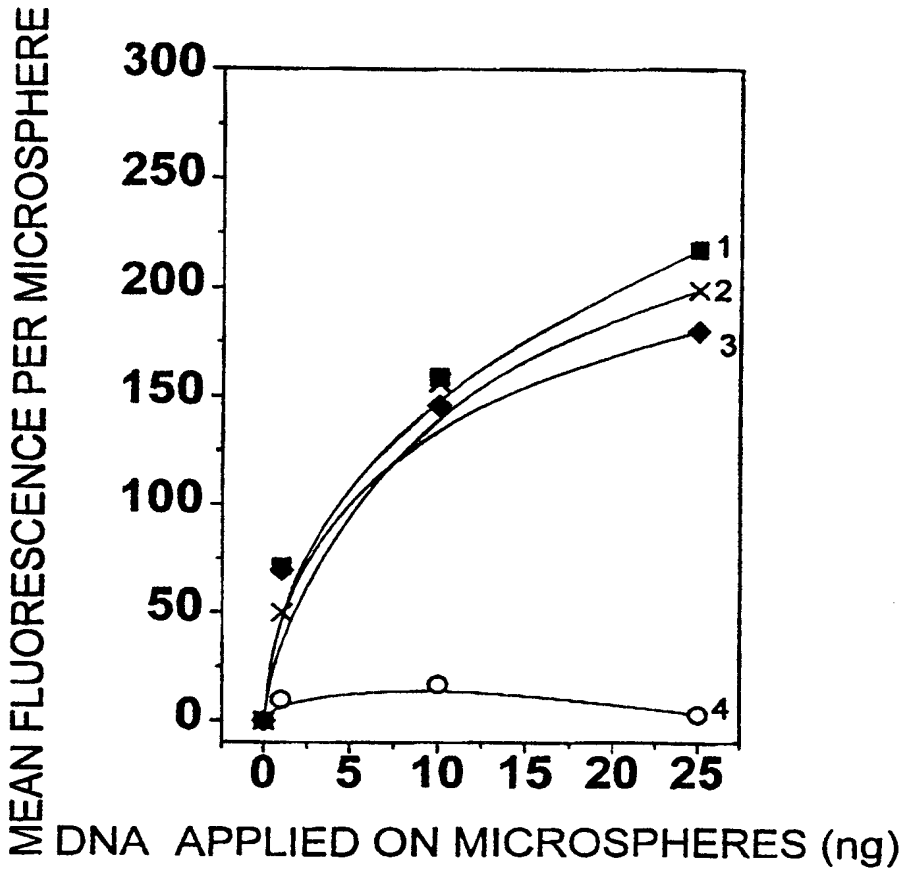
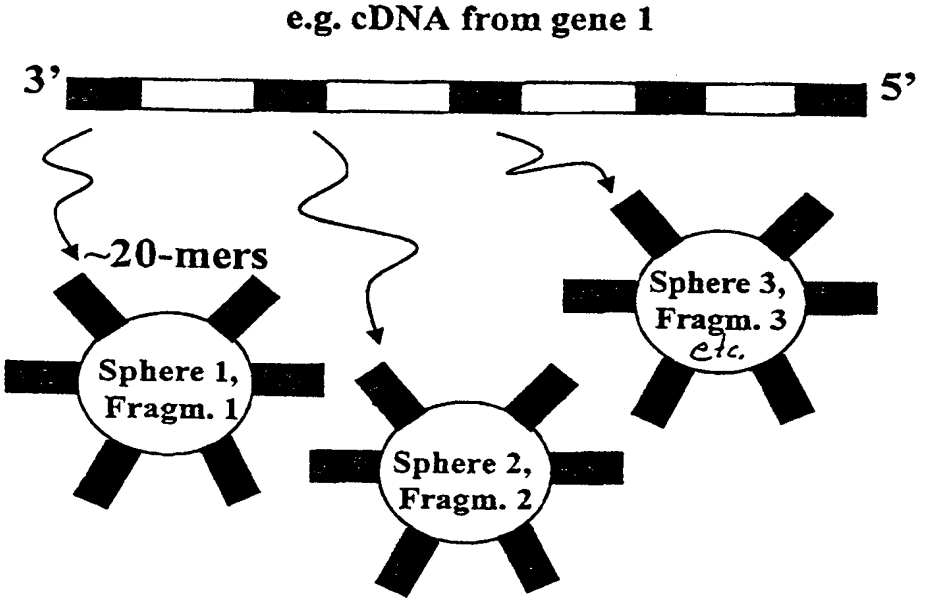

MEAN FLUORESCENCE PER MICROSPHERE
300
250
200
150
100
50
0
0 5 10 15 20 25
DNA APPLIED ON MICROSPHERES (ng)
1
2
3
4

FIGURE 14

MICROSPHERE-BASED MUTATION SCANNING ARRAYS

Each coding sequence is represented by oligonucleotides (■) sampling the mRNA at regular intervals.



THE OUTPUT INFO IS IN BINARY FORMAT

FIGURE 15

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/31177

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) : C12Q 1/68; C12P 19/34
US CL : 435/6, 91.2
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/6, 91.2

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	HACIA, J.G. et al. Strategies for Mutational Analysis of the Large Multiexon ATM Gene Using High-Density Oligonucleotide Arrays. Genome Research. 1998, Vol. 8, pages 1245-1258, entire document.	1-15
Y	LIPSHUTZ, R.J. et al. Using Oligonucleotie Probe Arrays to Access Genetic Diversity. Biotechinques. 1995,Vol. 19, No. 3, pages 442-448, entire document.	1-15
Y	ZIMMERMANN et al. DNA stretching on functionalized gold surfaces. Nucleic Acids Research. 1994, Vol. 22. No. 3, pages 492-497, entire document.	1-15
Y	WO 98/30883 (AFFYMETRIX, INC.) 16 July 1998 (16.05.98), pages 12-14, especially p. 13, lines 26-30.	1-15
Y	US 5, 418,149 A (GELFAND et al) 23 May 1995 (23.05.95), columns 2-4.	1-15
Y	BOTURYN et al. Synthesis of fluorescent probes for the detection of abasic sites in DNA. Tetrahedron. 1997, Vol. 53, No. 15, pages 5485-5492, entire document.	1-15
Y	ASAEDA, A. et al. Highly sensitive assay of DNA abasic sites in mammalian cells-optimization of the aldehyde reactive probe method. Analytica Chimica Acta. 1998, Bol. 365, pages 35-41, see entire document.	1-15
Y	MAKRIGIORGOS, G.M. et al. Fluorescent labelling of abasic sites: a novel methodology to detect closely-spaced damage sites in DNA. Int. J. Radiat. Biol. 1998, Vol. 74, No. 1, pages 99-109, see entire document.	1-15

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

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

"E" earlier application or patent published on or after the international filing date

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search
20 March 2000 (20.03.2000)

Date of mailing of the international search report
06 APR 2000

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703)305-3230

Authorized officer
Juliet C. Einsmann
JOYCE BRIDGERS
PARALEGAL SPECIALIST
CHEMICAL MATT.
Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

IDT Exhibit 1020.7951

INTERNATIONAL SEARCH REPORT

In...national application No.
PCT/US99/31177

C (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,736,330 A (FULTON) 07 April 1998 (07.04.98), columns 2-3.	1-15

Form PCT/ISA/210 (continuation of second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

ernational application No.
PCT/US99/31177

Continuation of B. FIELDS SEARCHED Item 3: US PATENTS (WEST), MEDLINE, BIOSIS, CAPLUS search terms: abasic site, flow cytometry, heteroduplex, array, whole gene array

IDT Exhibit 1020.7953

Espacenet

Bibliographic data: WO2004011665 (A2) — 2004-02-05

METHODS FOR FRAGMENTATION, LABELING AND IMMOBILIZATION OF NUCLEIC ACIDS

Inventor(s): KURN NURITH [US]; DAFFORN GEOFFREY A [US] ± (KURN, NURITH, ; DAFFORN, GEOFFREY, A)

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

Classification: - international: ***C12M1/00; C12N15/09; C12Q1/68; G01N33/58;*** (IPC1-7): C12Q
- cooperative: **C12Q1/68; C12Q1/6806;** B01J2219/00608; B01J2219/0061; B01J2219/00612; B01J2219/00621; B01J2219/00626; B01J2219/00637; C12Q1/6837 more

Application number: WO2003US15825 20030519 Global Dossier

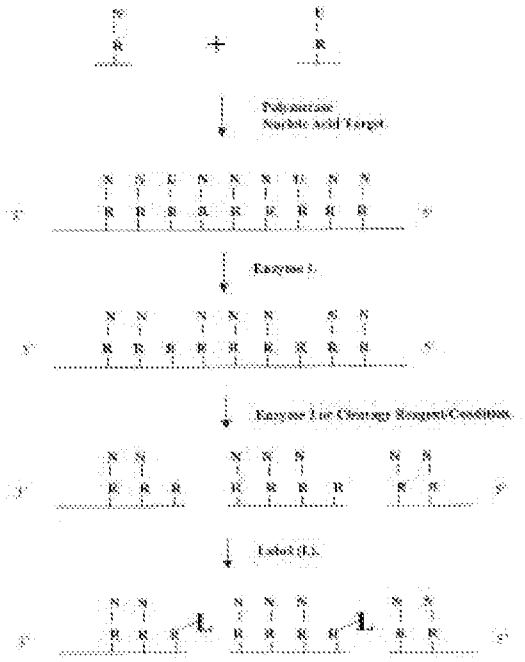
Priority number (s): US20020381457P 20020517

Also published as: WO2004011665 (A3) AU2003279697 (A1) CA2486283 (A1) EP1573056 (A2) EP1573056 (A4) more

Abstract of WO2004011665 (A2)

The invention relates to methods for fragmentation and/or labeling and/or immobilization of nucleic acids. More particularly, the invention relates to methods for fragmentation and/or labeling and/or immobilization of nucleic acids comprising of nucleic acids comprising labeling and/or cleavage and/or immobilization at abasic sites.

IDT Exhibit 1020.7954



IDT Exhibit 1020.7955

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 February 2004 (05.02.2004)

PCT

(10) International Publication Number
WO 2004/011665 A2

(51) International Patent Classification[7]: C12Q

(21) International Application Number:
PCT/US2003/015825

(22) International Filing Date: 19 May 2003 (19.05.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/381,457 17 May 2002 (17.05.2002) US

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

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **KURN, Nurith** [IL/US]; 2876 Lamona Street, Palo Alto, CA 94306 (US). **DAFFORN, Geoffrey, A.** [US/US]; 1669 Christina Drive, Los Altos, CA 94024 (US).

(74) Agents: **COBURN, Cara, M.** et al.; Morrison & Foerster LLP, 755 Page Mill Road, Palo Alto, CA 94304 (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, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, 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 2004/011665 A2

(54) **Title:** METHODS FOR FRAGMENTATION, LABELING AND IMMOBILIZATION OF NUCLEIC ACIDS

(57) **Abstract:** The invention relates to methods for fragmentation and/or labeling and/or immobilization of nucleic acids. More particularly, the invention relates to methods for fragmentation and/or labeling and/or immobilization of nucleic acids comprising of nucleic acids comprising labeling and/or cleavage and/or immobilization at abasic sites.

IDT Exhibit 1020.7956

METHODS FOR FRAGMENTATION, LABELING AND IMMOBILIZATION OF NUCLEIC ACIDS

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the priority benefit of provisional application U.S. Serial No. 60/381,457, filed May 17, 2002, the contents of which is incorporated by reference in its entirety.

TECHNICAL FIELD

[0002] The invention relates to methods for fragmentation and/or labeling and/or immobilization of nucleic acids. More particularly, the invention relates to methods for fragmentation and/or labeling and/or immobilization of nucleic acids comprising labeling and/or cleavage and/or immobilization at abasic sites.

BACKGROUND ART

[0003] Fragmentation and labeling of nucleic acids are important for the analysis of genetic information contained within the nucleic acid sequence. For example, fragmentation and/or labeling are commonly required for detection of sequences by binding of a sample nucleic acid to complementary sequences immobilized on a surface, for example, on a microarray. Cleavage of sample nucleic acid into small fragments (e.g., 50-100 base pairs) facilitates diffusion of nucleic acid onto the surface, and may facilitate hybridization. It is known, for example, that steric and charge hindrance effects increase with the size of nucleic acids that are hybridized. Moreover, cleavage of sample nucleic acids into small fragments may ensure that two sequences of interest in the sample do not appear to bind to the same template nucleic acid simply by virtue of their proximity on the test nucleic acid. Cleavage of nucleic acids also facilitates detection of hybridized nucleic acid when, as in many detection methods, the size of the signal is proportional to the size of the bound fragment and thus, control of fragment size is desirable. Labeling of nucleic acids is necessary in many methods of nucleic acid analysis because there are presently few techniques for direct detection of unlabeled nucleic acid with the requisite sensitivity for analysis on chips. Methods for fragmenting and/or labeling nucleic acids are known in the art. *See, e.g.*, U.S. Pat. Nos. 5,082,830; 4,996,143; 5,688,648; 6,326,142; WO02/090584, and references cited therein.

[0004] Immobilization of nucleic acids to create, for example, microarrays or tagged analytes, is useful for, e.g., detection and analysis of nucleic acids and tagged

analytes. Methods for immobilizing nucleic acids are known in the art *See, e.g.*, U.S. Patent Nos. 5,667,979; 6,077,674; 6,280,935; and references cited therein.

[0005] There is a serious need for improved methods for labeling and/or fragmenting and/or immobilizing nucleic acids to a surface, for example a microarray.

[0006] All references cited herein, including patent applications and publications, are incorporated by reference in their entirety.

SUMMARY OF THE INVENTION

[0007] The invention provides novel methods and kits for labeling and/or fragmenting and/or immobilizing polynucleotides to a substrate.

[0008] In one aspect, the invention provides methods for fragmenting and labeling a polynucleotide, said method comprising: (a) synthesizing a polynucleotide from a template in the presence of at least one non-canonical nucleotide, whereby a polynucleotide comprising a non-canonical nucleotide is generated; (b) contacting the synthesized polynucleotide with an enzyme capable of cleaving a base portion of the non-canonical nucleotide from the synthesized polynucleotide (i.e. cleaving a base portion of a non-canonical nucleotide with an enzyme capable of cleaving a base portion of a non-canonical nucleotide), whereby an abasic site is created; (c) cleaving a phosphodiester backbone at the abasic site; and (d) contacting the synthesized polynucleotide with an agent capable of labeling the abasic site (i.e. labeling an abasic site), whereby a labeled polynucleotide fragment is generated.

[0009] In one aspect, the invention provides methods for fragmenting and labeling a polynucleotide, said method comprising (a) contacting a polynucleotide comprising a non-canonical nucleotide with an enzyme capable of cleaving a base portion of the non-canonical nucleotide, whereby an abasic site is created, wherein the polynucleotide comprising a non-canonical nucleotide is synthesized from a template in the presence of at least one non-canonical nucleotide; (b) cleaving a phosphodiester backbone at the abasic site; and (c) contacting the polynucleotide with an agent capable of labeling the abasic site (i.e. labeling at the abasic site); whereby a labeled polynucleotide fragment is generated.

[0010] In another aspect, the invention provides methods for fragmenting and labeling a polynucleotide, said method comprising (a) cleaving a phosphodiester backbone at an abasic site of a polynucleotide comprising the abasic site, wherein the polynucleotide comprising the abasic site is generated by contacting a polynucleotide comprising a non-canonical nucleotide with an enzyme capable of cleaving a base portion of the non-canonical nucleotide, whereby an abasic site is created, wherein the polynucleotide

IDT Exhibit 1020.7958

comprising a non-canonical nucleotide is synthesized from a template in the presence of at least one non-canonical nucleotide; and (b) contacting the polynucleotide with an agent capable of labeling the abasic site; whereby labeled fragments of the polynucleotide are generated.

[0011] In another aspect, the invention provides methods for fragmenting and labeling a polynucleotide, said method comprising contacting a polynucleotide comprising an abasic site with an agent capable of labeling the abasic site; wherein the polynucleotide is generated by cleaving a phosphodiester backbone at an abasic site of a polynucleotide comprising the abasic site, wherein the polynucleotide comprising the abasic site is generated by contacting a polynucleotide comprising a non-canonical nucleotide with an enzyme capable of cleaving a base portion of the non-canonical nucleotide, whereby an abasic site is created, wherein the polynucleotide comprising a non-canonical nucleotide is synthesized from a template in the presence of at least one non-canonical nucleotide; whereby labeled fragments of the polynucleotide are generated.

[0012] In another aspect, the invention provides method for fragmenting and labeling a polynucleotide comprising: (a) incubating a reaction mixture, said reaction mixture comprising: (i) a template and (ii) a non-canonical nucleotide; wherein the incubation is under conditions that permit formation of a polynucleotide comprising a non-canonical nucleotide; (b) incubating a reaction mixture, said reaction mixture comprising: (i) a polynucleotide comprising a non-canonical nucleotide; and (ii) an agent capable of specifically cleaving a base portion of a non-canonical nucleotide; wherein the incubation is under conditions that permit cleavage of the base portion of the non-canonical nucleotide, whereby a polynucleotide comprising an abasic site is generated; (c) incubating a reaction mixture, said reaction mixture comprising: (i) a polynucleotide comprising an abasic site; and (ii) an agent capable of effecting (generally, specific) cleavage of a phosphodiester backbone at the abasic site; wherein the incubation is under conditions that permit cleavage of the phosphodiester backbone at the abasic site; whereby fragments of the polynucleotide are generated; (d) incubating a reaction mixture, said reaction mixture comprising: (i) a polynucleotide comprising an abasic site; and (ii) an agent capable of labeling the abasic site; wherein the incubation is under conditions that permit labeling at the abasic site; whereby labeled fragments are generated.

[0013] In another aspect, the invention provides methods for labeling and fragmenting a polynucleotide, said method comprising: (a) incubating a reaction mixture, said reaction mixture comprising: (i) a template and (ii) a non-canonical nucleotide; wherein the incubation is under conditions that permit formation of a polynucleotide comprising a non-canonical nucleotide; (b) incubating a reaction mixture, said reaction

IDT Exhibit 1020.7959

mixture comprising: (i) the polynucleotide comprising the non-canonical polynucleotide; (ii) an enzyme capable of cleaving a base portion of a non-canonical nucleotide; (iii) an agent capable of cleaving a polynucleotide at the abasic site; wherein the incubation is under conditions that permit cleavage of a base portion of a non-canonical nucleotide and optionally, cleavage of the polynucleotide at the abasic site; whereby fragments of the polynucleotide are generated; and (b) incubating a reaction mixture, said reaction mixture comprising: (i) a polynucleotide fragment comprising an abasic site; and (ii) an agent capable of labeling the abasic site; wherein the incubation is under conditions that permit labeling at the abasic site; whereby a labeled fragment is generated.

[0014] In another aspect, the invention provides the invention provides methods for labeling and fragmenting a polynucleotide, said method comprising (a) incubating a reaction mixture, said reaction mixture comprising: (i) a template; (ii) a non-canonical nucleotide; (iii) an enzyme capable of cleaving a base portion of a non-canonical nucleotide; and (iv) an agent capable of cleaving a polynucleotide at the abasic site; wherein the incubation is under conditions that permit formation of a polynucleotide comprising a non-canonical nucleotide, cleavage of a base portion of a non-canonical nucleotide and cleavage of the polynucleotide at the abasic site; whereby fragments of the polynucleotide are generated; and (b) incubating a reaction mixture, said reaction mixture comprising: (i) a polynucleotide fragment comprising an abasic site or optionally, fragments of a polynucleotide comprising an abasic site; and (ii) an agent capable of labeling the abasic site; wherein the incubation is under conditions that permit labeling at the abasic site; whereby a labeled fragment is generated.

[0015] 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. The reaction mixtures may be combined (thus reducing the number of incubations) in any way, with one or more reaction mixtures above combined.

[0016] Accordingly, in some embodiments, synthesizing a polynucleotide comprising a non-canonical nucleotide and cleaving a base portion of a non-canonical nucleotide are conducted in the same reaction mixture. In other embodiments, synthesizing a polynucleotide comprising a non-canonical nucleotide, cleaving a base portion of a non-canonical nucleotide, and cleaving the backbone at an abasic site are conducted in the same reaction mixture. In still another embodiment, synthesizing a polynucleotide comprising a non-canonical nucleotide and cleaving a base portion of a non-canonical nucleotide are conducted in the same reaction mixture, and cleaving the

IDT Exhibit 1020.7960

backbone at an abasic site and labeling at the abasic site are conducted in same reaction mixture. In another embodiment, synthesizing a polynucleotide comprising a non-canonical nucleotide, cleaving a base portion of a non-canonical nucleotide, cleaving the backbone at an abasic site, and labeling at the abasic site are conducted in same reaction mixture. In other embodiments, synthesizing a polynucleotide comprising a non-canonical nucleotide, cleaving a base portion of a non-canonical nucleotide, cleaving the backbone at an abasic site, and labeling at the abasic site are conducted in same reaction mixture. In other embodiments, cleaving a base portion of a non-canonical nucleotide, and cleaving the backbone at an abasic site are conducted in the same reaction mixture. In other embodiments, cleaving the backbone at an abasic site, and labeling at the abasic site are conducted in the same reaction mixture. In another embodiment, cleaving a base portion of a non-canonical nucleotide and labeling at the abasic site are conducted in the same reaction mixture. In another embodiment, synthesizing a polynucleotide comprising a non-canonical nucleotide, cleaving a base portion of a non-canonical nucleotide, and labeling at an abasic site are conducted in the same reaction mixture. 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. As explained herein, labeling can occur before fragmentation (i.e. cleavage of the phosphodiester backbone at an abasic site), fragmentation can occur before labeling, or fragmentation and labeling can occur simultaneously.

[0017] In another aspect, the invention provides methods for labeling a polynucleotide, said method comprising: (a) synthesizing a polynucleotide from a template in the presence of at least one non-canonical nucleotide, whereby a polynucleotide comprising a non-canonical nucleotide is generated; (b) contacting the synthesized polynucleotide with an enzyme capable of effecting cleavage of a base portion of the non-canonical nucleotide from the synthesized polynucleotide, whereby an abasic site is created; (c) contacting the synthesized polynucleotide with an agent capable of labeling the abasic site; whereby the synthesized polynucleotide is labeled.

[0018] In one aspect, the invention provides methods for labeling a polynucleotide, said method comprising: (a) contacting a polynucleotide comprising a non-canonical nucleotide with an enzyme capable of cleaving a base portion of the non-canonical nucleotide, whereby an abasic site is created, wherein the polynucleotide comprising a non-canonical nucleotide is synthesized from a template in the presence of at least one non-canonical nucleotide; (b) contacting the polynucleotide with an agent capable of labeling the abasic site; whereby the polynucleotide is labeled.

IDT Exhibit 1020.7961

[0019] In another aspect, the invention provides methods for labeling a polynucleotide, said method comprising contacting a polynucleotide comprising an abasic site with an agent capable of labeling the abasic site; wherein the polynucleotide comprising the abasic site is generated by contacting a polynucleotide comprising a non-canonical nucleotide with an enzyme capable of cleaving a base portion of the non-canonical nucleotide, whereby an abasic site is created, wherein the polynucleotide comprising a non-canonical nucleotide is synthesized from a template in the presence of at least one non-canonical nucleotide; whereby the polynucleotide is labeled.

[0020] In another aspect, the invention provides methods for labeling a polynucleotide, said method comprising: (a) preparing an aminooxy derivative of Alexa Fluor 555; and (b) contacting a polynucleotide comprising an abasic site (prepared using methods described herein) with the aminooxy derivative of Alexa Fluor 555; whereby the polynucleotide is labeled. In another aspect, the invention provides methods for labeling a polynucleotide comprising contacting a polynucleotide comprising an abasic site (prepared using methods described herein) with an aminooxy derivative of Alex Fluor 555; whereby the polynucleotide is labeled.

[0021] In some embodiments of the methods of generating polynucleotides immobilized to a surface (i.e., a substrate), the polynucleotide comprising an abasic site is labeled at an abasic site.

[0022] In another aspect, the invention provides methods for labeling a polynucleotide comprising: (a) incubating a reaction mixture, said reaction mixture comprising: (i) a template and (ii) a non-canonical nucleotide; wherein the incubation is under conditions that permit formation of a polynucleotide comprising a non-canonical nucleotide; (b) incubating a reaction mixture, said reaction mixture comprising: (i) a polynucleotide comprising a non-canonical nucleotide; and (ii) an agent capable of specifically cleaving a base portion of a non-canonical nucleotide; wherein the incubation is under conditions that permit cleavage of the base portion of the non-canonical nucleotide, whereby a polynucleotide comprising an abasic site is generated; (c) incubating a reaction mixture, said reaction mixture comprising: (i) a polynucleotide comprising an abasic site; and (ii) an agent capable of labeling the abasic site; wherein the incubation is under conditions that permit labeling at the abasic site; whereby labeled polynucleotides are generated.

[0023] 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. The reaction mixtures may be combined (thus reducing

IDT Exhibit 1020.7962

the number of incubations) in any way, with one or more reaction mixtures above combined.

[0024] Accordingly, in some embodiments, synthesizing a polynucleotide comprising a non-canonical nucleotide and cleaving a base portion of a non-canonical nucleotide are conducted in the same reaction mixture. In other embodiments, synthesizing a polynucleotide comprising a non-canonical nucleotide, cleaving a base portion of a non-canonical nucleotide, and labeling at the abasic site are conducted in same reaction mixture. In other embodiments, cleaving a base portion of a non-canonical nucleotide, and labeling at the abasic site are conducted in same reaction mixture. 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.

[0025] In another aspect, the invention provides methods for labeling and optionally fragmenting a polynucleotide, said method comprising: (a) synthesizing a polynucleotide from a polynucleotide template in the presence of a non-canonical nucleotide, whereby a polynucleotide comprising the non-canonical nucleotide is generated; (b) cleaving a base portion of a non-canonical nucleotide from the synthesized polynucleotide with an enzyme capable of cleaving the base portion of the non-canonical nucleotide, whereby an abasic site is generated; (c) optionally, cleaving a phosphodiester backbone of the polynucleotide comprising the abasic site at the abasic site; and (d) labeling the polynucleotide or the fragment of the polynucleotide at the abasic site; whereby a labeled polynucleotide, or optionally, a labeled polynucleotide fragment is generated.

[0026] In another aspect, the invention provides methods for labeling and optionally fragmenting a polynucleotide, said method comprising: (a) incubating a reaction mixture, said reaction mixture comprising: (i) a polynucleotide template; and (ii) a non-canonical nucleotide; wherein the incubation is under conditions that permit synthesis of a polynucleotide comprising the non-canonical nucleotide, whereby a polynucleotide comprising the non-canonical nucleotide is generated; (b) incubating a reaction mixture, said reaction mixture comprising: (i) the polynucleotide comprising the non-canonical nucleotide; and (ii) an enzyme capable of cleaving a base portion of the non-canonical nucleotide, wherein the incubation is under conditions that permit cleavage of the base portion of the non-canonical nucleotide, whereby a polynucleotide comprising an abasic site is generated; (c) optionally incubating a reaction mixture, said reaction mixture comprising: (i) the polynucleotide comprising the abasic site; and (ii) an agent capable of cleaving a phosphodiester backbone of the polynucleotide comprising the abasic site at the

IDT Exhibit 1020.7963

abasic site, wherein the incubation is under conditions that permit cleavage of the phosphodiester backbone of the polynucleotide at the abasic site, whereby a fragment of the polynucleotide is generated; (d) incubating a reaction mixture, said reaction mixture comprising: (i) the polynucleotide comprising the abasic site or optionally, the fragment of the polynucleotide comprising the abasic site; and (ii) an agent capable of labeling the abasic site, wherein the incubation is under conditions that permit labeling at the abasic site; whereby a labeled polynucleotide or optionally, a labeled polynucleotide fragment, is generated.

[0027] In another aspect, the invention provides methods for labeling and optionally fragmenting a polynucleotide, said method comprising (a) incubating a reaction mixture, said reaction mixture comprising: (i) the polynucleotide comprising the non-canonical polynucleotide of step (a) of claim 1; (ii) an enzyme capable of cleaving a base portion of the non-canonical nucleotide; and (iii) optionally, an agent capable of cleaving a phosphodiester backbone of the polynucleotide comprising the abasic site at the abasic site, wherein the incubation is under conditions that permit cleavage of the base portion of the non-canonical nucleotide and optionally, cleavage of the phosphodiester backbone of the polynucleotide at the abasic site; whereby polynucleotide comprising the abasic site, or optionally, a fragment of the polynucleotide comprising the abasic site, is generated; and (b) incubating a reaction mixture, said reaction mixture comprising: (i) the polynucleotide comprising the abasic site or optionally, the fragment of the polynucleotide comprising the abasic site; and (ii) an agent capable of labeling the abasic site, wherein the incubation is under conditions that permit labeling at the abasic site, whereby a labeled polynucleotide or optionally, a labeled fragment of the polynucleotide, is generated.

[0028] In another aspect, the methods of the invention provide methods for generating polynucleotides immobilized to a surface, said methods comprising immobilizing a polynucleotide comprising an abasic site to a surface, wherein the polynucleotide is immobilized at the abasic site. In some embodiments, the polynucleotide comprising an abasic site is generated by contacting a polynucleotide comprising a non-canonical nucleotide with an enzyme capable of cleaving a base portion of the non-canonical nucleotide from the polynucleotide, whereby an abasic site is created. In further embodiments, the polynucleotide comprising a non-canonical nucleotide is synthesized from a template in the presence of at least one non-canonical nucleotide.

[0029] In another aspect, the methods of the invention provide methods for generating polynucleotides immobilized to a surface, said method comprising: (a) contacting a polynucleotide comprising a non-canonical nucleotide with an enzyme capable of cleaving a base portion of the non-canonical nucleotide from the

IDT Exhibit 1020.7964

polynucleotide, whereby an abasic site is created; (b) optionally cleaving a phosphodiester backbone at the abasic site; whereby fragments of the polynucleotide are generated; and (c) immobilizing the polynucleotide comprising an abasic site, or fragments thereof, to a surface, wherein the polynucleotide is immobilized at an abasic site. In some embodiments, the polynucleotide is synthesized from a template in the presence of at least one non-canonical nucleotide.

[0030] In another aspect, the invention provides methods for generating a polynucleotide immobilized to a surface, said methods comprising: (a) cleaving a phosphodiester backbone at an abasic site of a polynucleotide comprising the abasic site; whereby fragments of the polynucleotide are generated; and (b) immobilizing the fragments of the polynucleotide to a surface, wherein the polynucleotide is immobilized at the abasic site. In some embodiments, the polynucleotide comprising an abasic site is generated by contacting a polynucleotide comprising a non-canonical nucleotide with an enzyme capable of cleaving a base portion of the non-canonical nucleotide from the polynucleotide, whereby an abasic site is created. In further embodiments, the polynucleotide comprising a non-canonical nucleotide is synthesized from a template in the presence of at least one non-canonical nucleotide.

[0031] In another aspect, the invention provides methods for immobilizing a polynucleotide comprising: (a) incubating a reaction mixture, said reaction mixture comprising: (i) a polynucleotide comprising a non-canonical nucleotide; and (ii) an agent capable of specifically cleaving a base portion of a non-canonical nucleotide; wherein the incubation is under conditions that permit cleavage of the base portion of the non-canonical nucleotide, whereby a polynucleotide comprising an abasic site is generated; (b) optionally incubating a reaction mixture, said reaction mixture comprising: (i) a polynucleotide comprising an abasic site; and (ii) an agent capable of effecting specific cleavage of a phosphodiester backbone at the abasic site; wherein the incubation is under conditions that permit cleavage of the phosphodiester backbone at the abasic site; whereby fragments of the polynucleotide are generated; (c) incubating a reaction mixture, said reaction mixture comprising: (i) a polynucleotide, or fragment thereof, comprising an abasic site; and (ii) a surface (i.e., a substrate); and (iii) an agent capable of immobilizing the polynucleotide, or fragment thereof, comprising the abasic site to the surface at the abasic site; wherein the incubation is under conditions that permit immobilization of the polynucleotide, or fragment thereof, to the surface at the abasic site; whereby immobilized polynucleotides, or fragments thereof, are generated. In some embodiments, the polynucleotide is synthesized from a template in the presence of at least one non-canonical nucleotide.

IDT Exhibit 1020.7965

[0032] 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. The reaction mixtures may be combined (thus reducing the number of incubations) in any way, with one or more reaction mixtures above combined. 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

[0033] Various embodiments of the methods of the inventions are described herein. For example, in embodiments involving synthesis of a polynucleotide comprising a non-canonical nucleotide from a template, the synthesizing can be by PCR, primer extension, reverse transcription, DNA replication, strand displacement amplification (SDA), multiple displacement amplification (MDA), and the like. In some embodiments, the polynucleotide is synthesized using single primer isothermal amplification, for example, wherein a polynucleotide sequence complementary to a target polynucleotide is amplified using methods comprising the following steps of: (a) hybridizing a single stranded DNA template comprising the target sequence with a composite primer, said composite primer comprising a RNA portion and a 3' DNA portion; (b) optionally hybridizing a polynucleotide comprising a termination polynucleotide sequence to a region of the template which is 5' with respect to hybridization of the composite primer to the template; (c) extending the composite primer with DNA polymerase; and (d) cleaving the RNA portion of the annealed composite primer with an enzyme that cleaves RNA from an RNA/DNA hybrid such that another composite primer hybridizes to the template and repeats primer extension by strand displacement, whereby multiple copies of the complementary sequence of the target sequence are produced. In another embodiment, the polynucleotide is synthesized using methods comprising the following steps of: (a) extending a composite primer in a complex comprising (i) a polynucleotide template; and (ii) the composite primer, said composite primer comprising an RNA portion and a 3' DNA portion, wherein the polynucleotide template is hybridized to the composite primer; and (b) cleaving the RNA portion of the annealed composite primer with an enzyme that cleaves RNA from an RNA/DNA hybrid such that another composite primer hybridizes to the template and repeats primer extension by strand displacement, whereby multiple copies of the complementary sequence of the target sequence are produced. In some embodiments, the RNA portion of the composite primer is 5' with respect to the 3' DNA portion, the 5' RNA portion is adjacent to the 3' DNA portion, the RNA portion of the

IDT Exhibit 1020.7966

composite primer consists of about 10 to about 20 nucleotides and the DNA portion of the composite primer consists of about 7 to about 20 nucleotides.

[0034] In other embodiments, the polynucleotide is synthesized using Ribo-SPIA™, for example wherein multiple copies of a polynucleotide sequence complementary to an RNA sequence of interest (template) are generated using methods comprising the following 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 and a RNA-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 some embodiment, RNA in a complex of step (b) is cleaved with an agent (such as heat or basic conditions) that cleaves RNA from an RNA/DNA hybrid.

[0035] In some embodiments, the polynucleotide that is synthesized is single stranded. In other embodiments, the polynucleotide that is synthesized is double-stranded. In still other embodiments, the polynucleotide that is synthesized is partially double stranded. In still other embodiments, the polynucleotide that is synthesized comprises a cDNA. In still other embodiments, the template comprises RNA, mRNA, genomic DNA, plasmid DNA, synthetic DNA, cDNA. In other embodiments, the template comprises a cDNA library, a genomic library, or a subtractive hybridization library. In still other embodiments, the polynucleotide comprising a non-canonical nucleotide is synthesized using a labeled primer. In still other embodiments, the polynucleotide comprising a non-canonical nucleotide is synthesized using a primer comprising a non-canonical nucleotide. In other embodiments, the polynucleotide comprising a non-canonical nucleotide is synthesized in the presence of two or more different non-canonical nucleotides, whereby a polynucleotide comprising two or more different non-canonical nucleotide is synthesized.

IDT Exhibit 1020.7967

In other embodiments, the polynucleotide comprising a non-canonical nucleotide is synthesized from two or more different polynucleotide templates.

[0036] In some embodiments, the non-canonical nucleotide is dUTP. In other embodiments, the non-canonical nucleotide is dUTP and the enzyme capable of cleaving a base portion of the non-canonical nucleotide from the synthesized polynucleotide is Uracil N-Glycosylase (interchangeably termed "UNG").

[0037] In embodiments involving fragmentation, the phosphodiester backbone can be cleaved by an agent, such as an enzyme or an amine, capable of effecting cleavage of a phosphodiester backbone at an abasic site. In some embodiments, the enzyme is E. coli Endonuclease IV. In other embodiments, the agent is N, N'-dimethylethylenediamine. In still other embodiments, the agent is heat, basic conditions, or acidic conditions.

[0038] In embodiments involving fragmentation, the fragments can be about 10, about 15, about 20, about 25, about 30 about 35 about 40, about 50, about 65, about 75, about 85, about 100, about 125, about 150, about 175, about 200, about 225, about 250, about 300, about 350, about 400, about 450, about 500, about 550, about 600, about 650 or more nucleotides in length. In some embodiments, the fragments can be at least about 15, about 20, about 25, about 30 about 35 about 40, about 50, about 65, about 75, about 85, about 100, about 125, about 150, about 175, about 200, about 225, about 250, about 300, about 350, about 400, about 450, about 500, about 550, about 600, about 650 or more nucleotides in length. In other embodiments, the fragments can be less than about 15, about 20, about 25, about 30 about 35 about 40, about 50, about 65, about 75, about 85, about 100, about 125, about 150, about 175, about 200, about 225, about 250, about 300, about 350, about 400, about 450, about 500, about 550, about 600, about 650 or more nucleotides in length. It is understood that these fragment lengths may represent an average size in the population of fragments generated using the methods of the invention.

[0039] In some embodiments, the fragments comprise an abasic site at the 3' end (terminus). In other embodiments, the fragments comprise an abasic site at the 5' end (terminus). In still other embodiments, the fragments comprise both abasic sites at the 3' ends and abasic sites at the 5' ends. It is understood that a polynucleotide fragment may additionally comprise internal abasic sites (i.e., abasic sites that are not at the 3' or 5' end of the fragment), as when, for example, fragmentation does not occur at every abasic site in a polynucleotide.

[0040] In embodiments involving labeling, the polynucleotide comprising a non-canonical nucleotide, or fragments thereof, is labeled at an abasic site, whereby a polynucleotide (or polynucleotide fragment) comprising a label is generated. In some embodiments, the polynucleotide, or fragments thereof, comprising an abasic site is

IDT Exhibit 1020.7968

contacted with an agent capable of labeling the abasic site. In various embodiments, the detectable moiety (label) is covalently or non-covalently associated or directly or indirectly associated with an abasic site. In some embodiments, the label is directly or indirectly detectable. In some embodiments, the label comprises an organic molecule, a hapten, or a particle (such as a polystyrene bead). In some embodiments, the label is detected using antibody binding, biotin binding, or via fluorescence or enzyme activity. In some embodiments, the detectable signal is amplified. In some embodiments, the detectable moiety comprises an organic molecule. In some embodiments, the label reacts with an aldehyde residue at the abasic site. In other embodiments, the label comprises a reactive group selected from: a hydrazine, or a hydroxylamine. In some embodiments, the label is 5-(((2-(carbohydrazino)-methyl)thio)acetyl)aminofluorescein, aminooxyacetyl hydrazide ("FARP"). In another embodiment, the label is N-(aminooxyacetyl)-N'-(D-biotinoyl) hydrazine, trifluoroacetic acid salt ("ARP"). In yet another embodiment, the label is Alexa 555. In yet another embodiment, the label is an aminooxy derivative of Alexa Fluor 555.

[0041] In another aspect, the invention provides an aminooxy derivative of Alexa Fluor 555, wherein the aminooxy derivative is generated as disclosed herein.

[0042] In some embodiments involving immobilization, the polynucleotide or fragment thereof, is immobilized on a substrate (used interchangeably herein with "surface") at the abasic site. In some embodiments, the substrate comprises a solid or semi-solid support. In some embodiments, the substrate is a microarray. In other embodiments, 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 other metals), and optical fiber. In still other embodiments, the polynucleotide, or fragment thereof, 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.

[0043] In other embodiments, a substrate which is an analyte is selected from the group consisting of a protein, a polypeptide, a peptide, a carbohydrate, an organic molecule, an inorganic molecule, a cell, a microorganism, and fragments and products thereof. In other embodiments, the analyte is selected from the group consisting of a polypeptide, an antibody, an organic molecule and an inorganic molecule.

[0044] The methods are applicable to generating labeled polynucleotides, labeled polynucleotide fragments, or immobilized polynucleotides (or fragments thereof), or labeled immobilized polynucleotides (or fragments thereof) from any polynucleotide target, including, for example, mRNA, genomic DNA, cDNA, cloned DNA, and synthetic

IDT Exhibit 1020.7969

DNA. 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. Methods of the invention encompass embodiments in which later, "downstream" steps are an initial step. 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, synthesizing a polynucleotide comprising a non-canonical nucleotide and cleaving a base portion of a non-canonical nucleotide are conducted in the same reaction mixture.. In other embodiments, synthesizing a polynucleotide comprising a non-canonical nucleotide, cleaving a base portion of a non-canonical nucleotide, and labeling at the abasic site are conducted in same reaction mixture. In other embodiments, synthesizing a polynucleotide comprising a non-canonical nucleotide, cleaving a base portion of a non-canonical nucleotide, labeling at the abasic site are conducted in same reaction mixture, and immobilizing at an abasic site are conducted in the same reaction mixture. In other embodiments, synthesizing a polynucleotide comprising a non-canonical nucleotide, cleaving a base portion of a non-canonical nucleotide, and immobilizing at an abasic site are conducted in the same reaction mixture. In other embodiments, cleaving a base portion of a non-canonical nucleotide, and labeling at the abasic site are conducted in same reaction mixture. In other embodiments, cleaving a base portion of a non-canonical nucleotide, and immobilizing at the abasic site are conducted in same reaction mixture. In other embodiments, labeling at an abasic site and immobilizing at an abasic site are conducted in the same reaction mixture. 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.

[0045] The invention also provides methods which employ (usually, analyze) the products of the labeling and/or labeling and/or immobilization methods of the invention, such as methods of detecting the presence or absence of nucleic acid sequence mutations; methods to characterize (for example, detect presence or absence of and/or quantify) a polynucleotide template; methods of preparing a hybridization probe; methods of hybridization using the hybridization probes; methods of detection using the hybridization probe; methods of determining a gene expression profile; method of comparative hybridization; methods of identifying a polynucleotide; and methods of preparing a subtractive hybridization probe.

IDT Exhibit 1020.7970

[0046] In one aspect, the invention provides methods of detecting presence or absence of a mutation in a template, comprising: (a) generating a labeled polynucleotide, or fragments thereof, by any of the methods described herein; and (b) analyzing the labeled polynucleotide, or fragments thereof, whereby presence or absence of a mutation is detected. In some embodiments, the labeled polynucleotide, or fragments thereof, is compared to a labeled reference template, or fragments thereof. Step (b) of analyzing the labeled polynucleotide, or fragments thereof, whereby presence or absence of a mutation is detected, can be performed by any method known in the art. In some embodiments, probes for detecting mutations are provided as a microarray.

[0047] In another aspect, the invention provides methods of characterizing a template, comprising: (a) generating a labeled polynucleotide, or fragments thereof, by any of the methods described herein; and (b) analyzing the polynucleotide, or fragments thereof. Step (b) of analyzing the labeled polynucleotide, or fragments thereof, can be performed by any method known in the art or described herein, for example by detecting and/or quantifying labeled polynucleotide, or fragments thereof, that are hybridized to a 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 some embodiments, step (b) of analyzing the labeled polynucleotide, or fragment thereof, comprises determining amount of said products, whereby the amount of the template present in a sample is quantified. In other embodiments, step (b) of analyzing the labeled polynucleotide, or fragment thereof, comprises determining the sequence of the labeled polynucleotide (or fragments thereof) for example, using sequencing by hybridization.

[0048] In another aspect, the invention provides methods for identifying a polynucleotide, comprising: (a) generating a labeled polynucleotide, or fragments thereof, from a polynucleotide template by any of the methods described herein; and (b) analyzing the polynucleotide, or fragments thereof, whereby the polynucleotide is identified. In some embodiments, step (b) of identifying the polynucleotide comprises hybridizing the labeled polynucleotide or fragments thereof to at least one probe.

[0049] In another aspect, the invention provides methods of determining gene expression profile in a sample, said method comprising: (a) generating a labeled polynucleotide, or fragments thereof, by any of the methods described herein; and (b)

IDT Exhibit 1020.7971

determining amount of labeled polynucleotide, or fragments thereof, generated from each template polynucleotide, wherein each said amount is indicative of amount of each template in the sample, whereby the gene expression profile in the sample is determined.

[0050] Any of these applications can use any of the methods (including various components and various embodiments of any of the components) as described herein.

[0051] The invention also provides compositions, kits, complexes, reaction mixtures and systems comprising various components (and various combinations of the components) used in the methods described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[0052] FIGURE 1: shows a diagrammatic illustration of a method for fragmenting and labeling a nucleic acid. "R" indicates a nucleotide residue.

[0053] FIGURE 2: shows a diagrammatic illustration of a method for labeling a nucleic acid. "R" indicates a nucleotide residue.

[0054] FIGURE 3: shows a diagrammatic illustration of a method for immobilizing a nucleic acid to a surface. "R" indicates a nucleotide residue.

[0055] FIGURE 4: shows a gel showing fragmented labeled polynucleotide fragments generated by (1) creating an abasic site by cleaving a base portion of a non-canonical nucleotide present in an oligonucleotide, (2) cleaving the phosphodiester backbone at the abasic site, and (3) labeling the abasic site using an agent capable of specifically labeling an abasic site.

[0056] FIGURE 5: shows a gel showing labeled polynucleotides generated by (1) creating an abasic site by cleaving a base portion of a non-canonical nucleotide present in an oligonucleotide, and (2) labeling the abasic site using an agent capable of specifically labeling an abasic site.

[0057] FIGURE 6: shows a gel showing labeled polynucleotide fragments generated according to the fragmentation and labeling methods of the invention, wherein the synthesized polynucleotides were amplified using the single primer amplification methods described in Kurn, U.S. Patent Publication No. 2003/0087251 A1, which is hereby incorporated by reference in its entirety.

[0058] FIGURE 7: shows an electropherogram showing labeled polynucleotide fragments generated according to the fragmentation and labeling methods of the invention, wherein the synthesized polynucleotides were amplified using the single primer amplification methods described in Kurn, U.S. Patent Publication No. 2003/0087251 A1, and the UNG treatment and amine fragmentation steps were performed in the same reaction mixture.

IDT Exhibit 1020.7972

[0059] FIGURE 8: shows a graph depicting the correlation observed between two populations of labeled fragments prepared from two independent RiboSPIA™ amplification reactions using the single primer amplification methods described in Kurn, U.S. Patent Publication No. 2003/0087251 A1. Each sample was hybridized to two identical arrays, and intensities observed for each spot on the arrays are plotted against each other. The Pearson correlation coefficient was calculated, and a statistically significant correlation between duplicate arrays was observed (correlation coefficient r = 0.98).

MODES FOR CARRYING OUT THE INVENTION

Methods of the invention

Methods for labeling and fragmenting a polynucleotide, and methods for labeling a polynucleotide

[0060] The invention provides novel methods and kits for labeling and fragmenting a polynucleotide, and novel methods and kits for labeling a polynucleotide. These methods are suitable for, for example, generation of labeled polynucleotides, or labeled polynucleotide fragments, for use as hybridization probes. Generally, the polynucleotide is labeled at an abasic site present in the polynucleotide, and fragmented at an abasic site present in the polynucleotide (in embodiments involving fragmentation). The abasic site present in the polynucleotide is generally prepared by cleavage of a base portion of a non-canonical nucleotide present in the polynucleotide. Thus, the spacing of the non-canonical nucleotide in the polynucleotide to be labeled and fragmented (in embodiments involving fragmentation), relates to and determines the size of fragments and intensity of labeling. This feature permits control of fragment size and/or site of labeling by use of conditions permitting controlled incorporation of non-canonical nucleotide, for example, during synthesis of the polynucleotide comprising the non-canonical nucleotide from a polynucleotide template.

[0061] Thus, in one aspect, the invention provides methods for labeling and fragmenting a polynucleotide. The methods generally comprise generation of a polynucleotide comprising a non-canonical nucleotide, cleavage of a base portion of the non-canonical nucleotide present in the polynucleotide with an agent (such as an enzyme) capable of cleaving a base portion of the non-canonical nucleotide (whereby an abasic site is generated); cleavage of the phosphodiester backbone at the abasic site, and labeling at the abasic site, whereby labeled polynucleotide fragments are generated. In another aspect, the invention provides methods for labeling a polynucleotide. The methods generally comprise generation of a polynucleotide comprising a non-canonical nucleotide, cleavage

IDT Exhibit 1020.7973

of a base portion of the non-canonical nucleotide present in the polynucleotide with an agent capable of cleaving a base portion of the non-canonical nucleotide (whereby an abasic site is generated); and labeling at the site of incorporation of the non-canonical nucleotide (i.e., at the abasic site), whereby a labeled polynucleotide(s) is generated.

[0062] The methods of labeling and fragmenting a polynucleotide and the methods of labeling a polynucleotide generally comprise synthesis of the polynucleotide comprising a non-canonical nucleotide from a polynucleotide template in the presence of a non-canonical nucleotide, whereby a polynucleotide comprising a non-canonical nucleotide(s) is generated.

[0063] Non-canonical nucleotides are known in the art and any suitable non-canonical polynucleotide can be used. In some embodiments, two or more different non-canonical nucleotides are used, such that a polynucleotide comprising two or more non-canonical nucleotides is generated. Method for synthesizing polynucleotides from a polynucleotide template are known in the art and described herein, and any suitable method can be used in the methods of the invention. In some embodiments, synthesis of the polynucleotide comprising the non-canonical nucleotides is using single primer isothermal amplification (*see* Kurn, U.S. Patent No. 6,251,639 B1), Ribo-SPIA™ (*see* Kurn, U.S. Patent Publication No. 2003/0087251 A1), PCR, primer extension, reverse transcription, strand displacement amplification (SDA), multiple displacement amplification (MDA), DNA replication, and the like. The polynucleotide that is synthesized can single stranded, double-stranded or partially double stranded, and either or both strands can comprise a non-canonical nucleotide. In some embodiments, the polynucleotide that is synthesized comprises a cDNA. The polynucleotide template (along which the polynucleotide comprising a non-canonical nucleotide is synthesized) is any template from which labeled polynucleotide or fragments thereof is desired to be produced. In some embodiments, the template comprises RNA, mRNA, genomic DNA, cDNA, or synthetic DNA. In other embodiments, the template comprises a cDNA library, a subtractive hybridization library, or a genomic library. Generally, the polynucleotide comprising the non-canonical nucleotide is synthesized using limited and/or controlled incorporation of the non-canonical nucleotide, which results in generation of a polynucleotide with a frequency or proportion of non-canonical nucleotides such that, in embodiments involving fragmentation, labeled fragments of a desired size (or size range) are generated (following production of an abasic site, labeling at an abasic site, and cleavage of the phosphodiester backbone at an abasic site (in embodiments involving fragmentation). Similarly, in embodiments involving labeling but not fragmentation,

IDT Exhibit 1020.7974

labeled polynucleotides are produced (following production of an abasic site, and labeling at an abasic site).

[0064] In some embodiment, a labeled primer is used during synthesis of the polynucleotide comprising a non-canonical nucleotide. In other embodiments, a primer comprising a non-canonical nucleotide (such as dUTP) is used during synthesis of the polynucleotide comprising a non-canonical nucleotide. In other embodiments, the primer is a composite primer, said composite primer comprising a RNA portion and a 3' DNA portion.

[0065] It is understood that a polynucleotide comprising a non-canonical nucleotide can be a multiplicity (from small to very large) of different polynucleotide molecules. Such populations can be related in sequence (e.g., members of a gene family or superfamily) or extremely diverse in sequence (e.g., generated from all mRNA, generated from all genomic DNA, etc.). Polynucleotides can also correspond to single sequences (which can be part or all of a known gene, for example a coding region, genomic portion, etc.).

[0066] A base portion of the non-canonical nucleotide is cleaved by an agent (such as an enzyme) capable of cleaving a base portion of a non-canonical nucleotide. Such agents are known in the art and described herein. In one embodiment, the agent capable of specifically cleaving a base portion of a non-canonical nucleotide is N-glycosylase. In another embodiment, the agent is Uracil N-Glycosylase (interchangeably termed "UNG" or "uracil DNA glyosylase").

[0067] The polynucleotide comprising an abasic site is labeled using an agent capable of labeling an abasic site, and, in embodiments involving fragmentation, the phosphodiester backbone of the polynucleotide comprising an abasic site is cleaved at the site of incorporation of the non-canonical nucleotide (i.e., the abasic site by an agent capable of cleaving the phosphodiester backbone at an abasic site, such that two or more fragments are produced. As used herein, "cleaving the backbone or phosphodiester backbone" is also termed "fragmentation" or fragmenting". In embodiments involving fragmentation, labeling can occur before fragmentation, fragmentation can occur before labeling, or fragmentation and labeling can occur simultaneously. For convenience, these steps are described separately below.

[0068] Agents capable of labeling (generally specifically labeling) an abasic site, whereby a polynucleotide (or polynucleotide fragment) comprising a labeled abasic site is generated, are known in the art. In some embodiments, the detectable moiety (label) is covalently or non-covalently associated with an abasic site. In some embodiments, the detectable moiety is directly or indirectly associated with an abasic site. In some

IDT Exhibit 1020.7975

embodiments, the detectable moiety (label) is directly or indirectly detectable. In some embodiments, the detectable signal is amplified. In some embodiments, the detectable moiety comprises an organic molecule. In other embodiments, the detectable moiety comprises an antibody. In other embodiments, the detectable signal is fluorescent. In other embodiments, the detectable signal is enzymatically generated. In some embodiments, the label is selected from 5-(((2-(carbohydrazino)-methyl)thio)acetyl)aminofluorescein, aminooxyacetyl hydrazide ("FARP"), N-(aminooxyacetyl)-N'-(D-biotinoyl) hydrazine, trifluoroacetic acid salt (ARP), Alexa Fluor 555, or an aminooxy-derivatized Alexa Fluor 555 (as described herein).

[0069] In embodiments involving fragmentation, the backbone of the polynucleotide comprising the abasic site is cleaved at the abasic site, whereby two or more fragments of the polynucleotide are generated. At least one of the fragments comprises an abasic site, which may be labeled and/or immobilized as described herein. Agents that cleave the phosphodiester backbone of a polynucleotide at an abasic site are known in the art. In some embodiments, the agent is E. coli AP endonuclease IV. In other embodiments, the agent is N, N'-dimethylethylenediamine (termed "DMED"). In other embodiments, the agent is heat, basic condition, acidic conditions, or an alkylating agent. Depending on the agent, the backbone can be cleaved 5' to the abasic site (e.g., cleavage between the 5'-phosphate group of the abasic residue and the deoxyribose ring of the adjacent nucleotide, generating a free 3' hydroxyl group), such that an abasic site is located at the 5' end of the resulting fragment. In other embodiments, cleavage can also be 3' to the abasic site (e.g., cleavage between the deoxyribose ring and 3'-phosphate group of the abasic residue and the deoxyribose ring of the adjacent nucleotide, generating a free 5' phosphate group on the deoxyribose ring of the adjacent nucleotide), such that an abasic site is located at the 3' end of the resulting fragment. In still other embodiments, more complex forms of cleavage are possible, for example, cleavage such that cleavage of the phosphodiester backbone and cleavage of a portion of the abasic nucleotide results. Selection of the fragmentation agent thus permits control of the orientation of the abasic site within the polynucleotide fragment, for example, at the 3' end of the resulting fragment or the 5' end of the resulting fragment. This feature has advantages, e.g., in embodiments involving immobilization as described below. Selection of reaction conditions also permits control of the degree, level or completeness of the fragmentation reactions. In some embodiments, reaction conditions can be selected such that the cleavage reaction is performed in the presence of a large excess of reagents and allowed to run to completion with minimal concern about excessive cleavage of the polynucleotide (i.e., while retaining a desired fragment size, which may be determined by spacing of the incorporated non-

IDT Exhibit 1020.7976

canonical nucleotide, during the synthesis step, above). By contrast, other methods known in the art, e.g., mechanical shearing, DNase cleavage, require careful titration of reaction conditions (including careful control of quantity of input DNA when DNase is used), to avoid excessive cleavage. In other embodiments, reaction conditions are selected such that fragmentation is not complete (in the sense that the backbone at some abasic sites remains uncleaved (unfragmented)), such that polynucleotide fragments comprising more than one abasic site are generated. Such fragments comprise internal (nonfragmented) abasic sites.

[0070] The methods of the invention include methods using the labeled polynucleotide fragments and labeled polynucleotides produced by the methods of the invention (so-called "applications"). The invention provides methods to characterize (for example, detect presence or absence of and/or quantify) a sequence of interest by analyzing the labeled and/or fragmented products by detection/quantification methods such as those based on array technologies or solution phase technologies. In some embodiments, the invention provides methods of detecting the presence or absence of mutations.

[0071] In other embodiments, the invention provides methods of producing a hybridization probe, hybridization using the hybridization probes; detection using the hybridization probes; characterizing and/or quantitating nucleic acid, preparing a subtractive hybridization probe, comparative genomic hybridization, and determining a gene expression profile, using the labeled and/or fragmented nucleic acids generated by the methods of the invention.

Methods for immobilizing a polynucleotide to a substrate at an abasic site

[0072] The invention also provides methods for the generation of polynucleotides, or fragments thereof, immobilized to a substrate (surface). In some embodiments, the immobilized polynucleotide, or immobilized polynucleotide fragment (in embodiments involving fragmentation) is labeled according to the labeling methods described herein. These methods are suitable for, for example, the production of microarrays or tagged analytes.

[0073] As described herein, the abasic site is generally prepared by cleavage of a base portion of a non-canonical nucleotide present in the polynucleotide, and, as such, the spacing of the non-canonical nucleotide in the polynucleotide to be immobilized, optionally fragmented and/or optionally labeled, relates to and determines the site of immobilization, size of fragments (in embodiments involving fragmentation) and intensity of labeling (in embodiments involving labeling). This feature permits control of fragment size and/or intensity and location of labeling (in embodiments involving labeling) by use of

IDT Exhibit 1020.7977

conditions permitting controlled incorporation of non-canonical nucleotide, for example, during synthesis of the polynucleotide comprising the non-canonical nucleotide from a polynucleotide template.

[0074] Thus, in one aspect, the invention provides methods for immobilizing a polynucleotide to a substrate comprising cleavage of a base portion of a non-canonical nucleotide present in a polynucleotide comprising a non-canonical nucleotide with an agent capable of cleaving a base portion of the non-canonical nucleotide (whereby an abasic site is created); optionally, cleaving the phosphodiester backbone of the polynucleotide at the abasic site, whereby fragments are generated; and immobilizing the polynucleotide, or fragments thereof (in embodiments involving fragmentation) on a substrate at the abasic site. Generally, the polynucleotide comprising a non-canonical nucleotide is prepared using any method known in the art and as described herein. Agents capable of cleaving a base portion of a non-canonical nucleotide and, in embodiments involving fragmentation, agents capable of cleaving a phosphodiester backbone at an abasic site, are as described herein.

[0075] Optionally, the polynucleotides, or fragments thereof, are labeled according to any of the labeling methods described herein. Thus, in some embodiments, the invention provides methods for generating labeled polynucleotides, or labeled polynucleotide fragments, that are immobilized to a substrate. In some embodiments, the polynucleotide, or polynucleotide fragments are labeled according to any of the labeling methods disclosed herein.

[0076] The polynucleotide (or fragment thereof) comprising an abasic site is immobilized to a substrate at the abasic site. The substrate can be a solid or semi-solid surface, e.g., a microarray. In other embodiments, the microarray comprises at least one polynucleotide (or fragment thereof) 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. In other embodiments, the polynucleotide (or fragment thereof) 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. In other embodiments, polynucleotide (or fragment thereof in embodiments involving fragmentation) comprising an abasic site is immobilized to a substrate selected from the group consisting of one or more of: protein, polypeptide, peptide, nucleic acid, carbohydrates, cells, microorganisms and fragments and products thereof, an organic molecule, and an inorganic molecule. In still other embodiment, the substrate is selected from a polypeptide, an antibody, an organic molecule and an inorganic molecule.

IDT Exhibit 1020.7978

[0077] Single stranded polynucleotides (including polynucleotide fragments) are particularly suitable for preparing microarrays comprising the single stranded polynucleotides. Single stranded polynucleotide fragments (in embodiments involving cleavage of the phosphodiester backbone at an abasic site) are advantageous, because the orientation of the fragment with respect to the surface (upon which the fragment is immobilized) can be controlled by selection of the method used to cleave the phosphodiester backbone, such that an abasic site is positioned at the 3' end of a fragment or at the 5' end of a fragment. Immobilizing polynucleotides in a defined orientation (e.g., at the 3' end, at the 5' end) enhances hybridization of complementary oligonucleotides, and permits a higher density of immobilization.

[0078] The methods of the invention include methods using the immobilized polynucleotides, or immobilized polynucleotide fragments produced by the methods of the invention (so-called "applications"). In some embodiments, the invention provides methods of detecting nucleic acid sequence mutations. .

[0079] The invention also provides methods to characterize (for example, detect presence or absence of and/or quantify) a sequence of interest using the immobilized polynucleotides, or fragments thereof.

[0080] In another embodiment, the invention provides methods of determining a gene expression profile, using the immobilized polynucleotides, or fragments thereof, generated by the methods of the invention.

General Techniques

[0081] 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).

[0082] Primers, oligonucleotides and polynucleotides employed in the invention can be generated using standard techniques known in the art.

Definitions

IDT Exhibit 1020.7979

[0083] A "template sequence," or "template nucleic acid" or "template" as used herein, is a polynucleotide comprising a sequence of interest, for which synthesis of a complement comprising a non-canonical nucleotide is desired. The template 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.

[0084] "Polynucleotide," or "nucleic acid," as used interchangeably herein, refer to polymers of nucleotides of any length, and include DNA. The nucleotides can be deoxyribonucleotides, modified nucleotides or bases, and/or their analogs, or any substrate that can be incorporated into a polymer by DNA polymerase. Nucleotides include canonical and non-canonical nucleotides and a polynucleotide can comprise canonical and non-canonical nucleotides. A polynucleotide may comprise modified (altered) nucleotides, such as, for example, modification to the nucleotide structure and or modification to the phosphodiester backbone. As discussed herein modified nucleotide can be canonical nucleotide or non-canonical (cleavable) nucleotides. It is understood, however, that modified nucleotides that are not non-canonical (cleavable) nucleotide under the reaction conditions used in the methods of the invention, if present, generally should not affect the ability of the polynucleotide to undergo cleavage of a base portion of non-canonical nucleotide, such that an abasic site is generated, and/or cleavage of a phosphodiester backbone at an abasic site, such that fragments are generated, and/or immobilization of a polynucleotide (or fragment thereof) to a substrate, as described herein. If present, modification to the nucleotide structure, such as methylated nucleotides may be imparted before or after assembly of the polymer. The sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component. Other types of modifications include, for example, "caps", substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as, for example, those with uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoamidates, carbamates, etc.) and with charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), those containing pendant moieties, such as, for example, proteins (e.g., nucleases, toxins, antibodies, signal peptides, ply-L-lysine, etc.), those with intercalators (e.g., acridine, psoralen, etc.), those containing chelators (e.g., metals, radioactive metals, boron, oxidative metals, etc.), those containing alkylators, those with modified linkages (e.g., alpha anomeric nucleic acids, etc.), as well as unmodified forms of the polynucleotide(s). It is understood that internucleotide modifications may, e.g., alter the efficiency and/or kinetics of cleavage of the phosphodiester backbone (as when, for example a phosphodiester backbone is cleaved at an abasic site, as described herein).

IDT Exhibit 1020.7980

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. The 5' and 3' terminal OH can be phosphorylated or substituted with amines or organic capping groups moieties of from 1 to 20 carbon atoms. Other hydroxyls may also be derivatized to standard protecting groups. Polynucleotides can also contain analogous forms of ribose or deoxyribose sugars that are generally known in the art, including, for example, 2'--O-methyl-, 2'-O-allyl, 2'-fluoro- or 2'-azido-ribose, carbocyclic sugar analogs, α-anomeric sugars, epimeric sugars such as arabinose, xyloses or lyxoses, pyranose sugars, furanose sugars, sedoheptuloses, acyclic analogs and abasic nucleoside analogs. 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 DNA. It is understood, however, that modified nucleotides and/or internucleotide linkages and/or, if present, generally should not affect the ability of the polynucleotide to undergo cleavage of a base portion of a non-canonical nucleotide, such that an abasic site is generated, and/or the ability of a polynucleotide to undergo cleavage of a phosphodiester backbone at an abasic site, such that fragments are generated, and/or the ability of a polynucleotide to be immobilized at an abasic site (such as an abasic site at an end of a polynucleotide and/or an abasic site that is not at an end of a polynucleotide) to a surface, as described herein.

[0085] "Oligonucleotide," as used herein, generally refers to short, generally single stranded, generally synthetic polynucleotides that are generally, but not necessarily, less than about 200 nucleotides in length. The terms "oligonucleotide" and "polynucleotide" are not mutually exclusive. The description above for polynucleotides is equally and fully applicable to oligonucleotides.

[0086] A "primer," as used herein, refers to a nucleotide sequence (a polynucleotide), generally with a free 3'-OH group, that hybridizes with a template sequence (such as a template 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 an RNA template created following RNase cleavage of a template RNA-DNA complex) that is hybridized to a

IDT Exhibit 1020.7981

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.

[0087] 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 polynucleotide fragments, labeled polynucleotide, labeled polynucleotide fragments, and/or immobilized polynucleotide or fragment thereof.

[0088] A "fragment" of a polynucleotide or oligonucleotide is a contiguous sequence of 2 or more bases. In other embodiments, a fragment (also termed "region" or "portion") is any of about 3, about 5, about 10, about 15, about 20, about 25, about 30 about 35 about 40, about 50, about 65, about 75, about 85, about 100, about 125, about 150, about 175, about 200, about 225, about 250, about 300, about 350, about 400, about 450, about 500, about 550, about 600, about 650 or more nucleotides in length. In some embodiments, the fragments can be at least about 3, about 5, about 10, about 15, about 20, about 25, about 30 about 35 about 40, about 50, about 65, about 75, about 85, about 100, about 125, about 150, about 175, about 200, about 225, about 250, about 300, about 350, about 400, about 450, about 500, about 550, about 600, about 650 or more nucleotides in length. In other embodiments, the fragments can be less than about 3, about 5, about 10, about 15, about 20, about 25, about 30 about 35 about 40, about 50, about 65, about 75, about 85, about 100, about 125, about 150, about 175, about 200, about 225, about 250, about 300, about 350, about 400, about 450, about 500, about 550, about 600, about 650 or more nucleotides in length. In some embodiment, these fragment lengths represent an average size in the population of fragments generated using the methods of the invention.

[0089] 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).

[0090] "A", "an" and "the", and the like, unless otherwise indicated include plural forms. "A" fragment means one or more fragments. "A" non-canonical nucleotide means one or more non-canonical nucleotides.

[0091] "Comprising" means including in accordance with well-established principles of patent law.

[0092] Conditions that "allow" an event to occur or conditions that are "suitable" for an event to occur, such as polynucleotide synthesis, cleavage of a base portion of a

IDT Exhibit 1020.7982

non-canonical nucleotide, cleavage of a phosphodiester backbone at an abasic site, and the like, or "suitable" conditions are conditions that do not prevent such events from occurring. Thus, these conditions permit, enhance, facilitate, and/or are conducive to the event. Such conditions, known in the art and described herein, depend upon, for example, the nature of the polynucleotide sequence, temperature, and buffer conditions. These conditions also depend on what event is desired, such as polynucleotide synthesis, cleavage of a base portion of a non-canonical nucleotide, cleavage of a phosphodiester backbone at an abasic site, labeling an abasic site, immobilizing a polynucleotide fragment or a polynucleotide, etc.

[0093] "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 and other metals, optical fiber or any other suitable solid or semi-solid support, and configured in a planar (e.g., glass plates, silicon chips) or three-dimensional (e.g., pins, fibers, beads, particles, microtiter wells, capillaries) configuration. Probes forming the arrays may be attached to the substrate by any number of ways including (i) *in situ* synthesis (e.g., high-density oligonucleotide arrays) using photolithographic techniques (see, Fodor et al., *Science* (1991), 251:767-773; Pease et al., *Proc. Natl. Acad. Sci. U.S.A.* (1994), 91:5022-5026; Lockhart et al., *Nature Biotechnology* (1996), 14:1675; U.S. Pat. Nos. 5,578,832; 5,556,752; and 5,510,270); (ii) spotting/printing at medium to low-density (e.g., cDNA probes) on glass, nylon or nitrocellulose (Schena et al, *Science* (1995), 270:467-470, DeRisi et al, *Nature Genetics* (1996), 14:457-460; Shalon et al., *Genome Res.* (1996), 6:639-645; and Schena et al., *Proc. Natl. Acad. Sci. U.S.A.* (1995), 93:10539-11286); (iii) by masking (Maskos and Southern, *Nuc. Acids. Res.* (1992), 20:1679-1684) and (iv) by dot-blotting on a nylon or nitrocellulose hybridization membrane (see, e.g., Sambrook et al., Eds., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., Vol. 1-3, Cold Spring Harbor Laboratory (Cold Spring Harbor, N.Y.)). Probes may also be noncovalently immobilized on the substrate by hybridization to 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.

IDT Exhibit 1020.7983

[0094] 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.

[0095] 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.

[0096] 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.

[0097] As used herein, "canonical" nucleotide means a nucleotide comprising one the four common nucleic acid bases adenine, cytosine, guanine and thymine that are commonly found in DNA. The term also encompasses the respective deoxyribonucleosides, deoxyribonucleotides or 2'-deoxyribonucleoside-5'-triphosphates that contain one of the four common nucleic acid bases adenine, cytosine, guanine and thymine (though as explained herein, the base can be a modified and/or altered base as discussed, for example, in the definition of polynucleotide). As used herein, the base portions of canonical nucleotides are generally not cleavable under the conditions used in the methods of the invention.

[0098] As used herein, "non-canonical nucleotide" (interchangeably called "non-canonical deoxyribonucleoside triphosphate") refers to a nucleotide comprising a base other than the four canonical bases. The term also encompasses the respective deoxyribonucleosides, deoxyribonucleotides or 2'-deoxyribonucleoside-5'-triphosphates that contain a base other than the four canonical bases. In the context of this invention, nucleotides containing uracil (such as dUTP), or the respective deoxyribonucleosides, deoxyribonucleotides or 2'-deoxyribonucleoside-5'-triphosphates, are a non-canonical nucleotides. As used herein, the base portions of non-canonical nucleotides are capable of being, generally, specifically or selectively cleaved (such that a nucleotide comprising an abasic site is created) under the reaction conditions used in the methods of the invention. As described herein, non-canonical nucleotides are generally also capable of being incorporated into a polynucleotide during synthesis of a polynucleotide (during e.g., primer extension and/or replication); capable of being generally, specifically or selectively cleaved by an agent that cleaves a base portion of a nucleotide, such that a polynucleotide

IDT Exhibit 1020.7984

comprising an abasic site is generated; comprise a suitable internucleotide connection (when incorporated into a polynucleotide) such that a phosphodiester backbone at an abasic site (i.e., the non-canonical nucleotide following cleavage of a base portion) is capable of being cleaved by an agent capable of such cleavage; capable of being labeled (following generation of an abasic site); and/or capable of immobilization to a surface (following generation of an abasic site), according to the methods described herein. It is understood that the non-canonical nucleotide may, but does not necessarily, require all of the features described above, depending on the particular method of the invention in which the non-canonical nucleotide is to be used. In some embodiments, non-canonical nucleotides are altered and/or modified nucleotides as described herein. Non-canonical nucleotide refers to a nucleotide that is incorporated into a polynucleotide as well as to a single nucleotide.

[0099] The term "analyte" as used herein refers to a substance to be detected or assayed by the method of the present invention, for example, a compound whose properties, location, quantity and/or identity is desired to be characterized. Typical analytes may include, but are not limited to proteins, peptides, nucleic acid segments, cells, microorganisms and fragments and products thereof, organic molecules, inorganic molecules, or any substance for which immobilization sites for binding partner(s) can be developed. As this disclosure clearly conveys, an analyte is a substrate.

[00100] As used herein, an "abasic site" refers to the site of incorporation of the non-canonical nucleotide following treatment with an agent capable of effecting cleavage of a base portion of the non-canonical nucleotide. An abasic site (interchangeably termed "AP site") can comprise a hemiacetal ring, and lacks a base portion of the non-canonical nucleotide. As used herein, "abasic site" encompasses any chemical structure remaining following treatment of a non-canonical nucleotide (present in a polynucleotide chain) with an agent (e.g., an enzyme, or heat or basic conditions) capable of effecting cleavage of a base portion of a non-canonical nucleotide. Thus, an abasic site as used herein includes a modified sugar moiety attached to the 3' terminus of nicked polynucleotide, as when, for example, endonuclease III or OGG1 protein are used to cleave the base portion of the non-canonical nucleotide. *See, e.g.,* Kow, (2000) *Methods* 22, 164-169 (e.g., Figure 4).

[00101] As used herein, "labeling at an abasic site" means association of a label with any chemical structure remaining following removal of a base portion (including the entire base) of a non-canonical nucleotide (present in a polynucleotide chain) by treatment with an agent (e.g., an enzyme, or heat)) capable of effecting cleavage of a base portion of a non-canonical nucleotide. In one embodiment, a reactive aldehyde form of a hemiacetal ring in an abasic site is labeled. In other embodiments, the label associate with a chemical

IDT Exhibit 1020.7985

structure remaining following treatment of a non-canonical nucleotide (present in a polynucleotide chain) with an agent (e.g., an enzyme, or heat or basic conditions) capable of effecting cleavage of a base portion of a non-canonical nucleotide and treatment of polynucleotide comprising an abasic site with an agent capable of effecting cleavage of the backbone at the abasic site (as described herein).

[00102] As used herein, cleavage of a backbone (e.g. phosphodiester backbone) "at" an abasic site means cleavage of the phosphodiester linkage 3' to the abasic site or 5' to the abasic site, or both. As the disclosure herein conveys, "at" an abasic site refers to proximate or near location (such as immediately 3' or immediately 5'). In still other embodiments, more complex forms of cleavage are possible, for example, cleavage such that cleavage of the phosphodiester backbone and cleavage of (a portion of) the abasic nucleotide results.

[00103] As used herein, a "label" (interchangeably called a "detectable moiety") refers to a moiety that is associated or linked with a polynucleotide (interchangeably called "labeling"). The labeled polynucleotide may be directly or indirectly detected, generally through a detectable signal. The detectable moiety (label) can be attached (or associated) either directly or through a non-interfering linkage group with other moieties capable of specifically associating with one or more sites to be labeled. The detectable moiety (label) may be covalently or non-covalently associated as well as directly or indirectly associated.

[0100] The following are examples of the methods of the invention. It is understood that various other embodiments may be practiced, given the general description provided herein. For example, reference to using an agent capable of cleaving a base portion of the non-canonical nucleotide means that any of the agents capable of cleaving a base portion of the non-canonical nucleotide described herein may be used.

Methods for labeling and fragmenting nucleic acids

[0101] The invention provides methods for generating labeled fragments of nucleic acid. The methods generally comprise generation of a polynucleotide comprising at least one non-canonical nucleotide, cleavage of a base portion of the non-canonical nucleotide present in the polynucleotide with an agent capable of cleaving a base portion of the non-canonical nucleotide; and cleavage of the phosphodiester backbone of the polynucleotide comprising the abasic site at the abasic site; and labeling at the abasic site, whereby labeled nucleic acid fragments are generated. Generally, the polynucleotide comprising a non-canonical nucleotide is fragmented and labeled at the site of incorporation of the non-canonical nucleotide(s) present in the synthesized polynucleotide. Thus, the frequency of non-canonical nucleotides in the synthesized polynucleotide

IDT Exhibit 1020.7986

generally relates to and determines the size range of the labeled fragments produced from the polynucleotide. The methods of the invention generate labeled nucleic acid fragments, which are useful for, for example, hybridization to a microarray and other uses described herein.

[0102] For convenience, the synthesis of a polynucleotide comprising a non-canonical nucleotide, and the treatment of that polynucleotide with an agent, such as an enzyme, capable of cleaving a base portion of the non-canonical nucleotide are described as separate steps. It is understood that these steps (e.g., one or more of these steps) may be performed simultaneously, except (generally) in the case when a polynucleotide comprising a non-canonical nucleotide must be capable of serving as a template for further amplification (as in exponential methods of amplification, e.g. PCR), in which case it is preferable to synthesize the polynucleotide comprising an abasic site prior to cleaving the base portion of the non-canonical nucleotide.

[0103] The methods involve the following steps: (a) synthesizing a polynucleotide from a template in the presence of a non-canonical nucleotide (interchangeably termed "non-canonical deoxyribonucleoside triphosphate" or "non-canonical deoxyribonucleotide"), whereby a polynucleotide comprising a non-canonical nucleotide is generated; (b) contacting the polynucleotide comprising a non-canonical nucleotide with an agent capable of cleaving a base portion of the non-canonical nucleotide (i.e., cleaving a base portion of the non-canonical nucleotide), whereby an abasic site is created; (c) cleaving the backbone of the polynucleotide comprising the abasic site at the abasic site; and (d) contacting the polynucleotide comprising the abasic site with an agent capable of labeling the abasic site (i.e., labeling the abasic site), whereby labeled polynucleotide fragments are generated.

[0104] For simplicity, individual steps of the labeling and fragmentation method are discussed below. It is understood, however, that the steps may be performed simultaneously and/or in varied order, as discussed herein.

Synthesis of a polynucleotide comprising a non-canonical nucleotide

[0105] The methods involve synthesizing a polynucleotide from a template in the presence of at least one non-canonical nucleotide (interchangeably termed "non-canonical deoxyribonucleoside triphosphate"), whereby a polynucleotide comprising a non-canonical nucleotide is generated. The frequency of incorporation of non-canonical nucleotides into the polynucleotide relates to the size of fragment produced using the methods of the invention because the spacing between non-canonical nucleotides in the polynucleotide comprising a non-canonical nucleotide, along with the reaction conditions used,

IDT Exhibit 1020.7987

determines the approximate size of the fragments resulting from generation of an abasic site from the non-canonical nucleotide and cleavage of the backbone at the abasic site, as described herein.

[0106] Generally, the polynucleotide is DNA, though, as noted herein, the polynucleotide can comprise altered and/or modified nucleotides, internucleotide linkages, ribonucleotides, etc. As generally used herein, it is understood that "DNA" applies to polynucleotide embodiments.

[0107] Methods for synthesizing polynucleotides, e.g., single and double stranded DNA, from a template are well known in the art, and include, for example, single primer isothermal amplification, Ribo-SPIA™, PCR, reverse transcription, primer extension, limited primer extension, replication (including rolling circle replication), strand displacement amplification (SDA), nick translation, multiple displacement amplification (MDA), and, e.g., any method that results in synthesis of the complement of a template sequence such that at least one non-canonical nucleotide can be incorporated into a polynucleotide. *See, e.g.,* Kurn, U.S. Patent No. 6,251,639 B1; Kurn, WO 02/00938; Kurn, U.S. Patent Publication No. 2003/0087251 A1; Mullis, U.S. Patent No. 4,582,877; Wallace, U.S. patent No. 6,027,923; U.S. Patent No. 5,508,178; 5,888,819; 6,004,744; 5,882,867; 5,710,028; 6,027,889; 6,004,745; 5,763,178; 5,011,769; *see also* Sambrook (1989) "Molecular Cloning: A Laboratory Manual", second edition; Ausebel (1987, and updates) "Current Protocols in Molecular Biology"; Mullis, (1994) "PCR: The Polymerase Chain Reaction". One or more methods known in the art can be used to generate a polynucleotide comprising a non-canonical nucleotide. It is understood that the polynucleotide comprising a non-canonical nucleotide can be single stranded or double stranded or partially double stranded, and that one or both strands of a double stranded polynucleotide can comprise a non-canonical nucleotide. For convenience, "DNA" is used herein to describe (and exemplify) a polynucleotide. Suitable methods include methods that result in one single- or double-stranded polynucleotide comprising a non-canonical nucleotide (for example, reverse transcription, production of double stranded cDNA, a single round of DNA replication), as well as methods that result in multiple single stranded or double stranded copies or copies of the complement of a template (for example, single primer isothermal amplification or Ribo-SPIA™ or PCR). In one embodiment, illustrated in Figure 1, a single-stranded polynucleotide comprising a non-canonical nucleotide is synthesized using single primer isothermal amplification. *See* Kurn, U.S. Patent No. 6,251,639 B1.

[0108] Generally, the polynucleotide comprising a non-canonical nucleotide is generated from a template in the presence of all four canonical nucleotides and at least one

IDT Exhibit 1020.7988

non-canonical nucleotide under reaction conditions suitable for synthesis of polynucleotides, including suitable enzymes and primers, if necessary. Reaction conditions and reagents, including primers, for synthesizing the polynucleotide comprising a non-canonical nucleotide are known in the art, and further discussed herein. As described herein, non-canonical nucleotides are generally capable of polymerization (i.e., are substrates for DNA polymerase), and capable of being rendered abasic following treatment with a suitable agent capable of generally, specifically or selectively cleaving a base portion of a non-canonical nucleotide. Suitable non-canonical nucleotides are well-known in the art, and include: deoxyuridine triphosphate (dUTP), deoxyinosine triphosphate (dITP), 5-hydroxymethyl deoxycytidine triphosphate (5-OH-Me-dCTP). *See, e.g.,* Jendrisak, U.S. Patent No. 6,190,865 B1; *Mol. Cell Probes* (1992) 251-6. Generally, in embodiments in which a polynucleotide comprising an non-canonical nucleotide serves as a template for further amplification (e.g., as when multiple copies of a double stranded polynucleotides comprising a non-canonical nucleotide are synthesized, e.g., by PCR amplification), a polynucleotide comprising a non-canonical nucleotide must be capable of serving as a template for further amplification.

[0109] It is understood that two or more different non-canonical nucleotides can be incorporated into the polynucleotide synthesized from the template by DNA polymerase, whereby a polynucleotide comprising at least two different non-canonical nucleotides is generated.

[0110] Conditions for limited and/or controlled incorporation of a non-canonical nucleotide are known in the art. *See, e.g.,* Jendrisak, U.S. Patent No. 6,190,865 B1; *Mol. Cell Probes* (1992) 251-6; *Anal. Biochem.* (1993) 211:164-9; *see also* Sambrook (1989) "Molecular Cloning: A Laboratory Manual", second edition; Ausebel (1987, and updates) "Current Protocols in Molecular Biology". The frequency (or spacing) of non-canonical nucleotides in the resulting polynucleotide comprising a non-canonical nucleotide, and thus the average size of fragments generated using the methods of the invention (i.e., following cleavage of a base portion of a non-canonical nucleotide, and cleavage of a phosphodiester backbone at a non-canonical nucleotide), is controlled by variables known in the art, including: frequency of nucleotide(s) corresponding to the non-canonical nucleotide(s) in the template (or other measures of nucleotide content of a sequence, such as average G-C content), ratio of canonical to non-canonical nucleotide present in the reaction mixture; ability of the polymerase to incorporate the non-canonical nucleotide, relative efficiency of incorporation of non-canonical nucleotide verses canonical nucleotide, and the like. It is understood that the average fragmentation size also relates to the reaction conditions used during fragmentation, as is further discussed herein. The

IDT Exhibit 1020.7989

reaction conditions can be empirically determined, for example, by assessing average fragment size generated using the methods of the invention taught herein. The level of labeling at an abasic site also relates to the frequency of incorporation of non-canonical nucleotides, as is further discussed herein.

[0111] Generally, a non-canonical base can be incorporated at about every 5, 10, 15, 20, 25, 30, 40, 50, 65, 75, 85, 100, 123, 150, 175, 200, 225, 250, 300, 350, 400, 450, 500, 550, 600, 650 or more nucleotides apart in the resulting polynucleotide comprising a non-canonical nucleotide. In one embodiment, the non-canonical nucleotide is incorporated about every 200 nucleotides, about every 100 nucleotide, or about every 50 nucleotide. In another embodiment, the non-canonical nucleotide is incorporated about every 50 to about 200 nucleotides. In some embodiments, a 1:5 ratio of dUTP and dTTP is used in the reaction mixture.

[0112] The polynucleotide template (along which the polynucleotide comprising a non-canonical nucleotide is synthesized) may be any template from which labeled polynucleotide fragments are desired to be produced. As is evident from the description herein, the labeled polynucleotide fragments are the complement of the sequence of the polynucleotide template. The template includes double-stranded, partially double-stranded, and single-stranded nucleic acids from any source in purified or unpurified form, which can be DNA (dsDNA and ssDNA) or RNA, including tRNA, mRNA, rRNA, mitochondrial DNA and RNA, chloroplast DNA and RNA, DNA-RNA hybrids, or mixtures thereof, genes, chromosomes, plasmids, the genomes of biological material such as microorganisms, e.g., bacteria, yeasts, viruses, viroids, molds, fungi, plants, animals, humans, and fragments thereof. Obtaining and purifying nucleic acids use standard techniques in the art. RNAs can be obtained and purified using standard techniques in the art. A DNA template (including genomic DNA template) can be transcribed into RNA form, which can be achieved using methods disclosed in Kurn, U.S. Patent No. 6,251,639 B1, and by other techniques (such as expression systems) known in the art. RNA copies of genomic DNA would generally include untranscribed sequences generally not found in mRNA, such as introns, regulatory and control elements, etc. DNA copies of an RNA template can be synthesized using methods described in Kurn, U.S. Patent Publication No. 2003/0087251 A1 or other techniques known in the art). Synthesis of polynucleotide comprising a non-canonical nucleotide from a DNA-RNA hybrid can be accomplished by denaturation of the hybrid to obtain a ssDNA and/or RNA, cleavage with an agent capable of cleaving RNA from an RNA/DNA hybrid, and other methods known in the art. The template can be only a minor fraction of a complex mixture such as a biological sample and can be obtained from various biological material by procedures well known in the art.

IDT Exhibit 1020.7990

The template can be known or unknown and may contain more than one desired specific nucleic acid sequence of interest, each of which may be the same or different from each other. Therefore, the methods of the invention are useful not only for producing one specific polynucleotide comprising a non-canonical nucleotide, but also for producing simultaneously more than one different specific polynucleotides comprising a non-canonical nucleotide. The template DNA can be a sub-population of nucleic acids, for example, a subtractive hybridization probe, total genomic DNA, restriction fragments, a cDNA library, cDNA prepared from total mRNA, a cloned library, or amplification products of any of the templates described herein. In some cases, the initial step of the synthesis of the complement of a portion of a template nucleic acid sequence is template denaturation. The denaturation step may be thermal denaturation or any other method known in the art, such as alkali treatment.

[0113] For simplicity, the polynucleotide comprising a non-canonical nucleotide is described as a single nucleic acid. It is understood that the polynucleotide can be a single polynucleotide, or a population of polynucleotides (from a few to a multiplicity to a very large multiplicity of polynucleotides). It is further understood that a polynucleotide comprising a non-canonical nucleotide can be a multiplicity (from small to very large) of different polynucleotide molecules. Such populations can be related in sequence (e.g., member of a gene family or superfamily) or extremely diverse in sequence (e.g., generated from all mRNA, generated from all genomic DNA, etc.). Polynucleotides can also correspond to single sequence (which can be part or all of a known gene, for example a coding region, genomic portion, etc.). Methods, reagents, and reaction conditions for generating specific polynucleotide sequences and multiplicities of polynucleotide sequences are known in the art.

[0114] Suitable methods of synthesis of a polynucleotide comprising a non-canonical nucleotide are generally template-dependent (in the sense that polynucleotide comprising a non-canonical nucleotide is synthesized along a polynucleotide template, as generally described herein). It is understood that non-canonical nucleotides can be incorporated into a polynucleotide as a result of template-independent methods. For example, one or more primer(s) can be designed to comprise one or more non-canonical nucleotides. *See, e.g.,* Richards, U.S. Patent Nos. 6,037,152, 5,427,929, and 5,876,976. As discussed herein, inclusion of at least one non-canonical nucleotide in a primer results in cleavage of a base-portion of a non-canonical nucleotide and labeling at the abasic site (i.e., following generation of an abasic site, as described herein), thus generating a polynucleotide fragment or a labeled polynucleotide fragment comprising a portion of the primer. Inclusion of a non-canonical nucleotide in a primer may be particularly suitable

IDT Exhibit 1020.7991

for methods such as single primer isothermal amplification. *See* Kurn, U.S. Patent No. 6,251,639 B1; Kurn, WO 02/00938; Kurn, U.S. Patent Publication No. 2003/0087251 A1. Non-canonical nucleotide(s) can also be added to a polynucleotide by template-independent methods such as tailing and ligation of a second polynucleotide comprising a non-canonical nucleotide. Methods for tailing and ligation are well-known in the art.

Cleaving a base portion of the non-canonical nucleotide to create an abasic site

[0115] The polynucleotide comprising a non-canonical nucleotide is treated with an agent, such as an enzyme, capable of generally, specifically, or selectively cleaving a base portion of the non-canonical deoxyribonucleoside to create an abasic site. The exemplary embodiment shown in Figure 1 illustrates cleavage of a base portion of the non-canonical nucleotides, by an enzyme, whereby an abasic site is created. As used herein, "abasic site" encompasses any chemical structure remaining following removal of a base portion (including the entire base) with an agent capable of cleaving a base portion of a nucleotide, e.g., by treatment of a non-canonical nucleotide (present in a polynucleotide chain) with an agent (e.g., an enzyme, acidic conditions, or a chemical reagent) capable of effecting cleavage of a base portion of a non-canonical nucleotide. In some embodiments, the agent (such as an enzyme) catalyzes hydrolysis of the bond between the base portion of the non-canonical nucleotide and a sugar in the non-canonical nucleotide to generate an abasic site comprising a hemiacetal ring and lacking the base (interchangeably called "AP" site), though other cleavage products are contemplated for use in the methods of the invention. Suitable agents and reaction conditions for cleavage of base portions of non-canonical nucleotides are known in the art, and include: N-glycosylases (also called "DNA glycosylases" or "glycosidases") including Uracil N-Glycosylase ("UNG"; specifically cleaves dUTP) (interchangeably termed "uracil DNA glyosylase"), hypoxanthine-N-Glycosylase, and hydroxy-methyl cytosine-N-glycosylase; 3-methyladenine DNA glycosylase, 3- or 7- methylguanine DNA glycosylase, hydroxymethyluracile DNA glycosylase; T4 endonuclease V. *See, e.g.*, Lindahl, *PNAS* (1974) 71(9):3649-3653; Jendrisak, U.S. Patent No. 6,190,865 B1. In one embodiment, uracil-N-glycosylase is used to cleave a base portion of the non-canonical nucleotide. In other embodiments, the agent that cleaves the base portion of the non-canonical nucleotide is the same agent that cleaves a phosphodiester backbone at the abasic site.

[0116] Generally, cleavage of base portions of non-canonical nucleotides is general, specific or selective cleavage (in the sense that the agent (such as an enzyme) capable of cleaving a base portion of a non-canonical nucleotide generally, specifically or selectively cleaves the base portion of a particular non-canonical nucleotide), whereby

IDT Exhibit 1020.7992

greater than about 98%, about 95%, about 90%, about 85%, or about 80% of the base portions cleaved are base portions of non-canonical nucleotides. However, extent of cleavage can be less. Thus, reference to specific cleavage is exemplary. General, specific or selective cleavage is desirable for control of the fragment size in the methods of generating labeled polynucleotide fragments of the invention (i.e., the fragments generated by cleavage of the backbone at an abasic site). Generally, reaction conditions are selected such that the reaction in which the abasic site(s) are created can run to completion.

[0117] In some embodiments, the polynucleotide comprising a non-canonical nucleotide is purified following synthesis of the non-canonical polynucleotide (to eliminate, for example, residual free non-canonical nucleotides that are present in the reaction mixture). In other embodiments (such as the embodiment described in Example 4), there is no intermediate purification between the synthesis of the polynucleotide comprising the non-canonical nucleotide and subsequent steps (such as cleavage of a base portion of the non-canonical nucleotide and cleavage of a phosphodiester backbone at the abasic site).

[0118] As noted herein, for convenience, cleavage of a base portion of a non-canonical nucleotide (whereby an abasic site is generated) has been described as a separate step. It is understood that this step may be performed simultaneously with synthesis of the polynucleotide comprising a non-canonical nucleotide (as described above), cleavage of the backbone at an abasic site (fragmentation) and/or labeling at an abasic site.

[0119] It is understood that the choice of non-canonical nucleotide can dictate the choice of enzyme to be used to cleave the base portion of that non-canonical enzyme, to the extent that particular non-canonical nucleotides are recognized by particular enzymes that are capable of cleaving a base portion of the non-canonical nucleotide..

Cleaving the backbone at the abasic site of the polynucleotide comprising an abasic site and labeling at the abasic site

[0120] The backbone of the polynucleotide is cleaved at the abasic site, and the abasic site is labeled, whereby labeled fragments of nucleotide are generated. It is understood that cleavage of the backbone and labeling can be performed in any order, or simultaneously. For convenience, however, these reactions are described as separate steps.

Cleaving the backbone at the abasic site of the polynucleotide comprising an abasic site

[0121] Following generation of an abasic site by cleavage of the base portion of the non-canonical nucleotide present in the polynucleotide, the backbone of the polynucleotide is cleaved at the site of incorporation of the non-canonical nucleotide (also

IDT Exhibit 1020.7993

termed the abasic site, following cleavage of the base portion of the non-canonical nucleotide) with an agent capable of effecting cleavage of the backbone at the abasic site. Cleavage at the backbone (also termed "fragmentation") results in at least two fragments (depending on the number of abasic sites present in the polynucleotide comprising an abasic site, and the extent of cleavage).

[0122] Suitable agents (for example, an enzyme, a chemical and/or reaction conditions such as heat) capable of cleavage of the backbone at an abasic site are well known in the art, and include: heat treatment and/or chemical treatment (including basic conditions, acidic conditions, alkylating conditions, or amine mediated cleavage of abasic sites, (*see e.g.,* McHugh and Knowland, *Nucl. Acids Res.* (1995) 23(10):1664-1670; *Bioorgan. Med. Chem* (1991) 7:2351; Sugiyama, *Chem. Res. Toxicol.* (1994) 7: 673-83; Horn, *Nucl. Acids. Res.*, (1988) 16:11559-71), and use of enzymes that catalyze cleavage of polynucleotides at abasic sites, for example AP endonucleases (also called "apurinic, apyrimidinic endonucleases") (e.g., *E. coli* Endonuclease IV, available from Epicentre Tech., Inc, Madison WI), *E. coli* endonuclease III or endonuclease IV, *E. coli* exonuclease III in the presence of calcium ions. *See, e.g.* Lindahl, *PNAS* (1974) 71(9):3649-3653; Jendrisak, U.S. Patent No. 6,190,865 B1; Shida, *Nucleic Acids Res.* (1996) 24(22):4572-76; Srivastava, *J. Biol Chem.* (1998) 273(13):21203-209; Carey, *Biochem.* (1999) 38:16553-60; *Chem Res Toxicol* (1994) 7:673-683. As used herein "agent" encompasses reaction conditions such as heat. In one embodiment, the AP endonuclease, *E. coli* endonuclease IV, is used the cleave the phosphodiester backbone at an abasic site. In another embodiment, cleavage is with an amine, such as N, N'-dimethylethylenediamine. *See, e.g.* McHugh and Knowland, *supra.*

[0123] Generally, cleavage is between the nucleotide immediately 5' to the abasic residue and the abasic residue, or between the nucleotide immediately 3' to the abasic residue and the abasic residue (though, as explained herein, 5' or 3' cleavage of the phosphodiester backbone may or may not result in retention of the phosphate group 5' or 3' to the abasic site, respectively, depending on the fragmentation agent used). As is well known in the art, cleavage can be 5' to the abasic site (such as endonuclease IV treatment which generally results in cleavage of the backbone at a location immediately 5' to the abasic site between the 5'-phosphate group of the abasic residue and the deoxyribose ring of the adjacent nucleotide, generating a free 3' hydroxyl group on the adjacent nucleotide), such that an abasic site is located at the 5' end of the resulting fragment. Cleavage can also be 3' to the abasic site (e.g., cleavage between the deoxyribose ring and 3' -phosphate group of the abasic residue and the deoxyribose ring of the adjacent nucleotide, generating a free 5' phosphate group on the deoxyribose ring of the adjacent nucleotide), such that an

IDT Exhibit 1020.7994

abasic site is located at the 3' end of the resulting fragment. Treatment under basic conditions or with amines (such as N, N'-dimethylethylenediamine) results in cleavage of the phosphodiester backbone immediately 3' to the abasic site. In addition, more complex forms of cleavage are also possible, for example, cleavage such that cleavage of the phosphodiester backbone and cleavage of (a portion of) the abasic nucleotide results. For example, under certain conditions, cleavage using chemical treatment and/or thermal treatment may comprise a β-elimination step which results in cleavage of a bond between the abasic site deoxyribose ring and its 3' phosphate, generating a reactive α,β-unsaturated aldehyde which can be labeled or can undergo further cleavage and cyclization reactions. *See, e.g.* Sugiyama, *Chem. Res. Toxicol.* (1994) 7: 673-83; Horn, *Nucl. Acids. Res.*, (1988) 16:11559-71. It is understood that more than one method of cleavage can be used, including two or more different methods which result in multiple, different types of cleavage products (e.g., fragments comprising an abasic site at the 3' end, and fragments comprising an abasic site at the 5' end).

[0124] Generally, cleavage of the backbone at an abasic site is general, specific or selective cleavage (in the sense that the agent (such as an enzyme) capable of cleaving the backbone at an abasic site specifically or selectively cleaves the base portion of a particular non-canonical nucleotide), whereby greater than about 98%, about 95%, about 90% , about 85%, or about 80% of the cleavage is at an abasic site. However, extent of cleavage can be less. Thus, reference to specific cleavage is exemplary. General, specific or selective cleavage is desirable for control of the fragment size in the methods of generating labeled polynucleotide fragments of the invention. In some embodiments, reaction conditions can be selected such that the cleavage reaction is performed in the presence of a large excess of reagents and allowed to run to completion with minimal concern about excessive cleavage of the polynucleotide (i.e., while retaining a desired fragment size, which is determined by spacing of the incorporated non-canonical nucleotide, during the synthesis step, above). In other embodiments, extent of cleavage can be less, such that polynucleotide fragments are generated comprising an abasic site at an end and an abasic site(s) within or internal to the polynucleotide fragment (i.e., not at an end). As disclosed herein, polynucleotide fragments comprising internal abasic sites are useful e.g., in embodiments involving immobilization of a labeled polynucleotide (wherein one abasic site is used for immobilization and another abasic site(s) are labeled at an abasic site).

[0125] As noted herein, the frequency of incorporation of non-canonical nucleotides into the polynucleotide relates to the size of fragment produced using the methods of the invention because the spacing between non-canonical nucleotides in the polynucleotide comprising a non-canonical nucleotide, as well as the reaction conditions

IDT Exhibit 1020.7995

selected, determines the approximate size of the resulting fragments (following cleavage of a base portion of a non-canonical nucleotide, whereby an abasic site is generated, and cleavage of the backbone at the abasic site as described herein). Generally, suitable fragment sizes are about 5, 10, 15, 20, 25, 30, 40, 50, 65, 75, 85, 100, 123, 150, 175, 200, 225, 250, 300, 350, 400, 450, 500, 550, 600, 650 or more nucleotides in length. In some embodiments, the fragment is about 200 nucleotides, about 100 nucleotides, or about 50 nucleotides in length. In another embodiment, the size of a population of fragments is about 50 to 200 nucleotides. It is understood that the fragment size is approximate, particularly when populations of fragments are generated, because the incorporation of a non-canonical nucleotide (which relates to the fragment size following cleavage) will vary from template to template, and also between copies of the same template. Thus, fragments generated from same starting material (such as a single polynucleotide template) may have different (and/or overlapping) sequence, while still having the same approximate size or size range.

[0126] Following cleavage of the polynucleotide backbone at the abasic site, every fragment will comprise one abasic site (if cleavage is completely efficient), except for either the 5'- or 3'-most fragment, which will lack an abasic site depending on the cleavage agent. If the cleavage is 5' to the abasic site, the 5' most fragment will not comprise an abasic site. If cleavage is 3' to the abasic site, the 3' most fragment will not comprise an abasic site. If it is desired to incorporate an abasic site into a 5'-most fragment, (if the synthesis step requires a primer(s)), a primer comprising a non-canonical nucleotide can be used, as discussed herein, and the resulting abasic site in the primer will be cleaved. Generally, if cleavage of the phosphodiester backbone is 5' to the abasic residue, the abasic site should be incorporated at the 5' end of the primer (or the DNA portion of the primer, if an RNA-DNA composite primer is used, see Kurn, U.S. patent No. 6,251,639 B1).

Labeling the abasic site and detection

[0127] The abasic site is labeled, whereby a polynucleotide (or polynucleotide fragment) comprising a label is generated. In some embodiments, a polynucleotide fragment comprising an abasic site is contacted with an agent capable of labeling at the abasic site; whereby labeled fragments of the polynucleotide are generated. As used herein, a "label" (interchangeably called a "detectable moiety") is associated with a polynucleotide, such that the polynucleotide comprising an abasic site is associated with a label.

IDT Exhibit 1020.7996

[0128] Thus, in some embodiments, the label associates with a chemical structure remaining following treatment of a non-canonical nucleotide (present in a polynucleotide chain) with an agent (e.g., an enzyme, or acidic conditions, or a chemical reagent) capable of effecting cleavage of a base portion of a non-canonical nucleotide. In embodiments involving fragmentation, the label associates with any chemical structure remaining following treatment of a non-canonical nucleotide (present in a polynucleotide chain) with an agent (e.g., an enzyme, or acidic conditions, or a chemical reagent) capable of effecting cleavage of a base portion of a non-canonical nucleotide, and following treatment with an agent capable of cleaving the backbone at the abasic site. In one embodiment, the label covalently bonds with a reactive aldehyde form of a hemiacetal ring in an abasic site. In some embodiments, labeling "at" an abasic site encompasses labels that bind to an abasic site, but do not bind to the intact (uncleaved) non-canonical nucleotide (whether incorporated or present as a single non-canonical nucleotide). In some embodiment, labeling "at" an abasic site specifically excludes labels that associate (e.g., covalently bind) with a phosphate group of a nucleotide (or polynucleotide) or a phosphate group of an abasic site. As made clear from the disclosure herein, "label" refers to any component of a labeling system.

[0129] The embodiment shown in Figure 1 illustrates cleavage of the phosphodiester backbone at an abasic site of the polynucleotide comprising the abasic site, whereby a cleaved polynucleotide fragment is produced, then covalently or non-covalently associating a label with the cleaved fragment, such that labeled polynucleotide fragments are generated. It is understood that cleavage of the phosphodiester backbone at the abasic site, and labeling at an abasic site can be performed in any order, or simultaneously (for example, as disclosed in Example 4, herein).

[0130] The label can be detectable, or the label can be indirectly detected, for example as when the label (attached at an abasic residue) is covalently or non-covalently associated with another moiety which is itself detected. For example, biotin can be attached to the label capable of associating with the abasic site. In another example, an antibody (that can be detectably labeled) binds the label that is attached at the abasic site. In some embodiments, the label comprises an organic molecule, a hapten, or a particle (such as a polystyrene bead). In some embodiments, the label is detected using antibody binding, biotin binding, or via fluorescence or enzyme activity. In some embodiments, the detectable signal is amplified.

[0131] Generally, labeling at an abasic site is general, specific, or selective labeling (in the sense that the agent capable of labeling at an abasic site specifically or selectively labels the abasic site), whereby greater than about 98%, about 95%, about 93%,

IDT Exhibit 1020.7997

about 90%, about 85%, or about 80% of the labels bind abasic sites. However, extent of labeling can be less. Thus, reference to specific labeling is exemplary. Generally, reaction conditions are selected such that the reaction in which the abasic site(s) are labeled can run to completion.

[0132] In some embodiments, labeled polynucleotide fragments are produced which each comprise a single label (to the extent that cleavage of the phosphodiester backbone is generally complete, in the sense that many or essentially all of the polynucleotide fragments comprise a single abasic site). This aspect is useful in quantitating level of hybridization, because signal is proportional to number of bound fragments, and does not relate to the length of the hybridizing fragment or the number of labels per fragment. Thus, hybridization intensity can generally be directly compared, regardless of fragment length. This offers an advantage over prior art methods in which nucleic acid fragments are labeled with multiple detectable moieties, e.g., incorporation of labeled nucleotides, and other methods of directly and indirectly detecting incorporated nucleotides. These methods generally result in multiple labels per hybridizing fragment, and thus are generally less suitable for quantitative applications. Multiple labels per nucleic acid can result in quenching, and potential interference with hybridization kinetics (due to the presence of multiple labeled moieties per nucleic acid).

[0133] In another embodiment, labeled fragments are produced which comprise a labeled abasic site at an end (such as the 3' end and/or the 5' end) and a labeled internal abasic site.

[0134] Methods and reaction conditions for labeling abasic sites are known in the art. For example, a common functional group exposed in an abasic site (and therefore suitable for use in labeling) is the highly reactive aldehyde form of the hemiacetal ring which can be covalently or noncovalently attached to a label using reaction conditions that are known in the art. Many labels comprise substituted hydrazines or hydroxylamines which readily form imine bonds with aldehydes, for example, 5-(((2-(carbohydrazino)-methyl)thio)acetyl)aminofluorescein, aminooxyacetyl hydrazide (FARP). *See* Makrogiorgos, WO 00/39345. The stable oxime formed by this compound can be detected directly by fluorescence or the signal can be amplified using an antibody-enzyme conjugate. *See, e.g.,* Srivastava, *J. Biol. Chem.* (1998) 273(33): 21203-209; Makrigiorgos, *Int J. Radiat. Biol.* (1998) 74(1):99-109; Makriogiorgos, U.S. Patent No. 6,174,680 B1; Makrogiorgos, WO 00/39345. Suitable sidechains (present on the substrate) to react with the aldehyde (of the abasic site) include at least the following: substituted hydrazines, hydrazides, or hydroxylamines (which readily form imine bonds with aldehydes), and the related semicarbazide and thiosemicarbazide groups, and other amines which can form

IDT Exhibit 1020.7998

stable carbon-nitrogen double bonds, that can catalyze simultaneous cleavage and binding (*see* Horn, *Nucl. Acids. Res.*, (1988) 16:11559-71), or can be coupled to form stable conjugates, e.g. by reductive amination. Other methods for attaching a reactive group present in an abasic site to a reactive group present on a label are known in the art. In another example, the abasic site may be chemically modified, then the modified abasic site covalently or non-covalently attached to a suitable reactive group on a substrate. For example, the aldehyde (in the abasic site) can be oxidized or reduced (using methods known in the art), then covalently immobilized to a substrate using, e.g., reductive amination or various oxidative processes.

[0135] Other suitable reagents are known in the art, e.g., fluorescein aldehyde reagents. *See, e.g.,* Boturyn (1999) *Chem. Res. Toxicol.* 12:476-482. *See, also,* Adamczyk (1998) *Bioorg.Med. Chem. Lett.* 8(24):3599-3602; Adamczyk (1999) *Org. Lett.* 1(5):779-781; Kow (2000) *Methods* 22(2):164-169; Molecular Probes Handbook, Section 3.2 (www.probes.com). For example, detectable moieties comprising aminooxy groups can be used. *See*, Boturyn, *supra.* The aminoooxy group readily reacts with the highly reactive aldehyde form of the hemiacetal ring of an abasic site. In one embodiment, the label comprising an aminooxy reactive group is N-(aminooxyacetyl)-N'-(D-biotinoyl) hydrazine, trifluoroacetic acid salt (ARP) (available from Molecular Probes, Eugene OR, catalog No. A-10550). *See, e.g.*, Kubo et al., *Biochem* 31:3703-3708 (1992); Ide et al., *Biochem.* 32:8276-8283 (1993).

[0136] In yet another example, labels comprising a hydrazide linker can be converted to an aminooxy derivative, then used to label abasic sites as described herein. In one embodiment, the label comprises an aminooxy derivatized Alexa Fluor 555 reagent. As shown in Figure 5, use of the aminooxy-derivatized Alexa Fluor 555 resulted in greater labeling efficiency, as well as increased fluorescence as compared to labeling with unmodified Alexa Fluor 555 hydrazide (Order No. A-20501, Molecule Probes, Eugene OR).

[0137] In another example, the abasic site may be chemically modified (before, during or after cleavage of the phosphodiester backbone as described herein), then the modified abasic site detected directly or indirectly. For example, fluorescent cadaverine can be incorporated into an abasic site as described in Horn (*Nucl. Acids. Res.*, (1988) 16:11559-71). In another example, the abasic site may be chemically modified by reaction with NHBA (0-4-nitrobenzyl hydroxylamine), then the NBHA-modified abasic site is detected with an antibody that specifically binds to the NBHA-modified abasic sites *See* Kow et al, WO 92/07951 (1992).

IDT Exhibit 1020.7999

[0138] In another example, the abasic site may be labeled with an antibody (such as a monoclonal or polyclonal antibody or antigen binding fragment). Methods for detecting specific antibody binding are well known in the art.

[0139] In another example, the aldehyde and/or hemiacetal ring may itself be detected, as when for example, detectable signal is generated using chemical or electrochemical reactions specific to those chemical structures, including for example, oxidation reactions, enzymes with dehydrogenase or oxidase activity, and the like. In another example, many aldehydes are substrates for enzymes, such that a detectable product is generated in the presence of the aldehyde. For example, dehydrogenases typically couple oxidation of an aldehyde with reduction of NAD+ which can be detected spectrophotometrically. In another example, glucose oxidases generate hydrogen peroxide in the presence of sugar aldehydes. Hydrogen peroxide is readily detectable by coupling to horseradish peroxidase with suitable substrates. Thus, the invention provides methods for detecting an abasic site.

[0140] Methods of signal detection are known in the art. Signal detection may be visual or utilize a suitable instrument appropriate to the particular label used, such as a spectrometer, fluorimeter, or microscope. For example, where the label is a radioisotope, detection can be achieved using, for example, a scintillation counter, or photographic film as in autoradiography. Where a fluorescent label is used, detection may be by exciting the fluorochrome with the appropriate wavelength of light and detecting the resulting fluorescence, such as by microscopy, visual inspection or photographic film, fluorometer, CCD cameras, scanner and the like. Where enzymatic labels are used, detection may be by providing appropriate substrates for the enzyme and detecting the resulting reaction product. For example, many substrates of horseradish peroxidase, such as o-phenylenediamine, give colored products. Simple colorimetric labels can usually be detected by visual observation of the color associated with the label; for example, conjugated colloidal gold is often pink to reddish, and beads appear the color of the bead. Instruments suitable for high sensitivity detection are known in the art.

[0141] It is understood that the polynucleotide or polynucleotide fragments can be additionally labeled using other methods known in the art, such as incorporation of labeled nucleotide analogs during synthesis of the polynucleotide comprising a non-canonical nucleotide. In addition, following cleavage of the phosphodiester backbone of the polynucleotide comprising an abasic site, either the 5' most or the 3' most fragment will lack an abasic site, depending on the cleavage agent (in embodiments in which the fragmentation reaction goes to completion). However, as discussed herein, if the synthesis step requires primer(s), a labeled primer(s) can be used such that the resulting fragment

IDT Exhibit 1020.8000

comprising a primer is labeled. Suitable labels and methods of labeling primers are known. In addition, a primer comprising a non-canonical nucleotide can be used. Following generation of an abasic site, cleavage of the phosphodiester backbone at the abasic site, and labeling at the abasic site, the fragment comprising at least a portion of the primer will be labeled. Generally, if cleavage of the phosphodiester backbone is 5' to the abasic residue, the abasic site should be incorporated at the 5' end of the primer (or the DNA portion of the primer, if a composite primer is used, *see* Kurn , U.S. Patent No. 6,251,639 B1); U.S. Patent Publication No. 2003/0087251 A1.

[0142] Labeled polynucleotide fragments can be immobilized to a substrate, as described herein.

Methods for labeling nucleic acids

[0143] The invention provides methods for generating labeled nucleic acid(s). The methods generally comprise generation of a polynucleotide comprising at least one non-canonical nucleotide, cleavage of a base portion of the non-canonical nucleotide present in the polynucleotide with an agent capable of cleaving a base portion of the non-canonical nucleotide; and labeling the abasic site, whereby labeled polynucleotide(s) is generated. Generally, the polynucleotide comprising a non-canonical nucleotide is labeled at the site of incorporation of non-canonical nucleotides in the polynucleotide (following generation of an abasic site by cleavage of a base portion of a non-canonical nucleotide). The methods of the invention generate labeled polynucleotide(s), which are useful for, for example, hybridization to a microarray and other uses described herein.

[0144] The methods involve the following steps: (a) synthesizing a polynucleotide from a template in the presence of at least one non-canonical nucleotide (interchangeably termed "non-canonical deoxyribonucleoside triphosphate"), whereby a polynucleotide comprising a non-canonical nucleotide is generated; (b) contacting the polynucleotide comprising a non-canonical nucleotide with an agent capable of cleaving a base portion of the non-canonical nucleotide, whereby an abasic site is created; and (c) labeling the abasic site in the polynucleotide comprising the abasic site, whereby labeled polynucleotide(s) is generated. A schematic description of one embodiment of the labeling methods of the invention is given in Figure 2.

[0145] For simplicity, individual steps of the labeling methods are discussed below. It is understood, however, that the steps may be performed simultaneously and in varied order, as discussed herein.

Synthesis of a polynucleotide comprising a non-canonical nucleotide

IDT Exhibit 1020.8001

[0146] The methods involve synthesizing a polynucleotide from a template in the presence of at least one non-canonical nucleotide, whereby a polynucleotide comprising a non-canonical nucleotide is generated. The exemplary embodiment illustrated in Figure 2 illustrates synthesis of a single stranded polynucleotide from a template in the presence of non-canonical nucleotides, such that a single stranded polynucleotide comprising the non-canonical nucleotide is generated. The frequency of incorporation of non-canonical nucleotides into the polynucleotide relates to the frequency of labeled abasic site generated using the methods of the invention because the spacing between non-canonical nucleotides in the polynucleotide comprising a non-canonical nucleotide determines the approximate spacing of the labeled sites in the labeled nucleic acid.

[0147] Generally, the polynucleotide is DNA, though, as noted herein, the polynucleotide can comprise altered and/or modified nucleotides, internucleotide linkages, ribonucleotides, etc. As generally used herein, it is understood that "DNA" applies to polynucleotide embodiments.

[0148] Methods for synthesizing polynucleotides, e.g., single and double stranded DNA, from a template are well known in the art, and are described herein. For convenience, "DNA" is used herein to describe (and exemplify) a polynucleotide.

[0149] Generally, single or double stranded polynucleotide is generated from a template in the presence of all four canonical nucleotides and at least one non-canonical nucleotide under reaction conditions suitable for synthesis of DNA, including suitable enzymes and primers, if necessary. Reaction conditions and reagents, including primers, for synthesizing the polynucleotide comprising a non-canonical nucleotide are known in the art, and discussed herein. As described herein, non-canonical nucleotides are generally capable of polymerization, and capable of being rendered abasic following treatment with a suitable agent capable of generally, specifically or selectively cleaving a base portion of a non-canonical nucleotide. Suitable non-canonical nucleotides are well-known in the art and are described herein. In some embodiments, the polynucleotide comprising a non-canonical nucleotide is synthesized using single primer isothermal amplification, see Kurn, U.S. Patent No. 6,251,639 B1; Kurn, WO02/00938; and/or Ribo-SPIATM, see Kurn, U.S. Patent Publication No. 2003/0087251 A1.

[0150] Conditions for limited and/or controlled incorporation of a non-canonical nucleotide are known in the art and are described herein. The frequency (or proportion) of non-canonical bases in the resulting polynucleotide comprising a non-canonical nucleotide, and thus the frequency of labeling in the labeled polynucleotide, is controlled by variables known in the art, including: frequency of nucleotide(s) corresponding to the non-canonical nucleotide(s) in the template (or other measures of nucleotide content of a sequence, such

IDT Exhibit 1020.8002

as average G-C content), ratio of canonical to non-canonical nucleotide present in the reaction mixture; ability of the polymerase to incorporate the non-canonical nucleotide, relative efficiency of incorporation of non-canonical nucleotide verses canonical nucleotide, and the like.

[0151] Generally, the polynucleotide comprising a non-canonical nucleotide is labeled at the site of incorporation of the non-canonical nucleotide(s) (i.e., at an abasic site, as described herein) present in the synthesized polynucleotide. Thus, the frequency of non-canonical nucleotides in the synthesized polynucleotide generally determines the frequency of labels in the labeled polynucleotide. Generally, a non-canonical base can be incorporated at about every 5, 10, 15, 20, 25, 30, 40, 50, 65, 75, 85, 100, 123, 150, 175, 200, 225, 250, 300, 350, 400, 450, 500, 550, 600, 650 or more nucleotides apart in the resulting polynucleotide comprising a non-canonical nucleotide. In one embodiment, the non-canonical nucleotide is incorporated about every 500 nucleotides. In one embodiment, the non-canonical nucleotide is incorporated about every 100 nucleotides. In another embodiment, the non-canonical nucleotide is incorporated about every 50 nucleotides. In another embodiments, the non-canonical nucleotide is incorporated about every 50 to 200 nucleotides. It is understood that these length generally represent average lengths in a population of polynucleotides generated using the methods of the invention.

[0152] Methods of synthesis are generally template-dependent (as described herein). However, it is understood that non-canonical nucleotides can be incorporated into a polynucleotide as a result of template-independent methods (e.g. ligation, tailing), as described herein.

[0153] The template may be any template from which labeled polynucleotides are desired to be produced. The template includes double-stranded, partially double stranded and single-stranded nucleic acids from any source in purified or unpurified form, as described herein.

[0154] For simplicity, the polynucleotide comprising a non-canonical nucleotide is described as a single nucleic acid. It is understood, however, that the polynucleotide comprising a non-canonical nucleotide can be a single nucleic acid, for example, as produced by reverse transcription, first and second strand cDNA production, or a single cycle of DNA replication. The polynucleotide can also be a population of amplified products (from a few to very many), for example single stranded DNA products produced using single primer isothermal amplification and/or Ribo-SPIATM , see Kurn, U.S. Patent No. 6,251,639 B1; Kurn, U.S. Patent Publication No. 2003/0087251 A1, or double stranded DNA product produced by, for example, PCR. It is further understood that a polynucleotide comprising a non-canonical nucleotide can be a multiplicity (from small to

IDT Exhibit 1020.8003

very large) of different polynucleotide molecules. Such populations can be related in sequence (e.g., member of a gene family or superfamily) or extremely diverse in sequence (e.g., generated from all mRNA, generated from all genomic DNA, etc.). Polynucleotides can also correspond to single sequence (which can be part or all of a known gene, for example a coding region, genomic portion, etc.). Methods, reagents, and reaction conditions for generating specific polynucleotide sequences and multiplicities of polynucleotide sequences are known in the art

Cleaving a base portion of the non-canonical nucleotide to create an abasic site

[0155] The polynucleotide comprising a non-canonical nucleotide (synthesized from a template, as described herein) is treated with an agent (such as an enzyme) capable of generally, specifically or selectively cleaving a base portion of the non-canonical nucleotide to create an abasic site. The embodiment shown in Figure 2 illustrates cleavage of a base portion of the non-canonical nucleotides, whereby an abasic site is created. In some embodiments, the agent (such as an enzyme) catalyzes hydrolysis of the bond between the base portion of the non-canonical nucleotide and a sugar in the non-canonical nucleotide to generate an abasic site comprising a hemiacetal ring and lacking the base (interchangeably called "AP" site), though other cleavage products are contemplated for use in the methods of the invention. Suitable agents and reaction conditions for cleavage of base portions of non-canonical nucleotides are known in the art and are described herein. In one embodiment, uracil-N-glycosylase is used to cleave a base portion of the non-canonical nucleotide.

[0156] Generally, cleavage of base portions of non-canonical nucleotides is general, specific or selective cleavage, whereby greater than about 98%, about 95%, about 90%, about 85%, or about 80% of the base portions cleaved are base portions of non-canonical nucleotides. However, extent of cleavage can be less. Thus, reference to specific cleavage is exemplary. General, specific or selective cleavage is desirable for control of the number of potential labeling sites (and thus the intensity of labeling) in the methods of generating labeled polynucleotides of the invention. Generally, reaction conditions are selected such that the reaction in which the abasic site(s) are created can run to completion.

[0157] For convenience, the synthesis of a polynucleotide comprising a non-canonical nucleotide, and the cleavage of that polynucleotide by an enzyme capable of cleaving a base portion of the non-canonical nucleotide are described as separate steps. It is understood that these steps may be performed simultaneously, except (generally) in the case when a polynucleotide comprising a non-canonical nucleotide must be capable of

IDT Exhibit 1020.8004

serving as a template for further amplification (as in exponential methods of amplification, e.g. PCR).

Labeling the abasic site and detection

[0158] The abasic site is labeled, whereby a polynucleotide (or polynucleotide fragment) comprising a detectable moiety is generated. The embodiment shown in Figure 2 illustrates labeling at the abasic sites of a single stranded polynucleotide comprising abasic sites, such that labeled polynucleotides are produced. As used herein, "detectable moiety" (interchangeably called a label) refers to a covalent or non-covalent association of agent (interchangeably called "labeling") with an abasic site in a polynucleotide such that polynucleotides comprising an abasic site are associated with a detectable signal. Accordingly, in some embodiments, the detectable moiety (label) is covalently or non-covalently associated with an abasic site. In some embodiments, the detectable moiety (label) is directly or indirectly detectable. In some embodiments, the detectable signal is amplified. In some embodiments, the detectable moiety comprises an organic molecule. In other embodiments, the detectable moiety comprises an antibody. In other embodiments, the detectable signal is fluorescent. In other embodiments, the detectable signal is enzymatically generated. Other labeling embodiments are described herein.

[0159] Generally, labeling at an abasic site is general, specific or selective labeling (in the sense that the agent capable of labeling at an abasic site specifically or selectively labels the abasic site), whereby greater than about 98%, about 95%, about 93%, about 90%, about 85%, or about 80% of the labels bind abasic sites. However, extent of labeling can be less. Thus, reference to specific labeling is exemplary. Generally, reaction conditions are selected such that the reaction in which the abasic site(s) are labeled can run to completion.

[0160] Methods and reaction conditions for generally, specifically or selectively labeling abasic sites are known in the art and are described herein. Generally, methods for labeling abasic site which also result in cleavage of a phosphodiester backbone should be avoided, unless simultaneous cleavage and labeling is desired (see, e.g., Horn, *Nucl. Acids. Res.* (1988) 16:11559-71).

[0161] In some embodiments, labeled polynucleotide fragments are produced which each comprise a single label (to the extent that cleavage of the phosphodiester backbone is generally complete, in the sense that many or essentially all of the polynucleotide fragments comprise a single abasic site). In another embodiment, labeled fragments are produced which comprise a labeled abasic site at an end (such as the 3' end and/or the 5' end) and a labeled internal abasic site.

IDT Exhibit 1020.8005

[0162] Methods of detecting detectable signals are known in the art and are described herein. Signal detection may be visual or utilize a suitable instrument appropriate to the particular label used, such as a spectrometer, fluorometer, or microscope. For example, where the label is a radioisotope, detection can be achieved using, for example, a scintillation counter, or photographic film as in autoradiography. Where a fluorescent label is used, detection may be by exciting the fluorochrome with the appropriate wavelength of light and detecting the resulting fluorescence, such as by microscopy, visual inspection or photographic film. Where enzymatic labels are used, detection may be by providing appropriate substrates for the enzyme and detecting the resulting reaction product. Simple colorimetric labels can usually be detected by visual observation of the color associated with the label; for example, conjugated colloidal gold is often pink to reddish, and beads appear the color of the bead.

[0163] It is understood that the polynucleotide or polynucleotide can be additionally labeled using other methods known in the art, such as incorporation of labeled nucleotide analogs during synthesis of the polynucleotide comprising a non-canonical nucleotide. If the synthesis step requires primer(s), a labeled primer(s) can be used. Suitable labels and methods of labeling primers are known. In addition, a primer comprising a non-canonical nucleotide can be used. Following generation of an abasic site, cleavage of the phosphodiester backbone at the abasic site, and labeling at the abasic site, the primer will be labeled.

[0164] Labeled polynucleotide can be immobilized to a substrate as described herein.

Methods for preparing polynucleotides (or fragments thereof) immobilized on a substrate

[0165] The invention provides methods for generating polynucleotides or polynucleotide fragments immobilized on a substrate (interchangeably termed a "surface", herein). The methods generally comprise immobilization of a polynucleotide comprising an abasic site, or fragments thereof (in embodiments involving fragmentation), to a substrate at the abasic site. In some embodiments, the methods provide cleavage of a base portion of a non-canonical nucleotide present in a polynucleotide with an agent capable of cleaving a base portion of the non-canonical nucleotide, whereby an abasic site is created; optionally cleaving the phosphodiester backbone of the polynucleotide at the abasic site, whereby fragments of the synthesized nucleic acid are generated; and immobilizing the polynucleotide, or fragments thereof, on a substrate, wherein the polynucleotide or fragment thereof is immobilized at the abasic site. Optionally, the polynucleotide

IDT Exhibit 1020.8006

comprising an abasic site can be labeled at an abasic site according to the labeling methods described herein. The labeling may be anywhere on the immobilized fragment (for example, as when an internal abasic site is labeled, or an abasic site at a terminus of the polynucleotide is labeled). Generally, the polynucleotide comprising an abasic site is immobilized at the abasic site in the polynucleotide. Thus, as discussed above, the frequency of non-canonical nucleotides in the synthesized polynucleotide generally determines the number of abasic sites available for immobilization to a substrate (and the size range of the fragments produced from the polynucleotide, in embodiments involving cleavage of the phosphodiester backbone). The methods of the invention generate polynucleotides, and fragments thereof, immobilized on a substrate, for example, a microarray. In some embodiments, one or more abasic site(s) are labeled (as described herein) and one or more abasic site(s) are immobilized to a substrate.

[0166] The methods involve the following steps: (a) contacting a polynucleotide comprising a non-canonical nucleotide with an agent capable of cleaving a base portion of the non-canonical nucleotide, whereby an abasic site is created; (b) optionally cleaving a phosphodiester backbone at the abasic site; whereby fragments of the synthesized nucleic acid are generated; (c) optionally labeling a polynucleotide at the abasic site; and (d) immobilizing the polynucleotide (or polynucleotide fragments) on a substrate, wherein the polynucleotide is immobilized to the substrate through the abasic site. In some embodiments, the polynucleotide comprising a non-canonical nucleotide is synthesized from a template in the presence of at least one non-canonical nucleotide. In some embodiments, the polynucleotide comprising a non-canonical nucleotide is synthesized using single primer isothermal amplification or Ribo-SPIA™. *See* Kurn, U.S. Patent No. 6,251,639 B1; Kurn, WO02/00938; Kurn, U.S. Patent Publication No. 2003/0087251 A1. A schematic description of one embodiment of the immobilization methods of the invention is given in Figure 3.

[0167] For simplicity, individual steps of the methods are discussed below. It is understood, however, that the steps may be performed simultaneously and in varied order, as discussed herein. It is also understood that the invention encompasses methods in which the initial, or first, step is any of the steps described herein. For example, the method encompasses embodiments wherein a polynucleotide comprising an abasic site, or a polynucleotide fragment comprising an abasic site, are immobilized to a substrate as described herein.

Preparation of a polynucleotide comprising a non-canonical nucleotide

IDT Exhibit 1020.8007

[0168] In some embodiments, the polynucleotide comprising a non-canonical nucleotide is synthesized from a template in the presence of at least one non-canonical nucleotide, as discussed herein. The embodiment illustrated in Figure 3 illustrates synthesis of a single stranded polynucleotide from a template in the presence of non-canonical nucleotides, such that a single stranded polynucleotide comprising the non-canonical nucleotide is generated, though other embodiments are contemplated by the methods of the invention. Other methods for generating a polynucleotide comprising a non-canonical nucleotide are well known in the art, including tailing, ligation, oligonucleotide synthesis, and the like. *See, e.g.*, Sambrook (1989) "Molecular Cloning: A Laboratory Manual", second edition; Ausebel (1987, and updates) "Current Protocols in Molecular Biology". .

[0169] Generally, the polynucleotide is DNA, though, as noted herein, the polynucleotide can comprise altered and/or modified nucleotides, internucleotide linkages, ribonucleotides, etc. As generally used herein, it is understood that "DNA" applies to polynucleotide embodiments.

[0170] Methods for synthesizing polynucleotides, e.g., single and double stranded DNA, are well known in the art, and include template-dependent and template-independent methods. Examples of template-dependent methods include, for example, single primer isothermal amplification, Ribo-SPIA™, PCR, reverse transcription, primer extension, limited primer extension, replication (including rolling circle replication), strand displacement amplification (SDA), nick translation and, e.g., any method that results in synthesis of the complement of a template sequence such that at least one non-canonical nucleotide can be incorporated into a polynucleotide. *See, e.g.,* Kurn, U.S. Patent No. 6,251,639 B1; Kurn, WO02/00938; Kurn, U.S. Patent Publication No. 2003/0087251 A1; Mullis, U.S. Patent No. 4,582,877; Wallace, U.S. patent No. 6,027,923; U.S. Patent No. 5,508,178; 5,888,819; 6,004,744; 5,882,867; 5,710,028; 6,027,889; 6,004,745; 5,763,178; 5,011,769; *see also* Sambrook (1989) "Molecular Cloning: A Laboratory Manual", second edition; Ausebel (1987, and updates) "Current Protocols in Molecular Biology"; Mullis, (1994) "PCR: The Polymerase Chain Reaction". In one embodiment, the polynucleotide comprising a non-canonical nucleotide is synthesized using single primer isothermal amplification and/or Ribo-SPIA™. *See* Kurn, U.S. Patent No. 6,251,639 B1; Kurn, WO02/00938; Kurn, U.S. Patent Publication No. 2003/0087251 A1. Methods of template independent methods include oligonucleotide synthesis, ligation, and tailing, as described herein.

[0171] Suitable methods include methods that result in one single- or double-stranded (or partially double stranded) polynucleotide comprising a non-canonical

IDT Exhibit 1020.8008

nucleotide (for example, reverse transcription, production of double stranded cDNA, a single round of DNA replication), as well as methods that result in multiple single stranded or double stranded copies or copies of the complement of a template (for example, single primer isothermal amplification, Ribo-SPIA™ or PCR). *See* Kurn, U.S. Patent No. 6,251,639 B1; Kurn, WO02/00938; Kurn, U.S. Patent Publication No. 2003/0087251 A1. One or more methods known in the art can be used to generate a polynucleotide comprising a non-canonical nucleotide. It is understood that the polynucleotide comprising a non-canonical nucleotide can be single-stranded or double-stranded, e.g. single and double stranded DNA, or partially double stranded, and that each strand of a double-stranded polynucleotide can comprises a non-canonical nucleotide. For convenience, "DNA" is used herein to describe (and exemplify) a polynucleotide.

[0172] Reaction conditions and reagents, including primers, for producing the polynucleotide comprising a non-canonical nucleotide are known in the art, and described herein (see, e.g., methods for synthesizing a polynucleotide comprising an abasic site, as described herein).

[0173] Generally, the polynucleotide comprising an abasic site is immobilized to a substrate at the abasic sites, as described herein. Thus, the frequency of non-canonical nucleotides in the polynucleotide relates to the frequency of abasic site generated (in the polynucleotide comprising the non-canonical nucleotide following cleavage of a base portion of the non-canonical nucleotide), and thus the number of abasic sites in the polynucleotide comprising abasic site(s) useful (or available) for immobilization of the polynucleotide according to the method of the invention. Generally, a non-canonical base can be incorporated at about every 5, 10, 15, 20, 25, 30, 40, 50, 65, 75, 85, 100, 123, 150, 175, 200, 225, 250, 300, 350, 400, 450, 500, 550, 600, 650 or more nucleotides apart in the resulting polynucleotide comprising a non-canonical nucleotide. In one embodiment, the non-canonical nucleotide is incorporated about every 500 nucleotides. In one embodiment, the non-canonical nucleotide is incorporated about every 100 nucleotides. In another embodiment, the non-canonical nucleotide is incorporated about every 50 nucleotides. In still other embodiments, the non-canonical nucleotide is incorporated about every 50 to 200 nucleotides. It is understood that these length generally represent average lengths in a population of polynucleotides (or fragments thereof in embodiments involving fragmentation) generated using the methods of the invention.

[0174] In some embodiments, the polynucleotide comprising a non-canonical nucleotide is cleaved at the non-canonical nucleotide(s) (i.e., at an abasic site following cleavage of a base portion of the non-canonical nucleotide) present in the synthesized polynucleotide. Thus, the frequency of non-canonical nucleotides in the polynucleotide

IDT Exhibit 1020.8009

generally determines the size range of the fragments produced from the polynucleotide. Generally, a non-canonical nucleotide can be present at about every 5, 10, 15, 20, 25, 30, 40, 50, 65, 75, 85, 100, 123, 150, 175, 200, 225, 250, 300, 350, 400, 450, 500, 550, 600, 650 or more nucleotides apart in the resulting polynucleotide comprising a non-canonical nucleotide. In one embodiment, the non-canonical nucleotide is incorporated about every 500 nucleotides. In one embodiment, the non-canonical nucleotide is incorporated about every 100 nucleotides. In another embodiment, the non-canonical nucleotide is incorporated about every 50 nucleotides. In still another embodiment, the non-canonical nucleotide is incorporated about every 50 to about every 200 nucleotides. It is understood that these length generally represent average lengths in a population of polynucleotides (or fragments thereof in embodiments involving fragmentation) generated using the methods of the invention. Conditions for limited and/or controlled incorporation of a non-canonical nucleotide are known in the art and are described herein. The frequency (or proportion) of non-canonical bases in the resulting polynucleotide comprising a non-canonical nucleotide, and thus the average size of fragments generated using the methods of the invention (i.e., following cleavage of a base portion of a non-canonical nucleotide, and cleavage of a phosphodiester bond at a non-canonical nucleotide), is controlled by variables known in the art, including: frequency of nucleotide(s) corresponding to the non-canonical nucleotide(s) in the template (or other measures of nucleotide content of a sequence, such as average G-C content), ratio of canonical to non-canonical nucleotide present in the reaction mixture; ability of the polymerase to incorporate the non-canonical nucleotide, relative efficiency of incorporation of non-canonical nucleotide verses canonical nucleotide, and the like. The reaction conditions can be empirically determined, for example, by assessing average fragment size generated using the methods of the invention taught herein.

[0175] The template may be any template from which immobilized polynucleotides (polynucleotide fragments) are desired to be produced, as described herein.

[0176] For simplicity, the polynucleotide comprising a non-canonical nucleotide is described as a single nucleic acid. It is understood, however, that the polynucleotide comprising a non-canonical nucleotide can be a single nucleic acid, for example, as produced by reverse transcription, first and second strand cDNA production, or a single cycle of DNA replication. The polynucleotide can also be a population of amplified products (from a few to very many), for example single stranded DNA products produced using single primer isothermal amplification and/or Ribo-SPIA™, *see* Kurn, U.S. Patent No. 6,251,639 B1; Kurn, U.S. Patent Publication No. 2003/0087251 A1, or double

IDT Exhibit 1020.8010

stranded DNA product produced by, for example, PCR. It is further understood that a polynucleotide comprising a non-canonical nucleotide can be a multiplicity (from small to very large) of different polynucleotide molecules. Such populations can be related in sequence (e.g., member of a gene family or superfamily) or extremely diverse in sequence (e.g., generated from all mRNA, generated from all genomic DNA, etc.). Polynucleotides can also correspond to single sequence (which can be part or all of a known gene, for example a coding region, genomic portion, etc.). Methods, reagents, and reaction conditions for generating specific polynucleotide sequences and multiplicities of polynucleotide sequences are known in the art.

Generating a polynucleotide comprising an abasic site

[0177] A polynucleotide comprising an abasic site can be generated using methods known in the art (e.g., Makrigiorgos, *Int J. Radiat. Biol.* (1998) 74(1):99-109), and as described herein. Generally, a polynucleotide comprising a non-canonical nucleotide (which can be synthesized from a template, as described herein) is treated with an enzyme capable of generally, specifically or selectively cleaving a base portion of the non-canonical nucleotide to create an abasic site. The embodiment shown in Figure 3 illustrates cleavage of a base portion of the non-canonical nucleotides, whereby an abasic site is created. Generally, an agent, such as an enzyme, catalyzes hydrolysis of the bond between the base portion of the non-canonical nucleotide and a sugar in the non-canonical nucleotide to generate an abasic site comprising a hemiacetal ring and lacking the base (interchangeably called "AP" site), though other cleavage products are contemplated for use in the methods of the invention. Suitable agents and reaction conditions for cleavage of base portions of non-canonical nucleotides are known in the art and described herein.

[0178] Generally, cleavage of base portions of non-canonical nucleotides is general, specific or selective cleavage, whereby greater than about 98%, about 95%, about 90% , about 85%, or about 80% of the base portions cleaved are bases portions of non-canonical nucleotides. However, extent of cleavage can be less. Thus, reference to specific cleavage is exemplary. In embodiments involving generation of polynucleotide fragments, specific or selective cleavage is desirable for control of the fragment size in the methods of generating immobilized nucleotide fragments of the invention (i.e., the fragments generated by cleavage of the phosphodiester backbone at an abasic site). Generally, reaction conditions are selected such that the reaction in which the abasic site(s) are created can run to completion.

[0179] For convenience, the synthesis of a polynucleotide comprising a non-canonical nucleotide, and the cleavage of that polynucleotide by an enzyme capable of

IDT Exhibit 1020.8011

cleaving a base portion of the non-canonical nucleotide are described as separate steps. It is understood that these steps may be performed simultaneously, except (generally) in the case when a polynucleotide comprising a non-canonical nucleotide must be capable of serving as a template for further amplification (as in exponential methods of amplification, e.g. PCR).

Cleaving the phosphodiester backbone at the abasic site of the polynucleotide comprising an abasic site

[0180] In some embodiments, the phosphodiester backbone of the polynucleotide is cleaved at the abasic site with an agent capable of effecting cleavage of a backbone at the abasic site, whereby polynucleotide fragments are generated. The embodiment shown in Figure 3 illustrates cleavage of the backbone immediately 5' to the abasic sites of the polynucleotide comprising the abasic sites, whereby cleaved fragments are produced. Cleavage of the backbone at an abasic site is described herein. Suitable enzymes and/or reaction conditions for cleavage of the backbone are well known in the art, and are described herein.

[0181] As noted herein, the frequency of incorporation of non-canonical nucleotides into the polynucleotide relates to the size of fragment produced using the methods of the invention because the spacing between non-canonical nucleotides in the polynucleotide comprising a non-canonical nucleotide determines the approximate size of the resulting fragments (following generation of an abasic site from the non-canonical nucleotide and cleavage of the phosphodiester backbone at the site of incorporation of the non-canonical nucleotide (also termed the abasic site), as described herein). Generally, suitable fragment sizes are about 5, 10, 15, 20, 25, 30, 40, 50, 65, 75, 85, 100, 123, 150, 175, 200, 225, 250, 300, 350, 400, 450, 500, 550, 600, 650 or more nucleotides in length. It is understood that the fragment size is approximate, particularly when populations of fragments are generated, because the incorporation of a non-canonical nucleotide (which relates to the fragment size following cleavage) will vary from template to template, and also between copies of the same template. Thus, fragments generated from same starting material may have different (and/or overlapping) sequence, while still having the same approximate size or size range.

[0182] Generally, cleavage of the backbone at an abasic site is general, specific or selective cleavage (in the sense that the agent (such as an enzyme) capable of cleaving the backbone at an abasic site specifically or selectively cleaves the base portion of a particular non-canonical nucleotide), whereby greater than about 98%, about 95%, about 90%, about 85%, or about 80% of the cleavage is at an abasic site. However, extent of cleavage can be

IDT Exhibit 1020.8012

less. Thus, reference to specific cleavage is exemplary. Generally, about 98%, about 95%, about 90%, about 85%, or about 80% of the abasic sites at the backbone are cleaved. However, extent of cleavage can be less (such that fragments comprising uncleaved abasic sites are produced). In some embodiments, abasic sites are labeled (either before or after immobilization to a substrate, as described herein).

Labeling at an abasic site

[0183] In some embodiments, the polynucleotide, or fragment thereof, is labeled at an abasic site. Labeling is as described herein. As the disclosure herein makes clear, it is understood that labeling and fragmentation steps or labeling and immobilization steps, or labeling and immobilization, and fragmentation steps, can be performed in any order, or simultaneously.

Immobilizing a polynucleotide comprising an abasic site to a substrate

[0184] After generation of the polynucleotide comprising an abasic site, the polynucleotide (or polynucleotide fragment, if the backbone is cleaved) is immobilized to a substrate at the abasic site. In embodiments involving cleavage of the backbone at an abasic site (whereby fragments of the synthesized nucleic acid are generated), the cleaved fragments are immobilized to a substrate at the cleaved abasic site. Figure 3 diagrammatically depicts an embodiment in which a polynucleotide fragment is immobilized to a substrate at the cleaved abasic site. Immobilizing a polynucleotide(s) is useful, for example, to tag an analyte, or to create a microarray. Single stranded polynucleotides (including polynucleotide fragments) are particularly suitable for preparing microarrays comprising the single stranded polynucleotides. Single stranded polynucleotide fragments (in embodiments involving cleavage of the phosphodiester backbone at an abasic site) are advantageous, because the orientation of the fragment with respect to the substrate (upon which the fragment is immobilized) can be controlled by selection of the method used to cleave the phosphodiester backbone, such that an abasic site is positioned at the 3' end of a fragment or at the 5' end of a fragment. Immobilizing polynucleotides in a defined orientation (e.g., at the 3' end, at the 5' end) enhances hybridization of complementary oligonucleotides, and permits a higher density of immobilization.

[0185] The polynucleotide comprising the abasic site is immobilized to a substrate as follows: generally, reagents are used that are capable of covalently or non-covalently attaching a reactive group present in the abasic site to a reactive group present on a substrate. For example, a common functional group exposed in an abasic site (and

IDT Exhibit 1020.8013

therefore suitable for use in labeling) is the aldehyde of the hemiacetal ring which can be covalently or noncovalently attached to a reactive group on a suitable substrate using reaction conditions that are known in the art. Suitable sidechains (present on the substrate) to react with the aldehyde (of the abasic site) include at least the following: substituted hydrazines, hydrazides, or hydroxylamines (which readily form imine bonds with aldehydes), and the related semicarbazide and thiosemicarbazide groups, and other amines which can form stable carbon-nitrogen double bonds, that can catalyze simultaneous cleavage and binding (*see* Horn, *Nucl. Acids. Res.*, (1988) 16:11559-71), or can be coupled to form stable conjugates, e.g. by reductive amination.

[0186] The substrate to which the polynucleotide is to be immobilized can be functionalized with suitable reactive groups using methods known in the art. For example, a solid or semi-solid substrate (e.g., silicon or glass slide) can be coated with polymers (e.g., polyacrylamide, dextran, acrylamide, or latex) comprising hydrazine, hydrazide, or amine derivatized substrates (e.g. semicarbazides). Methods for functionalizing substrates with suitable reactive groups are known in the art, and disclosed in, for example, Luktanov, U.S. Patent No. 6,339,147; Van Ness, U.S. Patent No. 5,667,976; Bangs Laboratories, Inc. TechNote 205 (available at bangslabs.com); Ghosh, *Anal. Biochem* (1989) 178:43-51; O'Shannessy, *Anal. Biochem.* (1990) 191:1-8; Wilchek, *Methods Enzymol.* (1987) 138:429-442; Baumgartner, *Anal. Biochem.* (1989) 181:182-189; Zalipsky, *Bioconjugate Chem.* (1995) 6: 150-165, and references cited therein.

[0187] Methods and reaction conditions for performing these reactions are known in the art. *See, e.g.* Luktanov, U.S. Patent No. 6,339,147; Van Ness, U.S. Patent No. 5,667,976; Bangs Laboratories, Inc. TechNote 205 (available at bangslabs.com); Ghosh, *Anal. Biochem* (1989) 178:43-51; O'Shannessy, *Anal. Biochem.* (1990) 191:1-8; Wilchek, *Methods Enzymol.* (1987) 138:429-442; Baumgartner, *Anal. Biochem.* (1989) 181:182-189; Zalipsky, *Bioconjugate Chem.* (1995) 6: 150-165, and references cited therein. It is appreciated that similar chemistry is described herein with respect to the methods of labeling an abasic site (i.e., embodiments in which a reactive group in the abasic site is covalently or non-covalently attached to a suitable reactive group on a label). *See, e.g.,* Srivastava, *J. Biol. Chem.* (1998) 273(33): 21203-209; Makrigiorgos, *Int J. Radiat. Biol.* (1998) 74(1):99-109; Makriogiorgos, U.S. Patent No. 6,174,680 B1; Makrogiorgos, WO 00/39345.

[0188] In another example, the abasic site may be chemically modified, then the modified abasic site covalently or non-covalently attached to a suitable reactive group on a substrate. For example, the aldehyde (in the abasic site) can be oxidized or reduced (using

IDT Exhibit 1020.8014

methods known in the art), then covalently immobilized to a substrate using, e.g., reductive amination or various oxidative processes.

[0189] The substrate may consist of many materials, limited primarily by capacity to immobilize (or, in some embodiments, capacity for derivatization to immobilize) any of a number of chemically reactive groups and compatibility with the synthetic chemistry used to immobilize the polynucleotide comprising an abasic site. The substrate can be a solid or semi-solid support, which may be made, e.g., from glass, plastic (e.g., polystyrene, polypropylene, nylon), polyacrylamide, nitrocellulose, or other materials such as metals. As described herein, the substrate can be functionalized, if necessary to add a suitable reactive group (to which the abasic site is covalently or non-covalently immobilized). The polynucleotides may also 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, assuming that the substrate is suitably functionalized, as described herein (Khrapko, et al., *DNA Sequence* (1991), 1:375-388)).

[0190] 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 template molecules. In one embodiment the substrate to which the polynucleotide (or fragments thereof) is immobilized is magnetic beads or particles. In another embodiment, the solid substrate comprises an optical fiber. In yet another embodiment, the polynucleotides are dispersed in fluid phase within a capillary which, in turn, is immobilized with respect to a solid phase.

[0191] In another embodiment, the substrate comprises a polypeptide, a protein, a peptide, carbohydrates, cells, microorganisms and fragments and products thereof, an organic molecule, an inorganic molecule, carrier molecules, PEG, amino-dextran, carbohydrates, supramolecular assemblies, organelles, cells, microorganisms, organic molecules, inorganic molecules, or any substance for which immobilization sites for polynucleotides comprising abasic sites naturally exist, can be created (e.g. by functionalizing or otherwise modifying the substrate) or can be developed. In one embodiment, the substrate is a polynucleotide.

[0192] The substrate may be an analyte. Typical analytes may include, but are not limited to antibodies, proteins (including enzymes), peptides, nucleic acid molecules or segments thereof, carrier molecules, PEG, amino-dextran, carbohydrates, supramolecular assemblies, organelles, cells, microorganisms, organic molecules, inorganic molecules, or

IDT Exhibit 1020.8015

any substance for which immobilization sites for polynucleotides comprising abasic sites naturally exist, can be created (e.g. by functionalizing the analyte) or can be developed.

[0193] It is understood that a substrate may be a member(s) of a binding pair. Non-limiting examples of a binding pair include a protein:protein binding pair, and a protein: antibody binding pair. In another embodiment, polynucleotides (or fragments thereof) are immobilized to (tag) a molecular library of substrates, e.g., a molecular library of chemical compounds, a phage peptide display library, or a library of antibodies.

[0194] In some embodiments, the substrate (to which the polynucleotide is immobilized) is an enzyme, such that enhanced detection of hybridization of the polynucleotide is provided. For example, a polynucleotide immobilized to an enzyme can be hybridized to a microarray, and hybridized polynucleotide detected by contacting the microarray with a defined substrate.

[0195] In embodiments of the invention involving cleavage of the phosphodiester backbone at an abasic site (whereby fragments of the synthesized nucleic acid are generated), the cleaved fragments can also be immobilized to a substrate using any method known in the art for immobilization of a nucleic acid to a substrate.

[0196] For example, single or double stranded polynucleotide fragments (generally single stranded) can be immobilized to a solid or semi-solid support or substrate, which may be made, e.g., from plastics, ceramics, metals, acrylamide, cellulose, nitrocellulose, glass, polystyrene, polyethylene vinyl acetate, polypropylene, polymethacrylate, polyethylene, polyethylene oxide, polysilicates, polycarbonates, Teflon®, fluorocarbons, nylon, silicon rubber, polyanhydrides, polyglycolic acid, polylactic acid, polyorthoesters, polypropylfumerate, collagen, glycosaminoglycans, and polyamino acids, and other materials. Substrates may be two-dimensional or three-dimensional in form, such as gels, membranes, thin films, glasses, plates, cylinders, beads, magnetic beads, optical fibers, woven fibers, microtiter well, capillaries, etc.. For example, the fragments can be contacted with a solid or semi-solid substrate, such as a glass slide, which is coated with a reactive group which will form a covalent link with the reactive group that is on the polynucleotide fragment and become covalently immobilized to the substrate.

[0197] Microarrays comprising the nucleotide fragments can be fabricated using a Biodot (BioDot, Inc. Irvine, CA) spotting apparatus and aldehyde-coated glass slides (CEL Associates, Houston, TX). Polynucleotide fragments can be spotted onto the aldehyde-coated slides following suitable functionalization, and processed according to published procedures (Schena et al., *Proc. Natl. Acad. Sci. U.S.A.* (1995) 93:10614-10619), provided suitable care is taken to avoid interfering with other desired reactions at the abasic sites.

IDT Exhibit 1020.8016

Arrays can also be printed by robotics onto glass, nylon (Ramsay, G., *Nature Biotechnol.* (1998), 16:40-44), polypropylene (Matson, et al., *Anal Biochem.* (1995), 224(1):110-6), and silicone slides (Marshall, A. and Hodgson, J., *Nature Biotechnol.* (1998), 16:27-31). Other approaches to array assembly include fine micropipetting within electric fields (Marshall and Hodgson, *supra*), and spotting the polynucleotides directly onto positively coated plates. Methods such as those using amino propyl silane surface chemistry are also known in the art, as disclosed at http://www.cmt.corning.com and http://cmgm.stanford.edu/pbrown/.

[0198] 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). 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.

[0199] Methods for immobilizing polynucleotide fragments to analytes (as described herein) are known in the art. *See, e.g.*, U.S. Patent Nos. 6,309,843; 6,306,365; 6,280,935; 6,087,103 (and methods discussed therein).

[0200] It is understood that the polynucleotide fragments prepared according to the method of the invention can comprise a free 3'-hydroxyl or a free 5'-hydroxyl group. Methods and reaction conditions for immobilization of nucleotide through free hydroxyl groups are well known in the art. See, e.g., U.S. Patent Nos. 6,169,194; 5,726,329.

Reaction conditions and detection

[0201] Appropriate reaction media and conditions for carrying out the methods of the invention are those that permit nucleic acid synthesis 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. 6,190,865; 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; PCT Pub. No. WO99/42618; *Mol. Cell Probes* (1992) 251-6; and *Anal. Biochem.* (1993) 211:164-9. For example, a buffer may be Tris buffer, although other buffers can also be used as long as the buffer components are non-inhibitory to enzyme components of the methods of the invention. The pH is preferably from about 5 to about 11, more preferably from about 6 to about 10, even more preferably from about 7 to about 9, and most preferably from about 7.5 to about 8.5. The reaction medium can also include bivalent metal ions such as Mg^{2+} or Mn^{2+}, at a final concentration of free ions that is within the range of from about 0.01 to about 15 mM, and most preferably from about 1 to 10 mM.

IDT Exhibit 1020.8017

The reaction medium can also include other salts, such as KCl or NaCl, that contribute to the total ionic strength of the medium. For example, the range of a salt such as KCl is preferably from about 0 to about 125 mM, more preferably from about 0 to about 100 mM, and most preferably from about 0 to about 75 mM. The reaction medium can further include additives that could affect performance of the amplification reactions, but that are not integral to the activity of the enzyme components of the methods. Such additives include proteins such as BSA, single strand binding proteins (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 (if any) can also be included. Any aspect of the methods of the invention can occur at the same or varying temperatures. The synthesis reactions (particularly, primer extension other than the first and second strand cDNA synthesis steps, and strand displacement) can be performed isothermally, which avoids the cumbersome thermocycling process. The synthesis reaction is carried out at a temperature that permits hybridization of the oligonucleotides (primer) of the invention to the template polynucleotide and primer extension products, and that does not substantially inhibit the activity of the enzymes employed. The temperature can be in the range of 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.

[0202] Nucleotides, including non-canonical nucleotides (or other nucleotide analogs), that can be employed for synthesis of the nucleic acid comprising a non-canonical nucleotide 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. The oligonucleotide components of the synthesis reactions of the invention are generally in excess of the number of template nucleic acid sequence to be replicated. They can be provided at about or at least about any of the following: 10, 10^2, 10^4, 10^6, 10^8, 10^{10}, 10^{12} times the amount of target nucleic acid. Composite primers can 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.

IDT Exhibit 1020.8018

[0203] Optionally, the polynucleotide comprising a non-canonical nucleotide can be treated with hydroxylamine (or any other suitable agent) to remove any aldehydes that may have formed spontaneously in the nucleic acid. *See, e.g.*, Makrogiorgos, WO00/39345.

[0204] For convenience, the synthesis of a polynucleotide comprising a non-canonical nucleotide, and the cleavage of a base portion of that polynucleotide by an enzyme capable of cleaving a base portion of the non-canonical nucleotide, and the cleavage of the phosphodiester backbone at the abasic site, are described as separate steps. It is understood that these steps may be performed simultaneously, except (generally) in the case when a polynucleotide comprising a non-canonical nucleotide must be capable of serving as a template for further amplification (as in exponential methods of amplification, e.g. PCR).

[0205] Appropriate reaction media and conditions for carrying out the cleavage of a base portion of a non-canonical nucleotide according to the methods of the invention are those that permit cleavage of a base portion of a non-canonical nucleotide. Such media and conditions are known to persons of skill in the art, and are described in various publications, such as Lindahl, *PNAS* (1974) 71(9):3649-3653; Jendrisak, U.S. Patent No. 6,190,865 B1; U.S. Patent No. 5,035,996; U.S. Patent No. 5,418,149. For example, buffer conditions can be as described above with respect to polynucleotide synthesis. In one embodiment, UDG (Epicentre Technologies, Madison WI) is added to a nucleic acid synthesis reaction mixture, and incubated at 37°C for 20 minutes. In one embodiment, the reaction conditions are the same for the synthesis of a polynucleotide comprising a non-canonical nucleotide and the cleavage of a base portion of the non-canonical nucleotide. In another embodiment, different reaction conditions are used for these reactions. In some embodiments, a chelating regent (e.g. EDTA) is added before or concurrently with UNG in order to prevent the polymerase from extending the ends of the cleavage products.

[0206] In embodiments involving cleavage of the phosphodiester backbone, appropriate reaction media and conditions for carrying out the cleavage of the phosphodiester backbone at an abasic site according to the methods of the invention are those that permit cleavage of the phosphodiester backbone at an abasic site. Such media and conditions are known to persons of skill in the art, and are described in various publications, such as *Bioorgan. Med. Chem* (1991) 7:2351; Sugiyama, *Chem. Res. Toxicol.* (1994) 7: 673-83; Horn, *Nucl. Acids. Res.*, (1988) 16:11559-71); Lindahl, *PNAS* (1974) 71(9):3649-3653; Jendrisak, U.S. Patent No. 6,190,865 B1; Shida, *Nucleic Acids Res.* (1996) 24(22):4572-76; Srivastava, *J. Biol Chem.* (1998) 273(13):21203-209; Carey, *Biochem.* (1999) 38:16553-60; *Chem Res Toxicol* (1994) 7:673-683. For example, *E. coli*

IDT Exhibit 1020.8019

AP endonuclease IV is added to reaction conditions as described above. AP Endonuclease IV can be added at the same or different time as the agent (such as an enzyme) capable of cleaving the base portion of a non-canonical nucleotide. For example, AP Endonuclease IV can be added at the same time as UNG, or at different times. A reaction mixture suitable for simultaneous UNG treatment and N, N'-dimethylethylenediamine treatment is described in Example 4 herein.

[0207] In another example, nucleic acids containing abasic sites are heated in a buffer solution containing an amine, for example, 25mM Tris-HCl and 1-5 mM magnesium ions, for 10-30 minutes at 70°C to 95°C. Alternatively, 1.0 M piperidine (a base) is added to polynucleotide comprising an abasic site which has been precipitated with ethanol and vacuum dried. The solution is then heated for 30 minutes at 90°C and lyophilized to remove the piperidine. In another example, cleavage is effected by treatment with basic solution, e.g., 0.2 M sodium hydroxide at 37° for 15 minutes. *See* Nakamura (1998) *Cancer Res.* 58:222-225. In yet another example, incubation at 37C with 100 mM N, N'-dimethylethylenediamine acetate, pH 7.4 is used to cleave. *See* McHugh and Knowland, (1995) *Nucl. Acids Res.* 23(10) 1664-1670.

[0208] In one embodiment, the reaction conditions are the same for the cleavage of a base portion of the non-canonical nucleotide and for the cleavage of the phosphodiester backbone at abasic sites. In another embodiment, different reaction conditions are used for these reactions.

[0209] In embodiments involving labeling at an abasic site, appropriate reaction media and conditions for carrying out the labeling at an abasic site according to the methods of the invention are those that permit labeling at an abasic site. Such reaction mixtures and conditions are known to persons of skill in the art, and are described in various publications, such as Makrogiorgos, WO 00/39345; Srivastava, *J. Biol. Chem.* (1998) 273(33): 21203-209; Makrigiorgos, *Int J. Radiat. Biol.* (1998) 74(1):99-109; Makriogiorgos, U.S. Patent No. 6,174,680 B1; Makrogiorgos, WO 00/39345; Boturyn (1999) *Chem. Res. Toxicol.* 12:476-482. *See, also*, Adamczyk (1998) *Bioorg. Med. Chem. Lett.* 8(24):3599-3602; Adamczyk (1999) *Org. Lett.* 1(5):779-781; Kow (2000) *Methods* 22(2):164-169; Molecular Probes Handbook, Section 3.2 (www.probes.com); Horn (*Nucl. Acids. Res.*, (1988) 16:11559-71). For example, 5-(((2-(carbohydrazino)-methyl)thio)acetyl)aminofluorescein, aminooxyacetyl hydrazide (FARP); N-(aminooxyacetyl)-N'-(D-biotinoyl) hydrazine, trifluorecetic acid salt (ARP); Alexa Fluor 555 (Molecular Probes); aminooxy-derivatized Alexa Fluor 555; and other aldehyde-reactive reagents can be reacted with a polynucleotide comprising abasic sites. The buffer can be sodium citrate or sodium phosphate buffer, though other buffers are acceptable as

IDT Exhibit 1020.8020

long as the buffer components are non-inhibitory to enzyme components and/or desired chemical reactions used in the methods of the invention. The pH is preferably from about 3 to about 11, more preferably from about 4 to about 10, even more preferably from about 4 to about 8, and most preferably from about 4 to about 7. The reaction can be conducted at room temperature to 37°, though other temperatures are suitable as long as the temperature is non-inhibitory to enzyme components and/or desired chemical reactions used in the methods of the invention. Generally, the label (e.g. ARP or FARP) is added at about 1-10 mM, preferable 2-5 mM, though other concentrations are suitable. If an antibody label is used, conditions for antibody binding are well-known in the art, and can be as described herein. Optionally, a stop buffer can be used that neutralizes the pH of the labeling reaction, thereby stopping the labeling reaction and optionally, facilitating subsequent purification of labeled product.

[0210] In embodiments involving immobilization of a polynucleotide at an abasic site, appropriate reaction media and conditions for carrying out the immobilization at an abasic site according to the methods of the invention are those that permit immobilization at an abasic site. Such reaction mixtures and conditions are known to persons of skill in the art, and are described in various publications, such as Luktanov, U.S. Patent No. 6,339,147; Van Ness, U.S. Patent No. 5,667,976; Bangs Laboratories, Inc. TechNote 205 (available at bangslabs.com); Ghosh, *Anal. Biochem* (1989) 178:43-51; O'Shannessy, *Anal. Biochem.* (1990) 191:1-8; Wilchek, *Methods Enzymol.* (1987) 138:429-442; Baumgartner, *Anal. Biochem.* (1989) 181:182-189; Zalipsky, *Bioconjugate Chem.* (1995) 6: 150-165, and references cited therein. In some cases, the initial product can be stabilized by reduction with sodium cyanoborohydride or similar agents known in the art. *See, e.g.*, O'Shannessy, supra.

[0211] In one embodiment, the foregoing components are added simultaneously at the initiation of the synthesis step of the fragmentation and/or labeling and/or immobilization processes. In another embodiment, components are added in any order prior to or after appropriate timepoints during the synthesis step. Such timepoints, some of which are noted below, can be readily identified by a person of skill in the art. In these embodiments, the reaction conditions and components may be varied between the different reactions.

[0212] The fragmenting and/or labeling and/or immobilization 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. 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.

IDT Exhibit 1020.8021

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 or other reagent. In some embodiments, one or more of the components of the reactions is replenished prior to, at, or following the resumption of the reactions. Alternatively, the reaction can be allowed to proceed (i.e., from start to finish) without interruption.

The 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.

Compositions and kits of the invention

[0213] 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 primer (which can be an RNA-DNA composite primer), non-canonical nucleotides, an agent (such as an enzyme) capable of cleaving a base portion of a non-canonical nucleotide, optionally an agent (such as an enzyme) capable of effecting cleavage of a phosphodiester backbone at an abasic site, and an agent capable of labeling an abasic site. In another example, the invention provides a composition comprising a polynucleotide comprising a non-canonical nucleotide, said polynucleotide synthesized from a template, and an agent capable of labeling an abasic site. In still another example, the composition comprises a primer (which can be a RNA-DNA composite primer), dUTP, UNG, (optionally) E. coli Endonuclease IV, and N-(aminooxyacetyl)-N'-(D-biotinoyl) hydrazine, trifluorecetic acid salt (ARP).

[0214] In another embodiment, the invention provides a composition comprising a composite primer, said composite comprising a DNA portion and a 5' RNA portion; and a non-canonical nucleotide (such as dUTP). In another embodiment, the composition comprises a composite primer, said composite primer comprising an RNA portion and a 3' DNA portion; and an agent (such as UNG) that is capable of cleaving a base portion from a non-canonical nucleotide. In another embodiment, the composition comprises a composite primer, said composite primer comprising an RNA portion and a 3' DNA portion; and an agent (such as an amine, such as N, N'-dimethylethylenediamine) capable of cleaving the phosphodiester back at an abasic site. In other embodiments, the composition comprises a composite primer, said composite primer comprising an RNA portion and a 3' DNA portion; and an agent that labels an abasic site (such as ARP). In other embodiments, the composition comprises a composite primer, said composite primer comprising an RNA portion and a 3' DNA portion; dUTP; and UNG. In still other embodiments, the

IDT Exhibit 1020.8022

composition comprises a composite primer, said composite primer comprising an RNA portion and a 3' DNA portion; dUTP; UNG; and ARP. In still other embodiments, the composition comprises a composite primer, said composite primer comprising an RNA portion and a 3' DNA portion; dUTP; UNG; and N, N'-dimethylethylenediamine. In still other embodiments, the composition comprises a composite primer, said composite primer comprising an RNA portion and a 3' DNA portion; dUTP; UNG; N, N'-dimethylethylenediamine; and ARP.

[0215] In still other embodiments, the invention provides a composition comprising a composite primer, said composite primer comprising an RNA portion and a 3' DNA portion; a non-canonical nucleotide; an agent (such as an enzyme) capable of cleaving a base portion of a non-canonical nucleotide; an agent (such as an enzyme) capable of cleaving a phosphodiester backbone at an abasic site; and an agent capable of labeling an abasic site. In some embodiment, the composition further provides a suitable substrate for immobilization. In some embodiments, the RNA portion is 5' to the DNA portion, the 5' RNA portion of the composite primer is adjacent to the 3' DNA portion, the RNA portion of the composite primer consists of about 10 to about 20 nucleotides and the DNA portion of the composite primer consists of about 7 to about 20 nucleotides. In still other embodiments, the composition comprises a second, different composite primer. In some embodiments, the RNA portion of the composite primer comprises the following ribonucleotide sequence: 5'- GACGGAUGCGGUCU-3'.

[0216] In still another embodiment, invention provides a composition comprising (a) UNG; (b) N, N'-dimethylethylenediamine; and (c) ARP. In other embodiments, the invention provides a composition comprising (a) UNG; (b) N, N'-dimethylethylenediamine; (c) ARP; (d) dUTP; (e) a mixture of dATP, dTTP, dCTP, and dGTP; (f) a DNA polymerase; (g) a composite primer, wherein the composite primer comprises a 5' RNA portion and a 3' DNA portion. In still other embodiments, the invention provides a composition comprising (a) UNG; (b) N, N'-dimethylethylenediamine; (c) ARP; (d) dUTP; (e) a mixture of dATP, dTTP, dCTP, and dGTP; (f) a DNA polymerase; (g) RNAse H; (h) a composite primer, wherein the composite primer comprises a 5' RNA portion and a 3' DNA portion. In yet another embodiment, the invention provides a composition comprising (a) UNG; (b) N, N'-dimethylethylenediamine; (c) ARP; (d) dUTP; (e) a mixture of dATP, dTTP, dCTP, and dGTP; (f) a DNA polymerase; (g) RNAse H; (h) a composite primer, wherein the composite primer comprises a 5' RNA portion and a 3' DNA portion (i) $MgCl_2$ solution; (j) acetic acid solution; and optionally, (k) a stop buffer comprising 1.5M Tris, pH 8.5. In some embodiments, the dUTP and the mixture of dATP, dTTP, dCTP, and cGTP are

IDT Exhibit 1020.8023

combined. In some embodiments, the DNA polymerase and RNAse H are provided as a mixture. In some embodiments, the RNA portion of the composite primer is 5' with respect to the 3' DNA portion, the 5' RNA portion is adjacent to the 3' DNA portion, the RNA portion of the composite primer consists of about 10 to about 20 nucleotides and the DNA portion of the composite primer consists of about 7 to about 20 nucleotides. In still other embodiments, the composition comprises a second, different composite primer. In some embodiments, the RNA portion of the composite primer comprises the following ribonucleotide sequence: 5'- GACGGAUGCGGUCU-3'.

[0217] In still other embodiments, the invention provides a composition comprising: (a) UNG; (b) ARP; (c) dUTP; (d) a DNA polymerase; (e) RNAse H; (f) a composite primer, wherein the composite primer comprises a 5' RNA portion and a 3' DNA portion. In other embodiments, the composition further comprises (g) $MgCl_2$ solution; (h) acetic acid solution; and optionally, (i) a stop buffer comprising 1.5M Tris, pH 8.5. In other embodiments, the invention provides a composition comprising (a) UNG; (b) an agent capable of labeling an abasic site (for example, Alexa Fluor 555 or an aminooxy-modified Alexa Fluor 555); (c) dUTP; (d) a DNA polymerase; (e) RNAse H; (f) a composite primer, wherein the composite primer comprises a 5' RNA portion and a 3' DNA portion. In some embodiments, the DNA polymerase and RNAse H are provided as a mixture. In some embodiments, the RNA portion of the composite primer is 5' with respect to the 3' DNA portion, the 5' RNA portion is adjacent to the 3' DNA portion, the RNA portion of the composite primer consists of about 10 to about 20 nucleotides and the DNA portion of the composite primer consists of about 7 to about 20 nucleotides. In still other embodiments, the composition comprises a second, different composite primer. In some embodiments, the RNA portion of the composite primer comprises the following ribonucleotide sequence: 5'- GACGGAUGCGGUCU-3'.

[0218] In another example, the invention provides compositions comprising a polynucleotide comprising an abasic site and a suitable substrate for attachment through an abasic site (e.g., a microarray; an analyte), which may be functionalized if necessary. In still another example, the invention provides a composition comprising a polynucleotide comprising a non-canonical nucleotide, UNG, (optionally) E. coli Endonuclease IV, and a suitable substrate for attachment through an abasic site, which may be functionalized if necessary.

[0219] The compositions are generally in lyophilized or aqueous form (if appropriate), preferably in a suitable buffer.

[0220] The invention also provides compositions comprising the labeled and/or fragmented products described herein. Accordingly, the invention provides a population of

IDT Exhibit 1020.8024

labeled and/or fragmented polynucleotides, which are produced by any of the methods described herein (or compositions comprising the products).

[0221] The invention also provides compositions comprising the immobilized polynucleotides or immobilized polynucleotide fragments described herein. In some embodiments, the immobilized polynucleotide (or immobilized fragment, in embodiments involving fragmentation) are labeled, as described herein. Accordingly, the invention provides a population of immobilized polynucleotides or immobilized polynucleotide fragments which are produced by any of the methods described herein (or compositions comprising the products).

[0222] 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).

[0223] 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 composite primer, said composite primer comprising an RNA portion and a 3' DNA portion; and a non-canonical nucleotide (such as dUTP). In another embodiment, the reaction mixture comprises a polynucleotide comprising an abasic site, wherein the polynucleotide was synthesized using a composite primer; and an agent (such as UNG) that is capable of cleaving a base portion from a non-canonical nucleotide. In another embodiment, the reaction mixture comprises a polynucleotide comprising an abasic site, wherein the polynucleotide was synthesized using a composite primer; and an agent (such as an amine, such as N, N'-dimethylethylenediamine) capable of cleaving the phosphodiester back at an abasic site. In other embodiments, the reaction mixture comprises a polynucleotide comprising an abasic site, wherein the polynucleotide was synthesized using a composite primer; and an agent that labels an abasic site (such as ARP). In other embodiments, the reaction mixture comprises a composite primer, said composite primer comprising an RNA portion and a 3' DNA portion; dUTP; and UNG. In still other embodiments, the reaction mixture comprises a polynucleotide comprising an abasic site, wherein the polynucleotide was synthesized using a composite primer; and ARP. In still other embodiments, the reaction mixture comprises a polynucleotide comprising an abasic site, wherein the polynucleotide was synthesized using a composite primer; and N, N'-dimethylethylenediamine. In still other embodiments, the reaction mixture comprises a polynucleotide comprising an abasic site, wherein the polynucleotide was synthesized using a composite primer; N, N'-dimethylethylenediamine; and ARP. In still another

IDT Exhibit 1020.8025

embodiment, invention provides a reaction mixture comprising (a) UNG; (b) N, N'-dimethylethylenediamine; and (c) ARP. In other embodiments, the invention provides a reaction mixture comprising (a) UNG; and (b) N, N'-dimethylethylenediamine. In still other embodiments, the invention provides a reaction mixture comprising: (a) dUTP; (b) a DNA polymerase; (c) RNAse H; and (d) a composite primer, wherein the composite primer comprises a 5' RNA portion and a 3' DNA portion. In still other embodiments, the invention provides a reaction mixture comprising (a) a composite primer, wherein the composite primer comprises an RNA portion and a 3' DNA portion; (b) dUTP; (c) a mixture of dATP, dTTP, dCTP, and dGTP; (d) a DNA polymerase; and (e) RNAse H. In still other embodiments, the invention provides a reaction mixture comprising a composite primer, said composite primer comprising an RNA portion and a 3' DNA portion; and a non-canonical nucleotide. In some embodiment, the reaction mixture further provides a suitable substrate for immobilization. In some embodiments, the 5' RNA portion of the composite primer is adjacent to the 3' DNA portion, the RNA portion of the composite primer consists of about 10 to about 20 nucleotides and the DNA portion of the composite primer consists of about 7 to about 20 nucleotides. In still other embodiments, the reaction mixture comprises a second, different composite primer. In some embodiments, the RNA portion of the composite primer comprises the following ribonucleotide sequence: 5'-GACGGAUGCGGUCU-3'.

[0224] In other embodiments, the reaction mixture comprises a polynucleotide comprising an abasic site, wherein the polynucleotide was synthesized using a composite primer, said composite primer comprising an RNA portion and a 3' DNA portion; and a substrate suitable for immobilization.

[0225] The invention provides kits for carrying out the methods of the invention. Accordingly, a variety of kits are provided in suitable packaging. The kits may be used for any one or more of the uses described herein, and, accordingly, may contain instructions for any one or more of the following uses: methods of producing a hybridization probe, characterizing and/or quantitating nucleic acid, detecting a mutation, preparing a subtractive hybridization probe, detection (using a hybridization probe), and determining a gene expression profile, using the labeled and/or fragmented nucleic acids generated by the methods of the invention.

[0226] 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: a primer (such as a RNA-DNA composite primer), a non-canonical nucleotide, an agent (such as an enzyme) capable of cleaving a base portion of a non-canonical nucleotide, an agent (such as an enzyme) capable of effecting cleavage of a

IDT Exhibit 1020.8026

phosphodiester backbone at an abasic site, and an agent capable of labeling an abasic site, which may or may not be separately packaged. In still another example, the kit comprises a primer (such as a composite primer as described herein), dUTP, UNG, (optionally) E. coli Endonuclease IV, and N-(aminooxyacetyl)-N'-(D-biotinoyl) hydrazine, trifluorecetic acid salt (ARP). In another embodiment, the kit comprises a primer (such as a composite primer), a non-canonical nucleotide, an agent (such as an enzyme) capable of cleaving a base portion of a non-canonical nucleotide, and an agent (such as an enzyme) capable of effecting cleavage of a phosphodiester backbone at an abasic site. In another embodiment, the kit comprises a polynucleotide comprising an abasic site, wherein the polynucleotide was generated by synthesis using a template, and an agent capable of labeling an abasic site. In still another example, the kit comprises a polynucleotide comprising a non-canonical nucleotide, UNG, (optionally) E. coli Endonuclease IV, and a suitable substrate for attachment through an abasic site (e.g. a microarray,; an analyte), which may be functionalized if necessary. In another embodiment, the kit comprises a polynucleotide comprising an abasic site and a suitable substrate (which may be functionalized if necessary) for attachment to an abasic site.

[0227] In other embodiments, the invention provides a kit comprising a primer (which can be an RNA-DNA composite primer), non-canonical nucleotides, an agent (such as an enzyme) capable of cleaving a base portion of a non-canonical nucleotide, optionally an agent (such as an enzyme) capable of effecting cleavage of a phosphodiester backbone at an abasic site, and an agent capable of labeling an abasic site. In another example, the invention provides a kit comprising a polynucleotide comprising a non-canonical nucleotide, said polynucleotide synthesized from a template, and an agent capable of labeling an abasic site. In still another example, the composition comprises a primer (which can be a RNA-DNA composite primer), dUTP, UNG, (optionally) E. coli Endonuclease IV, and N-(aminooxyacetyl)-N'-(D-biotinoyl) hydrazine, trifluorecetic acid salt (ARP).

[0228] In another embodiment, the invention provides a kit comprising a composite primer, said composite primer comprising an RNA portion and a 3' DNA portion; and a non-canonical nucleotide (such as dUTP). In another embodiment, the composition comprises a composite primer, said composite primer comprising an RNA portion and a 3' DNA portion; and an agent (such as UNG) that is capable of cleaving a base portion from a non-canonical nucleotide. In another embodiment, the kit comprises a composite primer, said composite primer comprising an RNA portion and a 3' DNA portion; and an agent (such as an amine, such as N, N'-dimethylethylenediamine) capable of cleaving the phosphodiester back at an abasic site. In other embodiments, the kit

IDT Exhibit 1020.8027

comprises a composite primer, said composite primer comprising an RNA portion and a 3' DNA portion; and an agent that labels an abasic site (such as ARP). In other embodiments, the kit comprises a composite primer, said composite primer comprising an RNA portion and a 3' DNA portion; dUTP; and UNG. In still other embodiments, the kit comprises a composite primer, said composite primer comprising an RNA portion and a 3' DNA portion; dUTP; UNG; and ARP. In still other embodiments, the kit comprises a composite primer, said composite primer comprising an RNA portion and a 3' DNA portion; dUTP; UNG; and N, N'-dimethylethylenediamine. In still other embodiments, the kit comprises a composite primer, said composite primer comprising an RNA portion and a 3' DNA portion; dUTP; UNG; N, N'-dimethylethylenediamine; and ARP.

[0229] In still other embodiments, the kit comprises an agent capable of cleaving RNA from a RNA-DNA hybrid (such as RNAse H); a non-canonical nucleotide (dUTP); and an agent capable of cleaving a base portion of a non-canonical nucleotide (UNG). In still other embodiments, the kit comprises an agent capable of cleaving RNA from a RNA-DNA hybrid (such as RNAse H); and an agent capable of labeling an abasic site (such as ARP, Alexa Fluor 555 hydrazide, or FARP). In still other embodiments, the kit comprises an agent capable of cleaving RNA from a RNA-DNA hybrid (such as RNAse H); and an agent capable of cleaving the backbone at an abasic site (such as an amine, such as N, N'-dimethylethylenediamine). In still other embodiments, the kit comprises RNAse H; N, N'-dimethylethylenediamine; and ARP.

[0230] In still other embodiments, the invention provides a kit comprising: a composite primer, said composite primer comprising an RNA portion and a 3' DNA portion; a non-canonical nucleotide; an agent (such as an enzyme) capable of cleaving a base portion of a non-canonical nucleotide; an agent (such as an enzyme) capable of cleaving a phosphodiester backbone at an abasic site; and an agent capable of labeling an abasic site. In some embodiment, the kit further provides a suitable substrate for immobilization. In some embodiments, the 5' RNA portion of the composite primer is adjacent to the 3' DNA portion, the RNA portion of the composite primer consists of about 10 to about 20 nucleotides and the DNA portion of the composite primer consists of about 7 to about 20 nucleotides. In still other embodiments, the kit comprises a second, different composite primer. In some embodiments, the RNA portion of the composite primer comprises the following ribonucleotide sequence: 5'- GACGGAUGCGGUCU-3'.

[0231] In still another embodiment, invention provides a kit comprising (a) UNG; (b) N, N'-dimethylethylenediamine; and (c) ARP. In other embodiments, the invention provides a kit comprising (a) UNG; (b) N, N'-dimethylethylenediamine; (c) ARP; (d) dUTP; (e) a mixture of dATP, dTTP, dCTP, and dGTP; (f) a DNA polymerase; (g) a

IDT Exhibit 1020.8028

composite primer, wherein the composite primer comprises a 5' RNA portion and a 3' DNA portion. In still other embodiments, the invention provides a kit comprising (a) UNG; (b) N, N'-dimethylethylenediamine; (c) ARP; (d) dUTP; (e) a mixture of dATP, dTTP, dCTP, and dGTP; (f) a DNA polymerase; (g) RNAse H; (h) a composite primer, wherein the composite primer comprises a 5' RNA portion and a 3' DNA portion. In yet another embodiment, the invention provides a kit comprising (a) UNG; (b) N, N'-dimethylethylenediamine; (c) ARP; (d) dUTP; (e) a mixture of dATP, dTTP, dCTP, and dGTP; (f) a DNA polymerase; (g) RNAse H; (h) a composite primer, wherein the composite primer comprises a 5' RNA portion and a 3' DNA portion (i) $MgCl_2$ solution; (j) acetic acid solution; and optionally, (k) a stop buffer comprising 1.5M Tris, pH 8.5. In some embodiments, the dUTP and the mixture of dATP, dTTP, dCTP, and cGTP are combined. In some embodiments, the DNA polymerase and RNAse H are provided as a mixture. In some embodiments, the RNA portion of the composite primer is 5' with respect to the 3' DNA portion, the 5' RNA portion is adjacent to the 3' DNA portion, the RNA portion of the composite primer consists of about 10 to about 20 nucleotides and the DNA portion of the composite primer consists of about 7 to about 20 nucleotides. In still other embodiments, the kit comprises a second, different composite primer. In some embodiments, the RNA portion of the composite primer comprises the following ribonucleotide sequence: 5'- GACGGAUGCGGUCU-3'.

[0232] In still other embodiments, the invention provides a kit comprising: (a) UNG; (b) ARP; (c) dUTP; (d) a DNA polymerase; (e) RNAse H; (f) a composite primer, wherein the composite primer comprises a 5' RNA portion and a 3' DNA portion. In other embodiments, the kit further comprises (g) $MgCl_2$ solution; (h) acetic acid solution; and optionally, (i) a stop buffer comprising 1.5M Tris, pH 8.5. In other embodiments, the invention provides a kit comprising (a) UNG; (b) an agent capable of labeling an abasic site (for example, Alexa Fluor 555 or an aminooxy-modified Alexa Fluor 555); (c) dUTP; (d) a DNA polymerase; (e) RNAse H; (f) a composite primer, wherein the composite primer comprises a 5' RNA portion and a 3' DNA portion. In some embodiments, the DNA polymerase and RNAse H are provided as a mixture. In some embodiments, the RNA portion of the composite primer is 5' with respect to the 3' DNA portion, the 5' RNA portion is adjacent to the 3' DNA portion, the RNA portion of the composite primer consists of about 10 to about 20 nucleotides and the DNA portion of the composite primer consists of about 7 to about 20 nucleotides. In still other embodiments, the kit comprises a second, different composite primer. In some embodiments, the RNA portion of the composite primer comprises the following ribonucleotide sequence: 5'-GACGGAUGCGGUCU-3'.

IDT Exhibit 1020.8029

[0233] Kits may also include one or more suitable buffers (as described herein). 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.

[0234] 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 methods of the invention, and/or, as appropriate, for using the products for purposes such as, for example preparing a hybridization probe, expression profiling, preparing a microarray, or characterizing a nucleic acid. 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.

[0235] 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.

[0236] 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.

Applications using the labeling and/or fragmentation and/or immobilization methods of the invention

[0237] The methods and compositions of the invention can be used for a variety of purposes. For purposes of illustration, methods of producing a hybridization probe, characterizing and/or quantitating nucleic acid, detecting a mutation, preparing a subtractive hybridization probe, detection (using the hybridization probe), and determining a gene expression profile, using the labeled and/or fragmented nucleic acids generated by the methods of the invention, are described.

[0238] Immobilized polynucleotides, for example on a microarray, prepared according to any of the methods of the invention, are also useful for methods of analyzing and characterizing nucleic acids, including methods of hybridizing nucleic acids, methods

IDT Exhibit 1020.8030

of characterizing and/or quantitating nucleic acids, methods of detecting a mutation in a nucleic acids, and methods of determining a gene expression profile, as described below, and these applications likewise apply to immobilized polynucleotides.

Method of producing a hybridization probe

[0239] The labeled polynucleotides obtained by the methods of the invention are useful as a hybridization probe. Accordingly, in one aspect, the invention provides methods for generating hybridization probes, comprising generating labeled polynucleotides using any of the methods described herein, and using the labeled polynucleotides as a hybridization probe. In another embodiment, the invention provides methods for generating a hybridization probe, comprising generating labeled polynucleotide fragments using any of the methods described herein, and using the labeled polynucleotide fragments as a hybridization probe. The labeled polynucleotide (or polynucleotide fragments) can be produced from any template known in the art, including RNA, DNA, genomic DNA (including global genomic DNA amplification), and libraries (including cDNA, genomic or subtractive hybridization library). The invention also provides methods of hybridizing using the hybridization probes described herein.

Characterization of nucleic acids

[0240] The labeled and/or fragmented nucleic acids obtained by the methods of the invention are amenable to further characterization.

[0241] The labeled and/or fragmented nucleic acids (i.e., products of any of the 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.

[0242] In one embodiment, the methods of the invention are utilized to generate labeled and/or fragmented nucleic acids, and analyze the labeled and/or fragmented nucleic acids by contact with a probe. The labeled and/or fragmented nucleic acid can be produced from any template known in the art, including RNA, DNA, genomic DNA (including global genomic DNA amplification), and libraries (including cDNA, genomic or subtractive hybridization library).

[0243] In one embodiment, the methods of the invention are utilized to generate labeled and/or fragmented nucleic acids which are analyzed (for example, detection and/or quantification) by contacting them with, for example, microarrays (of any suitable substrate, 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 1020.8031

methods to characterize (for example, detect and/or quantify and/or identify) a labeled and/or fragmented nucleic acid by analyzing the labeled products, for example, by hybridization of the labeled products to, for example, probes immobilized at, for example, specific locations on a solid or semi-solid substrate, probes immobilized on defined particles (including beads, such as Bead Array, Illumina), or probes immobilized on blots (such as a membrane), for example arrays, or arrays of arrays. Immobilized probes include immobilized probes generated by the methods described herein, and also include at least the following: cDNA and synthetic oligonucleotides, which can be synthesized directly on the substrate.

[0244] 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 products and probes from solution. The identity of the probes provides characterization of the sequence identity of the products, and thus by extrapolation can also provide characterization of the identity of a template from which the products were prepared (for example, the identity of an RNA in a solution). For example, hybridization of the labeled products is detectable, and the amount of specific labels that are detected is proportional to the amount of the labeled products prepared from 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 template RNA species in the sample.

[0245] Methods of characterization include sequencing by hybridization (see, e.g., Dramanac, U.S. Patent No. 6,270,961) and global genomic hybridization (also termed comparative genome hybridization) (*see, e.g.*, Pinkel, U.S. Patent No. 6,159,685).

[0246] In another aspect, the invention provides a method of quantitating labeled and/or fragmented nucleic acids comprising use of an oligonucleotide (probe) of defined sequence (which may be immobilized, for example, on a microarray).

Mutation detection utilizing the methods of the invention

[0247] The labeled and/or fragmented nucleic acids generated according to the methods of the invention are also suitable for analysis for the detection of any alteration in the template nucleic acid sequence (from which the labeled and/or fragmented nucleic acids are synthesized), as compared to a reference nucleic acid sequence which is identical to the template nucleic acid sequence other than the sequence alteration. The sequence

IDT Exhibit 1020.8032

alterations may be sequence alterations present in the genomic sequence or may be sequence alterations which are not reflected in the genomic DNA sequences, for example, alterations due to post transcriptional alterations, and/or mRNA processing, including splice variants. Sequence alterations (interchangeably called "mutations") include deletion, substitution, insertion and/or transversion of one or more nucleotide.

[0248] Accordingly, the invention provides methods of detecting presence or absence of a mutation in a template, comprising: (a) generating a labeled polynucleotide, or fragments thereof, by any of the methods described herein; and (b) analyzing the labeled polynucleotide, or fragments thereof, whereby presence or absence of a mutation is detected. In some embodiments, the labeled polynucleotide, or fragments thereof, is compared to a labeled reference template, or fragments thereof. Step (b) of analyzing the labeled polynucleotide, or fragments thereof, whereby presence or absence of a mutation is detected, can be performed by any method known in the art. In some embodiments, probes for detecting mutations are provided as a microarray.

[0249] 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.

[0250] Other art recognized methods of analysis for the detection of any alteration in the template nucleic acid sequence, as compared to a reference nucleic acid sequence, are suitable for use in the methods of the present invention. For example, essentially any hybridization-based method of detection of mutations is suitable for use with the labeled and/or fragmented nucleic acids produced by the methods of the invention.

Determination of gene expression profile

[0251] The labeled and/or fragmented nucleic acids produced by the methods of the invention are particularly suitable for use in determining the levels of expression of one or more genes in a sample. As described above, labeled and/or fragmented nucleic acids 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 products (for example amplification products) of the sequences, provides for determination of the gene expression profile of the sample source.

IDT Exhibit 1020.8033

[0252] Accordingly, the invention provides methods of determining gene expression profile in a sample, said method comprising: amplifying single stranded (or double stranded) product from at least one RNA sequence of interest in the sample, using any of the methods described herein, wherein non-canonical nucleotides are incorporated during synthesis of a polynucleotide; labeling and/or fragmenting the polynucleotide comprising the non-canonical nucleotide; and determining amount of labeled and/or fragmented nucleic acids produced from 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.

[0253] Accordingly, the invention provides of determining gene expression profile in a sample, said method comprising: (a) generating labeled polynucleotide or fragments thereof from at least one polynucleotide template in the sample using any of the methods described herein; and (b) determining amount of labeled polynucleotide or fragments thereof of each polynucleotide template, wherein each said amount is indicative of amount of each polynucleotide template in the sample, whereby the gene expression profile in the sample is determined.

[0254] It is understood that amount of labeled and/or fragmented nucleic acids produced (and thus the amount of product) may be determined using quantitative and/or qualitative methods. Determining amount of labeled and/or fragmented nucleic acids includes determining whether labeled and/or fragmented nucleic acids are present or absent. Thus, an expression profile can include 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.

[0255] The methods of expression profiling are useful in a wide variety of molecular diagnostics, and especially in the study of gene expression in essentially any 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.

Methods of preparing a subtractive hybridization probe

[0256] The labeled and/or fragmented nucleic acids methods of the invention are particularly suitable for use in preparation of labeled and/or fragmented subtractive hybridization probes. For example, two nucleic acid populations, one sense and one antisense, can be allowed to mix together with one population present in molar excess

IDT Exhibit 1020.8034

("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. Labeled and/or fragmented subtractive hybridization probe is then labeled and/or fragmented according to the methods of the invention described herein.

Comparative hybridization

[0257] In another aspect, the invention provides methods for comparative hybridization (such as comparative genomic hybridization), said method comprising: (a) preparing a first population of labeled polynucleotides or fragments thereof from a first template polynucleotide sample using any of the methods described herein; (b) comparing hybridization of the first population to at least one probe with hybridization of a second population of labeled polynucleotides or fragments thereof. In some embodiments, the at least one probe is a chromosomal spread. In still other embodiments, the at least one probe is provided as a microarray. In some embodiments, the first and second population comprise detectably different labels. In other embodiments, the second population of labeled polynucleotides, or fragments thereof, are prepared from a second polynucleotide sample using any of the methods described herein. The method according to claim 57, wherein the first population and second population comprise detectably different labels. In some embodiments, step (b) of comparing comprises determining amount of said products, whereby the amount of the first and second polynucleotide templates is quantified.

[0258] In some embodiments, comparative hybridization comprises preparing a first population of labeled polynucleotides (which can be polynucleotide fragments) according to any of the methods described herein, wherein the template from which the first population is synthesized is genomic DNA. A second population of labeled polynucleotides (to which the first population is desired to be compared) is prepared from a second genomic DNA template. The first and second populations are labeled with different labels. The hybridized first and second populations are mixed, and hybridized to an array or chromosomal spread. The different labels are detected and compared.

[0259] The following Examples are provided to illustrate, but not limit, the invention.

IDT Exhibit 1020.8035

EXAMPLES

Example 1: Demonstration of Fragmentation and Labeling with Biotin of an Abasic Site of a Synthetic Oligonucleotide.

[0260] A synthetic 75mer oligodeoxynucleotide with a single deoxyuridine incorporated at the 49th position from the 5' end (Sequence 1) was obtained from Operon (Alameda, CA) and dissolved in TE buffer (10 mM Tris/1 mM EDTA, pH 8.0) at a concentration of 0.4 mg/mL.

Sequence 1:

5'-GGA CCA CCG TTC CGC CGA CCA GAC TCT GCA TAT CTT CCG CCA TCC CGG UGA CCA TAC CGT AAA AAA AAA AAA AAA-3' (SEQ ID NO:1).

[0261] Uracil was removed (creating an abasic site) by mixing 5 μL of the oligonucleotide stock with 35 μL of Isotherm® buffer (Epicentre, Madison, WI) and 2 Units of UNG (Epicentre, Madison, WI), and incubating the mixture at 37°C for 60 minutes in a thin-walled polypropylene tube in a thermal cycler. Next, the oligonucleotide comprising an abasic site was fragmented (cleaved at the phosphodiester backbone at the abasic site) at the abasic site by incubating the mixture at 99°C for 30 minutes. The cleaved oligonucleotide product was purified with a QIAquick Nucleotide Removal Kit (Qiagen, Valencia, CA) following the manufacturer's instructions, and recovered in approximately 35 μL of water. The fragmented product was labeled by adding 4 μL of 100 mM acetic acid/tetramethylethylenediamine buffer, pH 3.8 (the buffer was prepared by preparing 100 mM acetic acid, and adjusting the pH to 3.8 with TEMED), and 4 μL of ARP (N-(aminooxyacetyl)-N'-(D-biotinoyl) hydrazine, trifluoroacetic acid salt), 22.5 mM in water (Molecular Probes, Eugene, OR), and incubating for 60 minutes at 37°C. The labeling reaction was terminated by adding 5 μL of 1 M Tris buffer, pH 8.5, and the product was again purified as above and recovered in approximately 35 μL water. Appropriate controls were included which omitted either the UNG (data not shown) or the labeling reagent (ARP).

[0262] Incorporation of biotin in the product (via the labeling of abasic site with ARP) was detected by mixing 5 μL of product with 3 μL of a 2.5 mg/mL aqueous solution of streptavidin (Sigma, St. Louis. MO) before electrophoresis. The reaction products were analyzed on a PAGE gel (4-20%; InVitrogen, San Diego, CA). DNA was visualized using ethidium bromide.

IDT Exhibit 1020.8036

[0263] The results of this experiment are shown in the gel photograph of Figure 4. Lane 1 shows 50 and 100 bp double stranded DNA marker, lane 2 (labeled "no label") shows the no-label control, lane 3 (labeled "L3.8+Strep") shows the fragmented and labeled oligonucleotide treated with streptavidin, and lane 4 (labeled "L3.8") shows the fragmented and labeled oligonucleotide. Note that the single stranded oligonucleotide runs more slowly than the double stranded marker.

[0264] Excision of uracil and cleavage of the oligonucleotide were found to be nearly complete in the No Label control (lane 1), as evidenced by appearance of a strong band at ca. 50 nucleotide length and near disappearance of the starting material band at ca. 75 nucleotides. Reaction product treated with label (shown in lane 4, labeled L3.8) was similar in appearance, but the product additionally treated with streptavidin (shown in Lane "L3.8 + Strep) was strongly retarded, appearing as fuzzy bands with apparent lengths of several hundred nucleotides. Only a fraction of fragmented product did not react with streptavidin. It was concluded that fragment was nearly completely labeled with ARP.

Example 2: Labeling of an abasic site in a Synthetic Oligonucleotide with Biotin without Fragmentation.

[0265] The experiment in Example 1 was repeated, except that the 99°C fragmentation step was omitted and the starting oligodeoxynucleotide was Sequence 2. An additional reaction was performed in which the labeling reaction was performed as described in Example 1, except that the buffer was 100 mM acetic acid/tetramethylethylenediamine buffer, pH 6 (the buffer was prepared by preparing 100 mM acetic acid, and adjusting the pH to 6 with TEMED).

Sequence 2:
5'-GGA CCA CCG TTC CGC CGA CCA GAC UCT GCA TAT CTT CCG CCA TCC CGG TGA CCA TAC CGT AAA AAA AAA AAA AAA-3' (SEQ ID NO:2).

[0266] The results of this experiment are shown in Figure 5. Lane 1 shows molecular weight marker (as described in figure 1). Note that the single stranded oligonucleotide runs more slowly than the double stranded marker. "NL" refers to the no-label control, "L6" refers to reactions in which labeling was performed at pH 6, and "L3.8" refers to reactions in which labeling was performed at pH 3.8. Lanes marked "-" show

IDT Exhibit 1020.8037

reaction samples that were not treated with streptavidin, and lanes marked "+" show reaction samples that were treated with streptavidin. The lower arrow marks the expected molecular weight of oligonucleotide not retarded by streptavidin, and the upper arrow marks the expected position of gel retarded product treated with streptavidin.

[0267] As shown in lanes "L6" and "L3.8", nearly all of the product reacted with label could be retarded by streptavidin treatment. Only a fraction of labeled product did not react with streptavidin. It was concluded that fragment was nearly completely labeled with ARP.

[0268] By contrast, product not reacted with label ("NL", or the no label control) was not capable of being retarded by streptavidin treatment.

Example 3: Demonstration of Fragmentation and Labeling with Biotin of Ribo-SPIA™ Product.

[0269] A mixture of DNA products incorporating deoxyuridine was prepared using Ribo-SPIA™ amplification using commercial total RNA preparation from breast cancer tumor (CLONTECH; cat. no.: 64015-1) as follows:

Primer sequences:

MTB4 : 5'- *GAC GGA UGC GGU CUC CAG UGU* dTdTdT dTdTdT dTdTdT dTdTdT dTdNdN-3' (SEQ ID NO:3)

where dN denotes a degenerate nucleotide (i.e., it can be dA, dT, dC, and dG), and italicized and underlined letters denote ribonucleotides.

MTA4: 5'-*GAC GGA UGC GGU CUC C*dAdG dTdGdT dTdT-3' (SEQ ID NO:4)

where italicized and underlined letters denote ribonucleotides.

[0270] 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)

MTB4 primer @1μM

25 mM dNTPs

0.2 μl RNasin Ribonuclease Inhibitor (Promega N2511, 40u/μl)

1 μl 0.1 M DTT

20 ng of total RNA per reaction

DEPC- treated water to a total volume of 19 μl

IDT Exhibit 1020.8038

[0271] The reaction mixtures were pre-incubated at 75°C for 2 minutes, and then cooled down to 42°C. 1 μL Sensiscript per reaction (Qiagen, Valencia, CA, Cat No. 205211) was added to each reaction, and the reactions were incubated at 42°C for 50 minutes.

[0272] 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 synthesis 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 E. coli DNA ligase (BRL 18052-019, 10U/μl)

[0273] 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.

[0274] Step 3: Amplification of total cDNA. Amplification was carried out using 1 μl of the second strand cDNA reaction mixture above, using the MTA4 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 dATP, dGTP, dCTP (25 mM)

0.2 μL of a stock containing 20 mM dTTP and 5 mM dUTP

0.2 μl of MTA4 (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

[0275] Reactions were heated to 50C, 8 Units of Bst DNA Polymerase Large Fragment (New England Biolabs, Beverly, MA), 0.02U Hybridase Thermostable Rnase H (Epicentre H39100), and 0.3 μg T4 Gene 32 Protein (USB 70029Z)were added, and the reactions were further incubated at this temperature for 60 min.

[0276] Amplified single stranded DNA product was fragmented and labeled as follows: Approximately 2 μg of product DNA in 40 μL of Isotherm® buffer (Epicentre,

IDT Exhibit 1020.8039

Madison, WI) was treated with 2 Units of UNG, and fragmented, and labeled as described in Example 1. A control was performed lacking UNG and label (ARP), and without heat treatment. A portion of the fragmented and labeled product was treated with avidin, as described in Example 1. Reaction products were analyzed as described in Example 1 and the results are shown in Figure 6.

[0277] Figure 6 shows the following:

Lane 1: DNA molecular weight marker as described in Example 1

Lane 2: amplified single stranded DNA product

Lane 3: amplified single stranded DNA product treated with UNG, labeled with biotin, and cleaved by heat treatment

Lane 4: DNA molecular weight marker

Lane 5: streptavidin-treated amplified single stranded DNA product treated with UNG, labeled with biotin, and fragmented by heat treatment

Lane 6: No streptavidin control (contains amplified single stranded DNA product treated with UNG, labeled with biotin, and fragmented by heat treatment, as shown in Lane 3)

[0278] Analysis of average size of DNA in the reaction mixtures revealed that the control product of lane 2 was an average length of ca. 400 nucleotides (with the largest products over about 1000 bases). By contrast, the UNG-treated and heat-fragmented product of lane 3 was an average length of 150 nucleotides after UNG and heat treatment, and the largest products (over ca. 1,000 bases) disappeared almost entirely.

[0279] An aliquot of the UNG-treated, heat fragmented product was treated with streptavidin, and the results are shown in the Lane 5. Lane 6 shows the no-streptavidin control. Streptavidin treatment resulted in a shift of nearly the entire product band to larger size, indicating virtually complete labeling of the single stranded DNA products treated with UNG, labeled with biotin and fragmented by heat treatment.

Example 4: Efficient labeling and fragmentation of Ribo-SPIA product using a single reaction mixture for creation of abasic sites and fragmentation at abasic sites, with no intermediate purification steps.

[0280] A mixture of DNA products incorporating deoxyuridine was prepared using total RNA preparation from mouse brain (obtained from the Gladstone Institute, San Francisco CA; used with permission) as follows:

[0281] 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 $MgCl_2$)

MTB4 primer @ 0.25µM

IDT Exhibit 1020.8040

0.2 μL 25 mM dNTPs

0.25 μl RNasin Ribonuclease Inhibitor (Promega N2511, 40u/μl)

20 ng of total RNA per reaction

DEPC- treated water to a total volume of 19 μl

[0282] The reaction mixtures were pre-incubated at 65°C for 2 minutes, and then cooled down to 42°C. 1 μL Sensiscript per reaction (Qiagen, Valencia, CA, Cat No. 205211) was added to each reaction, and the reactions were incubated at 48°C for 60 minutes, then at 70°C for 15 minutes.

[0283] Step 2: Synthesis of second strand cDNA. The entire 20 μl of the first strand cDNA synthesis reaction mixture was aliquoted to individual reaction tubes, and 20 μl of second strand synthesis stock reaction mixture was added to each tube and mixed. The second strand synthesis stock reaction mixture contained the following:

1 μl of 10X Klenow reaction buffer (10X buffer: 500 mM Tris-HCl, pH 8.0; 100 mM $MgCl_2$, 500 mM NaCl)

1mM DTT

1 U/μl of exo- Klenow DNA polymerase (USB catalog number 70057Z)

0.02 U/μl of Ribonuclease H (USB catalog number 70054Z)

0.2 μl (25 mM) dNTPs

0.4 U/μl RNasin (USB catalog number 71571)

water to a total volume of 20μl.

[0284] The reaction mixtures were incubated at 37°C for 30 minutes. The reactions were stopped by addition of 2 μl of 0.5M EDTA.

[0285] Step 3: Amplification of total cDNA: 5 μl of the second strand cDNA reaction mixture was aliquoted into 8 20 μl reaction mixtures. 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 dATP, dGTP, dCTP (25 mM)

0.2 μL of a stock containing 20 mM dTTP and 5 mM dUTP

0.2 μl of MTA4 (100 μM)

5 μl of the second strand cDNA synthesis mixture

0.1 μl Rnasin

DEPC-treated water to a total volume of 18.8 μl

[0286] Reactions were placed on ice and 8 Units of Bst DNA Polymerase Large Fragment (New England Biolabs, Beverly, MA), 0.02U Hybridase Thermostable Rnase H

IDT Exhibit 1020.8041

(Epicentre H39100), and 0.3 μg T4 Gene 32 Protein (USB 70029Z) were added. Reactions were placed at 50°C for 60 minutes, then amplification was stopped by heating at 80°C for 5 minutes.

[0287] Removal of uracil (to create abasic sites) and fragmentation of abasic sites was conducted in the same reaction mixture as follows: 76 μl of unpurified amplification reaction product was mixed with 3.2 μL of N,N'-dimethylethylenediamine buffer (Aldrich Chemical, St. Louis, MO; prepared by diluting to 0.5 M solution in water, pH adjusted to 8.5 with HCl) and 4 Units of HK-UNG (Epicentre, Madison WI). This mixture was incubated at 37°C for 60 minutes, then 6.8 μL of 1 M acetic acid in water, 2 μL of 0.2 M $MgCl_2$ in water, and 8 μL of ARP solution (N-(aminooxyacetyl)-N'-(D-biotinoyl) hydrazine, trifluoroacetic acid salt; 22.5 mM in water; obtained from Molecular Probes, Eugene, OR; catalog no. A-10550) were added. This reaction mixture was incubated for 60 minutes more at 37°, then the reaction mixture was split into two tubes and purified as described in Examples 1-3. Fragmented and labeled reaction product was recovered by elution with water. A portion of the fragmented and labeled product was treated with streptavidin essentially as described in Example 1.

[0288] Control reaction mixtures were performed which lacked HK-UNG, or in which the Ribo-SPIA™ product was first purified using a QIAquick PCR purification kit (Qiagen, Valencia, CA) following the manufacturer's instructions before fragmentation and labeling as described. A portion of the fragmented and labeled product was treated with streptavidin essentially as described in Example 1.

[0289] The following reactions were analyzed using a PAGE gel as described in Example 1 (data not shown):

Lane 1: 50 and 100 bp double stranded DNA molecular weight marker.
Lane 2: No UNG control.
Lane 3: Same as Lane 2, but reacted with streptavidin.
Lane 4: Example 4 reaction product prepared with purified Ribo-SPIA™ product.
Lane 5: Same as Lane 4, reacted with streptavidin.
Lane 6: Example 4 reaction product prepared with unpurified Ribo-SPIA™ product.
Lane 7: Same as Lane 6, reacted with streptavidin.

[0290] Analysis of average size of DNA in the reaction mixtures revealed that the no-UNG control product of lane 2 was an average length of ca. 400 nucleotides (with the largest products over about 1000 bases. An aliquot of the no-UNG control product was treated with streptavidin. Streptavidin treatment did not result in a shift of the product band to larger sizes, indicating that the single stranded DNA products were not labeled

IDT Exhibit 1020.8042

with biotin, as expected, and that nonspecific interactions between streptavidin and DNA do not cause a shift on the gel.

[0291] By contrast, the UNG-treated and dimethylethylenediamine-fragmented product of lanes 4 and 6 was an average length of about 250 nucleotides after UNG and dimethylethylenediamine treatment, and the largest products (over ca. 1,000 bases) disappeared almost entirely. No difference in product length was observed between UNG-treated and dimethylethylenediamine-fragmented product prepared using unpurified Ribo-SPIA single stranded DNA (lane 6) and UNG-treated and dimethylethylenediamine-fragmented product prepared using purified Ribo-SPIA single stranded DNA (lane 4).

[0292] Aliquots of UNG-treated and dimethylethylenediamine-fragmented product prepared using unpurified Ribo-SPIA single stranded DNA (lane 6) and UNG-treated and dimethylethylenediamine-fragmented product prepared using purified Ribo-SPIA single stranded DNA (lane 4) were treated with streptavidin, and the results are shown in Lanes 5 and 7, respectively. Streptavidin treatment resulted in a shift of nearly the entire product band to larger size, indicating virtually complete labeling of the single stranded DNA products treated with UNG, labeled with biotin and fragmented by dimethylethylenediamine treatment.

[0293] Aliquots of no-UNG treatment control product (corresponding to that of lane 2, above) and UNG-treated and dimethylethylenediamine-fragmented product prepared using purified Ribo-SPIA single stranded DNA (corresponding to that of lane 4, above) were further analyzed by gel electrophoresis using an Agilent Bioanalyzer (Agilent, Mountain View, CA). Figure 7 shows the superimposed resulting electropherograms as follows:

- The closely spaced peaks at 19 seconds ("sec") (marked with "A") are an internal marker included in all samples. The closely matching elution times serve to demonstrate that the instrument is performing reproducibly.
- The sharp peak at 22 seconds is a synthetic single-stranded 75mer oligonucleotide used as a size marker (marked with "B").
- The broader peak centered at 21 seconds is the UNG-treated and dimethylethylenediamine-fragmented product prepared using purified Ribo-SPIA single stranded DNA (marked with "C"); material from Lane 6 appeared very similar.
- The much broader peak extending to about 42 seconds is the un-fragmented control (no-UNG treatment control product) (marked with "D").

[0294] For comparison, a series of RNA markers (Ambion, Austin TX) are also shown in Figure 7. Marker sizes are 0.2, 0.5, 1, 2, 4, and 6 kb (running at about 21, 23.5, 27, 30, 34, and 39 seconds, respectively).

IDT Exhibit 1020.8043

[0295] Both the conventional gel and the Bioanalyzer results establish that the UNG-treated Ribo-SPIA™ products are fragmented compared to a no-UNG-treated control. The difference is much more dramatic in the Bioanalyzer traces because the conventional gel was stained with ethidium bromide, which does not stain small single-stranded DNA well compared to the stain used in the Bioanalyzer.

Example 5: Labeling of Ribo-SPIA™ product with an aminooxy-derivatized dye.

[0296] The hydrazide of Alexa Fluor 555 ("AF555" or "Alexa Fluor hydrazide") (Molecular Probes, Eugene, OR) was converted to the aminooxy derivative Alexa Fluor 555-$NHNHCOCH_2ONH_2$ ("AF555-aminooxy") using the synthesis protocol disclosed in Ide et al, Biochemistry 32: 8276-83 (1993) (shown as the conversion of compound 2 to compound 5) The starting material shown in Ide as compound 2, BOC-aminooxy)acetic acid is available from Aldrich. The final product was purified using HPLC, and the identity of the product was verified by HPLC and mass spectrometry, which showed a mass that was 73 mass units higher than the starting material.

[0297] The aminooxy derivatized dye was dissolved in water to give a 2.1 mM solution. An aliquot was diluted at 1:1000 in water and analyzed on a Beckman DU520 spectrophotometer. The aminooxy derivatized dye retained a UV spectrum identical to unmodified Alexa Fluor 555 (data not shown).

[0298] Single stranded amplified DNA product containing dUTP was prepared from Universal Human Reference RNA (Strategene, catalog number A740000) (reaction "U") or Human Universal Reference Total RNA (Clontech catalog number 64115-1) (reaction "C"), essentially as described in Example 4. Single stranded DNA product (termed "Ribo-SPIA™" product) was then purified using a QIAquick column as described in Example 4.

[0299] Purified Ribo-SPIA™ product was labeled as follows: Approximately 10 μg of Ribo-SPIA™ DNA product from reactions U or C was concentrated to 80 μL in water using a SpeedVac. HK-UNG (10 Units; Epicentre; Madison WI) and 8 μL of 10x Isotherm® buffer (Epicentre; Madison WI) were added to each product and the mixtures were incubated at 37°C for 60 minutes. Each reaction mixture was then split into two 0.2 mL tubes. One tube of each sample received 1.7 μL of 1 M acetic acid and 3 μL of Alexa Fluor 555 hydrazide in water (7.1 mM or 21.3 nmol total); the other tube received 1.7 μL of 1 M acetic acid plus 10.2 μL of the aminooxy derivative of Alexa Fluor 555 in water (2.1 mM or 21.4 nmol total). After a further incubation at 37°C for 60 minutes and storage at -20°C overnight, 5 μL of 1 M Tris pH 8.5 was added to each tube. All products were

IDT Exhibit 1020.8044

purified using QIAquick PCR columns as described above, and each product was eluted into 60 μL of water.

[0300] Incorporation of dye into fragmented Ribo-SPIA product was analyzed by comparison of dye absorbance at 551 nm to DNA absorbance at 260 nm, using an extinction coefficient of 150,000 for dye and assuming 1 OD of DNA = 33 μg/mL. The results of this analysis were expressed as pmol dye/ug DNA and are shown in the column titled "Dye" in Table 1.

Table 1:

Sample	Dye (pmol dye/μg DNA)	Fluorescence
C + AF555 hydrazide	12.7	0.098×10^6
C + AF555 aminooxy	40.6	2.9×10^6
U + AF555 hydrazide	9.2	0.094×10^6
U + AF555 aminooxy	37.8	2.8×10^6

[0301] The fluorescence intensity of incorporated dye in fragmented Ribo-SPIA product was further analyzed as follows. 0.5 ug of sample in 15 ul of water was diluted in 4 volumes of GeneTac hybridization buffer (Genomic Solutions, Ann Arbor MI), re-purified using QIAquick columns as described above, and purified product was reduced to 8 ul under vacuum. A 2ul aliquot was diluted with 160 ul of water. Fluorescence of duplicate 80 ul aliquots was measured using a Wallac Victor2 fluorometer (Ex = 544 nm; Em = 595 nm), and the results were averaged. The results of this analysis are shown in the column titled "Fluorescence" in Table 1.

[0302] Both dye absorbance and fluorometry analysis reveal that dye was incorporated into Ribo-SPIA product. These results demonstrate that single stranded DNA products containing abasic sites prepared using UNG treatment can be labeled using commercially available dye-containing hydrazide reagents, such as Alexa Fluor 555 hydrazide.

[0303] 3.2-4.1-fold more dye was incorporated when the aminooxy-derivatized Alexa 555 was used, compared with dye incorporated using the unmodified Alexa-555-hydrazide dye. These results demonstrate that labeling is more efficient when the Alexa 555 dye is converted into an aminooxy derivative.

[0304] About 30 fold more fluorescence was detected from the Alexa-555-aminooxy (derivatized)-labeled samples compared to the Alexa-555 hydrazide

IDT Exhibit 1020.8045

(unmodified)-labeled samples. Thus, the aminooxy derivative of Alexa 555 dye shows greater brightness.

[0305] We note that these samples were subjected to additional purification. The higher fluorescence intensity ratio may indicate that some of the dye in the Alexa 555-hydrazide-labeled samples was not attached covalently and was removed by the additional purification step prior to the fluorescence analysis. Alternatively, or in addition, the fluorescence from the dye moiety in the aminooxy derivatives may be less quenched by interaction with DNA because of the longer linker present in the Alexa 555-aminooxy derivative.

Example 6: Detection of Hybridized Fragmented and Labeled polynucleotides on a Microarray.

[0306] Total mRNAs were amplified from total RNA from rat brain and rat kidney (Ambion, Austin, TX, Cat. Nos 7912 and 7926), fragmented, and labeled with biotin as described in the Example 4 control reaction in which the Ribo-SPIA™ product was purified before fragmentation and labeling. Fragmented and labeled probes were prepared for hybridization as follows: 2 μg aliquots of each fragmented and labeled product in 65 μL of water were mixed with 65 μL of formamide, denatured by heating for 2 minutes at 99°C in a 0.2 mL thin-wall PCR tube, then chilled on ice. An equal amount of 2x GeneTAC buffer (Genomic Solutions, Inc., Ann Arbor, MI) was added and the mixtures were applied to CodeLink microarrays (Uniset Rat 1, Part # 300012-03, Motorola Life Sciences, Inc., Northbrook IL.) and allowed to hybridize following manufacturer's instructions. Post-hybridization processing utilized two 30 minute incubations at 46°C rather than one incubation for one hour, but otherwise also followed manufacturer's instructions. Detection utilized a 1:100 dilution of Streptavidin-Alexa 647 as described by the manufacturer. Slides were scanned on an Axon GenePix 6000 scanner (Axon Instruments, Inc., Union City, CA).

[0307] Different spot patterns were observed depending on the starting mRNA sample (i.e., brain vs. kidney), indicating that expression analysis using labeled and fragmented polynucleotides prepared according to the methods of the invention can detect differential expression of RNA in different tissues. The wide range of signal intensities observed indicated that binding is specific for different capture sequences immobilized on different spots, rather than nonspecific binding to DNA on surfaces.

[0308] In a further experiment, total RNA from rat kidney (Ambion, Austin, TX, Cat. No 7926) was amplified, fragmented, and labeled with biotin in duplicate as described in the Example 4 control reaction in which the SPIA™ product was purified before

IDT Exhibit 1020.8046

fragmentation and labeling. Probes were prepared for hybridization as follows:2 μg aliquots of each fragmented and labeled product in 65 μL of water were mixed with 65 μL of formamide, denatured by heating for 2 minutes at 99°C in a 0.2 mL thin-wall PCR tube, then chilled on ice. An equal amount of 2x GeneTAC buffer (Genomic Solutions, Inc., Ann Arbor, MI) was added and the mixtures were applied to CodeLink microarrays (Uniset Rat 1, Part # 300012-03, Motorola Life Sciences, Inc., Northbrook IL.) and allowed to hybridize following manufacturer's instructions. Post-hybridization processing utilized two 30 minute incubations at 46°C rather than one incubation for one hour, but otherwise also followed manufacturer's instructions. Detection utilized a 1:100 dilution of Streptavidin-Alexa 647 as described by the manufacturer. Slides were scanned on an Axon GenePix 6000 scanner (Axon Instruments, Inc., Union City, CA).

[0309] Hybridization intensities were compared between the duplicate hybridizations. The correlation was calculated using the Pearson correlation coefficient calculated using the Codelink (TM) System Software available from Motorola Life Science. The correlation observed between the two independent hybridization reactions is shown in Figure 8, where the intensities observed for each spot on the arrays are plotted against each other. A useful signal range (dynamic range) of at least three orders of magnitude was obtained, demonstrating that the fragmentation and labeling reaction incorporated enough biotin label to enable detection of gene expression over a large range. The observation of signals three orders of magnitude over background demonstrated that binding to spots on the array is specific for sequences immobilized on individual spots. Good correlation between duplicate arrays (correlation coefficient $r = 0.98$) further confirmed the specificity of the entire amplification and detection process.

[0310] 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.

IDT Exhibit 1020.8047

CLAIMS

What is claimed is:

1. A method for labeling and optionally fragmenting a polynucleotide, said method comprising:

(a) synthesizing a polynucleotide from a polynucleotide template in the presence of a non-canonical nucleotide, whereby a polynucleotide comprising the non-canonical nucleotide is generated;

(b) cleaving a base portion of a non-canonical nucleotide from the synthesized polynucleotide with an enzyme capable of cleaving the base portion of the non-canonical nucleotide, whereby an abasic site is generated;

(c) optionally, cleaving a phosphodiester backbone of the polynucleotide comprising the abasic site at the abasic site; and

(d) labeling the polynucleotide or the fragment of the polynucleotide at the abasic site;

whereby a labeled polynucleotide, or optionally, a labeled polynucleotide fragment is generated.

2. A method for labeling and optionally fragmenting a polynucleotide, said method comprising:

(a) incubating a reaction mixture, said reaction mixture comprising:

(i) a polynucleotide template and

(ii) a non-canonical nucleotide; wherein the incubation is under conditions that permit synthesis of a polynucleotide comprising the non-canonical nucleotide, whereby a polynucleotide comprising the non-canonical nucleotide is generated;

(b) incubating a reaction mixture, said reaction mixture comprising:

(i) the polynucleotide comprising the non-canonical nucleotide; and

(ii) an enzyme capable of cleaving a base portion of the non-canonical nucleotide, wherein the incubation is under conditions that permit cleavage of the base portion of the non-canonical nucleotide, whereby a polynucleotide comprising an abasic site is generated;

(c) optionally incubating a reaction mixture, said reaction mixture comprising:

(i) the polynucleotide comprising the abasic site; and

(ii) an agent capable of cleaving a phosphodiester backbone of the polynucleotide comprising the abasic site at the abasic site, wherein the incubation is under

IDT Exhibit 1020.8048

conditions that permit cleavage of the phosphodiester backbone of the polynucleotide at the abasic site, whereby a fragment of the polynucleotide is generated;

(d) incubating a reaction mixture, said reaction mixture comprising:

(i) the polynucleotide comprising the abasic site or optionally, the fragment of the polynucleotide comprising the abasic site; and

(ii) an agent capable of labeling the abasic site, wherein the incubation is under conditions that permit labeling at the abasic site; whereby a labeled polynucleotide or optionally, a labeled polynucleotide fragment, is generated.

3. A method for labeling and optionally fragmenting a polynucleotide, said method comprising

(a) incubating a reaction mixture, said reaction mixture comprising:

(i) the polynucleotide comprising the non-canonical polynucleotide of step (a) of claim 1;

(ii) an enzyme capable of cleaving a base portion of the non-canonical nucleotide; and

(iii) optionally, an agent capable of cleaving a phosphodiester backbone of the polynucleotide comprising the abasic site at the abasic site, wherein the incubation is under conditions that permit cleavage of the base portion of the non-canonical nucleotide and optionally, cleavage of the phosphodiester backbone of the polynucleotide at the abasic site; whereby polynucleotide comprising the abasic site, or optionally, a fragment of the polynucleotide comprising the abasic site, is generated; and

(b) incubating a reaction mixture, said reaction mixture comprising:

(i) the polynucleotide comprising the abasic site or optionally, the fragment of the polynucleotide comprising the abasic site; and

(ii) an agent capable of labeling the abasic site, wherein the incubation is under conditions that permit labeling at the abasic site, whereby a labeled polynucleotide or optionally, a labeled fragment of the polynucleotide, is generated.

4. The method of any of the preceding claims, wherein the phosphodiester backbone of the polynucleotide comprising the abasic site is cleaved at the abasic site.

5. The method of claim any of the preceding claims, wherein the non-canonical nucleotide is selected from the group consisting of dUTP, dITP, and 5-OH-Me-dCTP.

IDT Exhibit 1020.8049

6. The method of claim any of the preceding claims, wherein the enzyme capable of cleaving a base portion of the non-canonical nucleotide is an N-glycosylase.

7. The method of claim any of the preceding claims, wherein the enzyme capable of cleaving a base portion of the non-canonical nucleotide is selected from the group consisting of Uracil N-Glycosylase (UNG), hypoxanthine-N-Glycosylase, and hydroxy-methyl cytosine-N-glycosylase.

8. The method of any of the preceding claims, wherein the non-canonical nucleotide is dUTP and the enzyme capable of cleaving a base portion of the non-canonical nucleotide is Uracil N-Glycosylase.

9. The method of any of the preceding claims, wherein the phosphodiester backbone is cleaved with an enzyme or an amine.

10. The method of any of the preceding claims, wherein the phosphodiester backbone is cleaved with N, N'-dimethylethylenediamine or AP endonuclease.

11. The method of any of the preceding claims, wherein the non-canonical nucleotide is dUTP, the enzyme capable of cleaving a base portion of the non-canonical nucleotide is Uracil N-Glycosylase, and the phosphodiester backbone is cleaved with N, N'-dimethylethylenediamine.

12. The method of any of the preceding claims, wherein the phosphodiester backbone is cleaved 3' to the abasic site.

13. The method of any of the preceding claims, wherein the phosphodiester backbone is cleaved 5' to the abasic site.

14. The method of any of the preceding claims, wherein the abasic site is labeled with N-(aminooxyacetyl)-N'-(D-biotinoyl) hydrazine, trifluoroacetic acid salt (ARP), Alexa Fluor 555, or aminooxy-derivatized Alexa Fluor 555.

15. The method of any of the preceding claims, wherein the label is capable of reacting with an aldehyde residue at the abasic site.

IDT Exhibit 1020.8050

16. The method of any of the preceding claims, wherein the non-canonical nucleotide is dUTP, the enzyme capable of cleaving a base portion of the non-canonical nucleotide is Uracil N-Glycosylase, the phosphodiester backbone is cleaved with N, N'-dimethylethylenediamine, and the abasic site is labeled with ARP.

17. The method of any of the preceding claims, wherein the polynucleotide template comprises DNA or RNA.

18. The method of any of the preceding claims, wherein the polynucleotide template is selected from the group consisting of RNA, mRNA, cDNA, and genomic DNA.

19. The method of any of the preceding claims, wherein the polynucleotide comprising a non-canonical nucleotide is single stranded.

20. The method of any of the preceding claims, wherein the polynucleotide comprising a non-canonical nucleotide is double-stranded.

21. The method of any of the preceding claims, wherein the polynucleotide comprising the non-canonical nucleotide is synthesized using a method comprising the following steps of:

(a) extending a composite primer in a complex comprising

(i) a polynucleotide template; and

(ii) the composite primer, said composite primer comprising an RNA portion and a 3' DNA portion, wherein the polynucleotide template is hybridized to the composite primer; and

(b) cleaving RNA of the annealed composite primer with an enzyme that cleaves RNA from an RNA/DNA hybrid such that another composite primer hybridizes to the template and repeats primer extension by strand displacement, whereby multiple copies of the complementary sequence of the polynucleotide template are produced.

22. The method of claim 21, wherein the complex of part (a) comprises:

(i) a complex of first and second primer extension products, wherein the first primer extension product is produced by extension of a first primer hybridized to a target RNA with at least one enzyme comprising RNA-dependent DNA polymerase activity, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA

IDT Exhibit 1020.8051

portion; wherein RNA in the complex of first and second primer extension products is cleaved with at least one enzyme that cleaves RNA from an RNA/DNA hybrid such that a composite primer hybridizes to the second primer extension product; and

(ii) the composite primer.

23. The method of any of claims 1-20, wherein the polynucleotide comprising a non-canonical nucleotide is synthesized by PCR, reverse transcription, primer extension, limited primer extension, replication, strand displacement amplification (SDA), or nick translation.

24. The method of any of the preceding claims, wherein the polynucleotide comprising a non-canonical nucleotide is synthesized using a labeled primer.

25. The method of any of the preceding claims, wherein the polynucleotide comprising a non-canonical nucleotide is synthesized using a primer comprising a non-canonical nucleotide.

26. The method of any of the preceding claims, wherein the polynucleotide comprising a non-canonical nucleotide is synthesized in the presence of two or more different non-canonical nucleotides, whereby a polynucleotide comprising two or more different non-canonical nucleotide is synthesized.

27. The method of any of the preceding claims, wherein the method comprises synthesizing a polynucleotide comprising a non-canonical nucleotide from two or more different polynucleotide templates.

28. The method of claim 1 or 2, wherein steps (a), (b) and (c) are performed simultaneously.

29. The method of claim 1 or 2, wherein steps (a), (b), (c), and (d) are performed simultaneously.

30. The method of claim 1 or 2, wherein steps (b) and (c) are performed simultaneously.

IDT Exhibit 1020.8052

31. The method of claim 1 or 2, wherein steps (b), (c), and (d) are performed simultaneously.

32. The method of claim 1 or 2, wherein steps (c) and (d) are performed simultaneously.

33. The method of claim 1 or 2, wherein step (c) is performed before step (d).

34. The method of claim 1 or 2, wherein step (d) is performed before step (c).

35. The method of any of the preceding claims, said method further comprising:

immobilizing the labeled polynucleotide comprising an abasic site, or a labeled fragment thereof, to a substrate, wherein the polynucleotide is immobilized at the abasic site.

36. A method for labeling a polynucleotide or a polynucleotide fragment, said method comprising incubating a reaction mixture, said reaction mixture comprising :

(a) the polynucleotide comprising the abasic site of step (b) of claim 1 or 2, or optionally, the fragment of the polynucleotide comprising the abasic site of step (c) of claim 1 or 2; and

(b) an agent capable of labeling the abasic site; wherein the incubation is under conditions that permit labeling at the abasic site; whereby a labeled polynucleotide or optionally, a labeled fragment of a polynucleotide, is generated.

37. The method of claim 36, wherein the polynucleotide or optionally, the polynucleotide fragment is generated according to any of claims 4-13 and 17-34.

38. A method for immobilizing a polynucleotide or a polynucleotide fragment, said method comprising immobilizing the polynucleotide comprising an abasic site of step (b) of claim 1, or optionally, the fragment of the polynucleotide comprising the abasic site of step (c) of claim 1, wherein the polynucleotide or the fragment of a polynucleotide is immobilized to a substrate at the abasic site.

39. The method of claim 38, wherein the polynucleotide or optionally, the polynucleotide fragment is generated according to any of claims 4-13 and 17-34.

IDT Exhibit 1020.8053

40. The method of claim 35 or 38, wherein the substrate is paper, glass, ceramic, plastic, polypropylene, polystyrene, nylon, polyacrylamide, nitrocellulose, silicon, or optical fiber.

41. The method of claim 35 or 38, wherein the substrate is a microarray.

42. The method of claim 41, wherein the substrate comprises pins, rods, fibers, tapes, threads, beads, particles, microtiter wells, capillaries, optical fiber, or cylinders.

43. The method of claim 35 or 38, wherein the substrate is selected from the group consisting of protein, polypeptide, peptide, nucleic acid, carbohydrate, a cell, an organic molecule, and an inorganic molecule.

44. A method of characterizing a polynucleotide of interest, comprising (a) generating labeled polynucleotide or fragments thereof using the method of any of claims 1-34 and 36; and (b) analyzing the labeled polynucleotide or fragments thereof.

45. The method of claim 44, wherein step (b) of analyzing the labeled polynucleotide or fragments thereof comprises determining amount of said polynucleotides or fragments thereof, whereby the amount of the polynucleotide template is quantified.

46. The method of claim 44, wherein step (b) comprises contacting the labeled polynucleotide or fragments thereof with at least one probe.

47. The method of claim 44, wherein the at least one probe is provided as a microarray.

48. The method of claim 47, 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.

49. The method of claim 48, 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.

IDT Exhibit 1020.8054

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

(a) generating labeled polynucleotide or fragments thereof from at least one polynucleotide template in the sample using the method of any of claims 1-34 and 36; and

(b) determining amount of labeled polynucleotide or fragments thereof of each polynucleotide template, wherein each said amount is indicative of amount of each polynucleotide template in the sample, whereby the gene expression profile in the sample is determined.

51. The method of claim 50, wherein the polynucleotide template is RNA or mRNA.

52. A method of generating hybridization probes, comprising generating labeled polynucleotide or fragments thereof using the method of any of claims according to any of claims 1-34 and 36.

53. A method of nucleic acid hybridization comprising

(a) generating labeled polynucleotide or fragments thereof using the method of any of claims according to any of claims 1-34 and 36; and

(b) hybridizing the labeled polynucleotide or fragments thereof with at least one probe.

54. A method for comparative hybridization, said method comprising:

(a) preparing a first population of labeled polynucleotides or fragments thereof from a first template polynucleotide sample using the method according to any of claims 1-34 and 36; and

(b) comparing hybridization of the first population to at least one probe with hybridization of a second population of labeled polynucleotides or fragments thereof.

55. The method according to claim 54, wherein the first population and second population comprise detectably different labels.

56. The method according to claim 54 or 55, wherein the second population of labeled polynucleotides, or fragments thereof, are prepared from a second polynucleotide sample using the method according to any of claims 1-34 and 36.

IDT Exhibit 1020.8055

57. The method of claim 54, 55, or 56, wherein step (b) of comparing comprises determining amount of said products, whereby the amount of the first and second polynucleotide templates is quantified.

58. The method of claim 54, 55, 56, or 57, wherein the first and second template polynucleotides comprise genomic DNA.

59. A method for detecting presence or absence of a mutation in a template, comprising:

(a) generating a labeled polynucleotide, or fragments thereof, by any of the methods of claims 1-34 and 36; and

(b) analyzing the labeled polynucleotide, or fragments thereof, whereby presence or absence of a mutation is detected.

60. The method of claim 59, wherein the labeled polynucleotide, or fragments thereof, is compared to a reference template.

61. The method of claim 59 or 60, wherein the mutation is selected from the group consisting of a base substitution, a base insertion, a base deletion, and a single nucleotide polymorphism.

63. A composition comprising (a) UNG; (b) N, N'-dimethylethylenediamine; and (c) ARP.

64. The composition of claim 63, wherein the composition further comprises (d) dUTP

65. The composition of claim 64, wherein the composition further comprises: (e) a DNA polymerase; (f) a composite primer, wherein the composite primer comprises a 5' RNA portion and a 3' DNA portion; and (g) an agent capable of cleaving RNA from an RNA-DNA hybrid.

66. A composition comprising: (a) a non-canonical nucleotide; (b) an agent capable of cleaving a base portion of a non-canonical nucleotide; (c) an agent capable of cleaving a phosphodiester backbone at an abasic site; (d) an agent capable of labeling an

IDT Exhibit 1020.8056

abasic site; and (e) a DNA polymerase; (f) a composite primer, wherein the composite primer comprises a 5' RNA portion and a 3' DNA portion; and (g) an agent capable of cleaving RNA from an RNA-DNA hybrid.

67. The composition of claim 66, wherein the composition further comprises: (h) an acetic acid solution; and (l) an $MgCl_2$ solution.

68. A composition comprising: (a) one or more of (i) a non-canonical nucleotide; (ii) an agent capable of cleaving a base portion of a non-canonical nucleotide; (iii) an agent capable of cleaving a phosphodiester backbone at an abasic site; and (iv) an agent capable of labeling an abasic site; and (b) a composite primer, wherein the composite primer comprises an RNA portion and a 3' DNA portion.

69. A composition comprising (a) one or more of (i) a non-canonical nucleotide; (ii) an agent capable of cleaving a base portion of a non-canonical nucleotide; (iii) an agent capable of cleaving a phosphodiester backbone at an abasic site; and (iv) an agent capable of labeling an abasic site; and (b) an agent capable of cleaving RNA from an RNA-DNA hybrid.

70. The composition of claim 68 or 69, wherein (i) is dUTP.

71. The composition of claim 68 or 69, wherein (ii) is UNG.

72. The composition of claim 68 or 69, wherein (iii) is N, N'-dimethylethylenediamine.

73. The composition of claim 68 or 69, wherein (iv) is ARP.

74. The composition of claim 68, wherein the RNA portion of the composite primer is 5' with respect to the 3' DNA portion, the 5' RNA portion is adjacent to the 3' DNA portion, the RNA portion of the composite primer consists of about 10 to about 20 nucleotides and the DNA portion of the composite primer consists of about 7 to about 20 nucleotide.

75. The composition of claim 69, wherein the agent that cleaves RNA from an RNA-DNA hybrid is RNAse H.

IDT Exhibit 1020.8057

76. A kit for use in the methods of any of claims 1-35 and 38-62, said kit comprising: (a) UNG; (b) N, N'-dimethylethylenediamine; and (c) ARP.

77. The kit of claim 76, wherein the kit further comprises (d) dUTP

78. The kit of claim 77, wherein the kit further comprises: (e) a DNA polymerase; (f) a composite primer, wherein the composite primer comprises a 5' RNA portion and a 3' DNA portion; and (g) an agent capable of cleaving RNA from an RNA-DNA hybrid.

79. A kit for use in the methods of any of claims 1-35 and 38-62, said kit comprising: (a) a non-canonical nucleotide; (b) an agent capable of cleaving a base portion of a non-canonical nucleotide; (c) an agent capable of cleaving a phosphodiester backbone at an abasic site; (d) an agent capable of labeling an abasic site; and (e) a DNA polymerase; (f) a composite primer, wherein the composite primer comprises a 5' RNA portion and a 3' DNA portion; and (g) an agent capable of cleaving RNA from an RNA-DNA hybrid.

80. The kit of claim 66, wherein the kit further comprises: (h) an acetic acid solution; and (I) an $MgCl_2$ solution.

81. A kit for use in the methods of any of claims 1-35 and 38-62, said kit comprising: (a) one or more of (i) a non-canonical nucleotide; (ii) an agent capable of cleaving a base portion of a non-canonical nucleotide; (iii) an agent capable of cleaving a phosphodiester backbone at an abasic site; and (iv) an agent capable of labeling an abasic site; and (b) a composite primer, wherein the composite primer comprises an RNA portion and a 3' DNA portion.

82. A kit for use in the methods of any of claims 1-35 and 38-62, said kit comprising: (a) one or more of (i) a non-canonical nucleotide; (ii) an agent capable of cleaving a base portion of a non-canonical nucleotide; (iii) an agent capable of cleaving a phosphodiester backbone at an abasic site; and (iv) an agent capable of labeling an abasic site; and (b) an agent capable of cleaving RNA from an RNA-DNA hybrid.

83. The kit of claim 80 or 81, wherein (i) is dUTP.

IDT Exhibit 1020.8058

84. The kit of claim 80 or 81, wherein (ii) is UNG.

85. The kit of claim 80 or 81, wherein (iii) is N, N'-dimethylethylenediamine.

86. The kit of claim 80 or 81, wherein (iv) is ARP.

87. The kit of claim 80, wherein the RNA portion of the composite primer is 5' with respect to the 3' DNA portion, the 5' RNA portion is adjacent to the 3' DNA portion, the RNA portion of the composite primer consists of about 10 to about 20 nucleotides and the DNA portion of the composite primer consists of about 7 to about 20 nucleotide.

88. The kit of claim 81, wherein the agent that cleaves RNA from an RNA-DNA hybrid is RNAse H.

89. The kit of claim 80 or 81, wherein (ii) is an enzyme.

IDT Exhibit 1020.8059

Figure 1

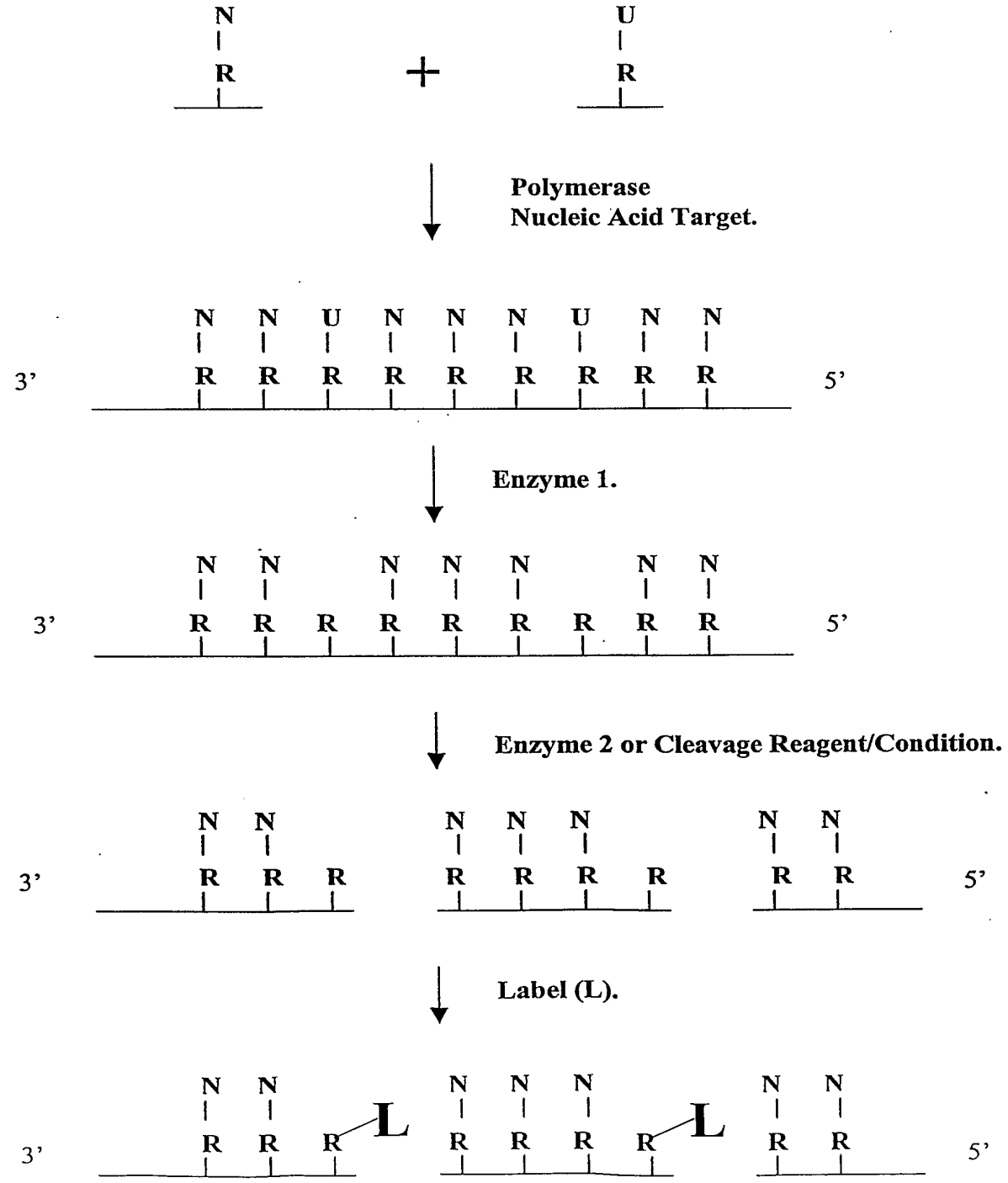

N
R
U
R
Polymerase
Nucleic Acid Target.
3'
N N U N N N U N N
R R R R R R R R R
5'
Enzyme 1.
3'
N N N N N N N
R R R R R R R R R
5'
Enzyme 2 or Cleavage Reagent/Condition.
3'
N N
R R R
N N N
R R R R
N N
R R
5'
Label (L).
3'
N N
R R R
L
N N N
R R R R
L
N N
R R
5'

Figure 2

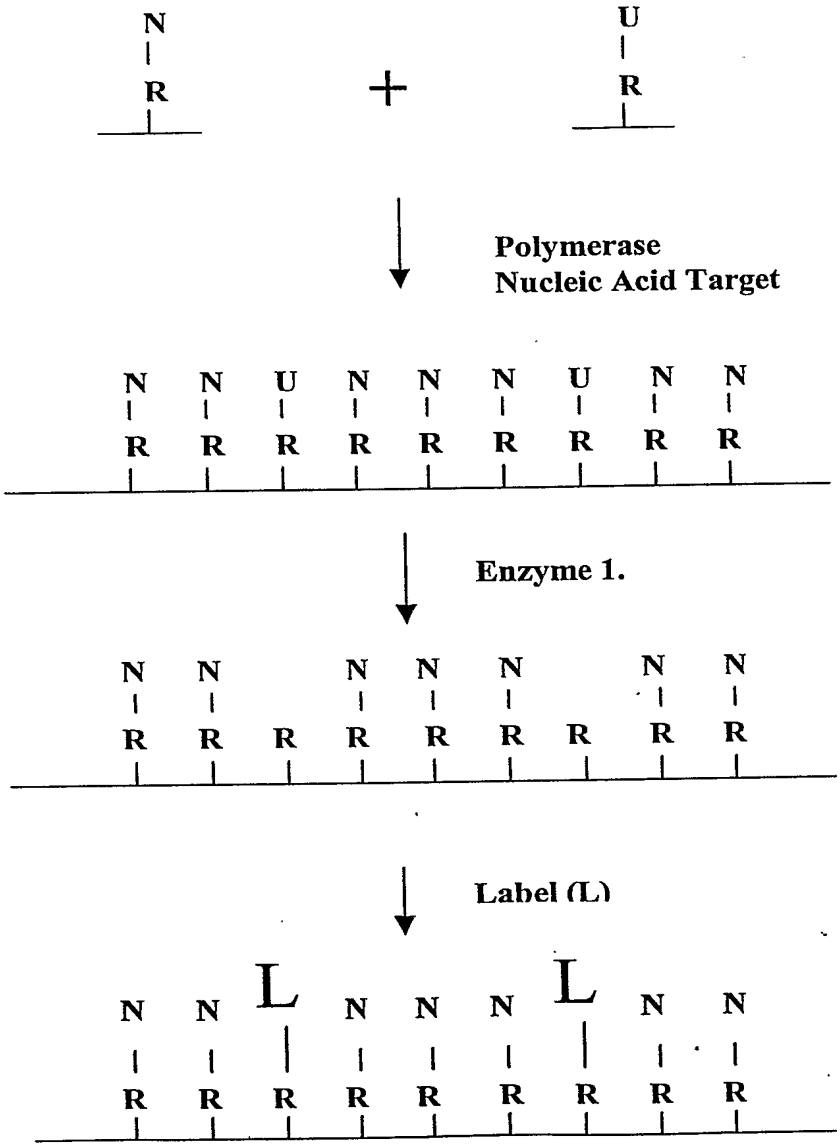

N
R
+
U
R
Polymerase
Nucleic Acid Target
N N U N N N U N N
R R R R R R R R R
Enzyme 1.
N N N N N N N
R R R R R R R R R
Label (L)
N N L N N N L N N
R R R R R R R R R

Figure 3

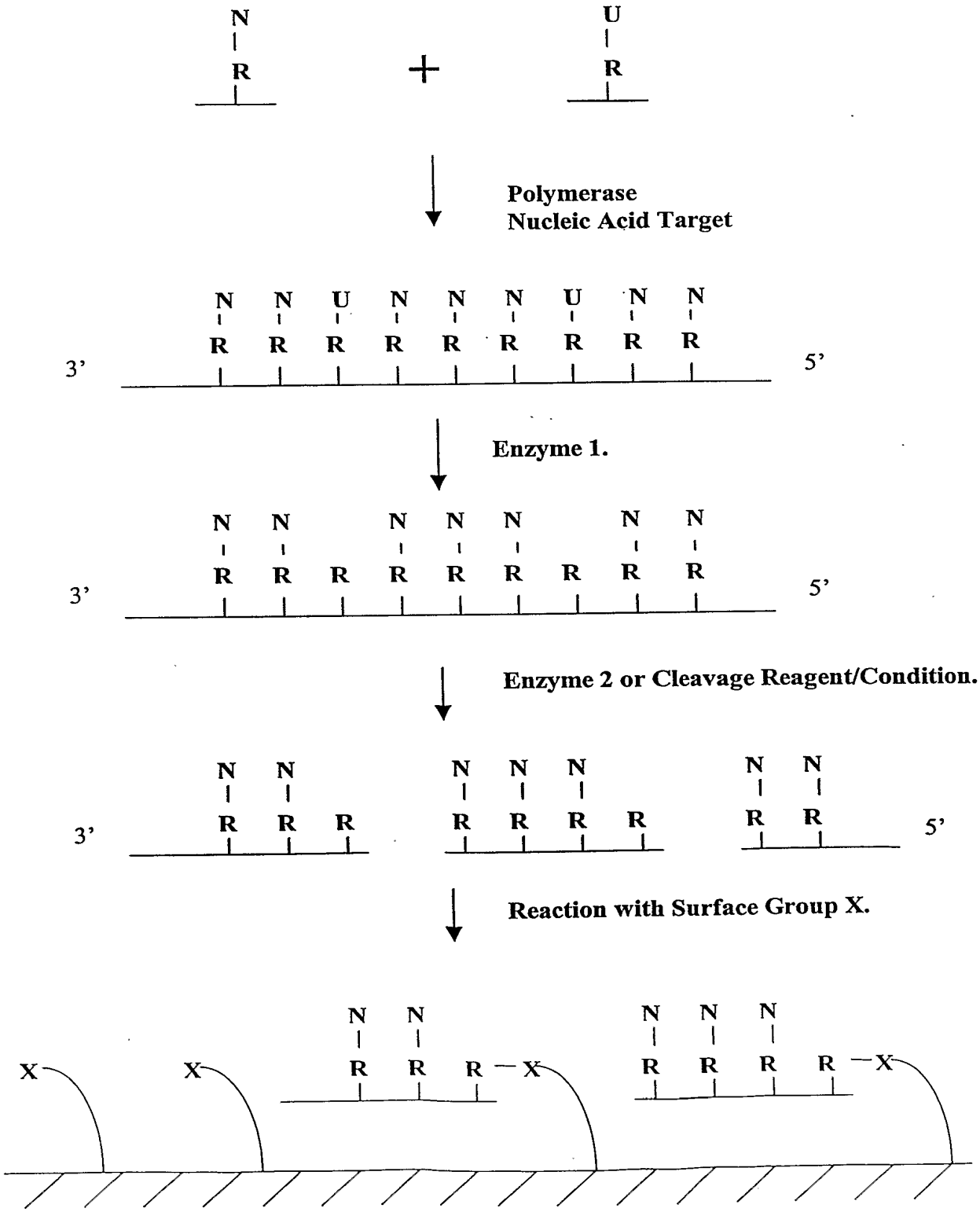
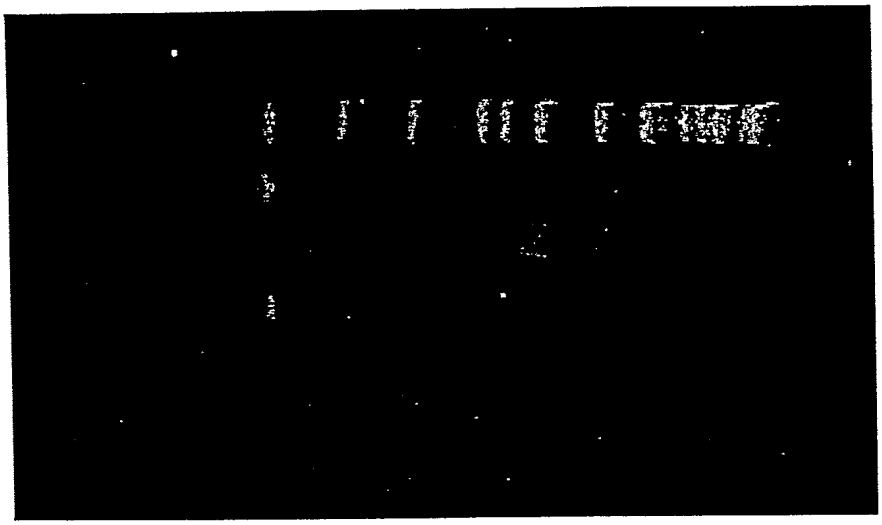

N
R
U
R
Polymerase
Nucleic Acid Target
3'
5'
Enzyme 1.
Enzyme 2 or Cleavage Reagent/Condition.
Reaction with Surface Group X.
X

Figure 4



No Label
L3.8+Strep
L3.8

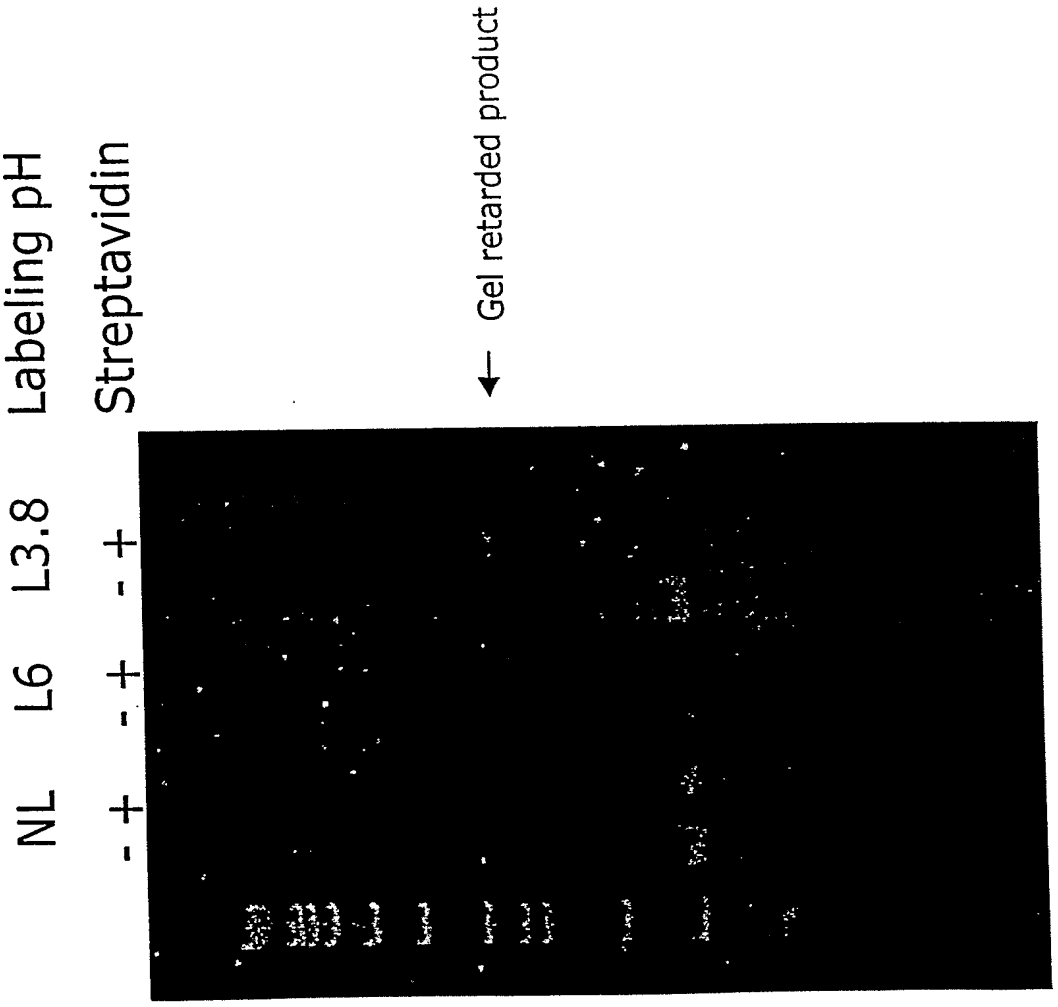


Figure 5

Figure 6

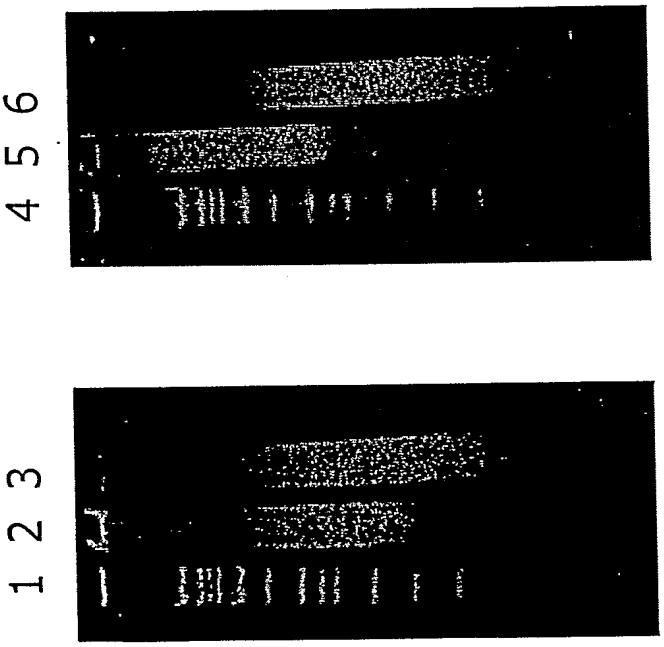
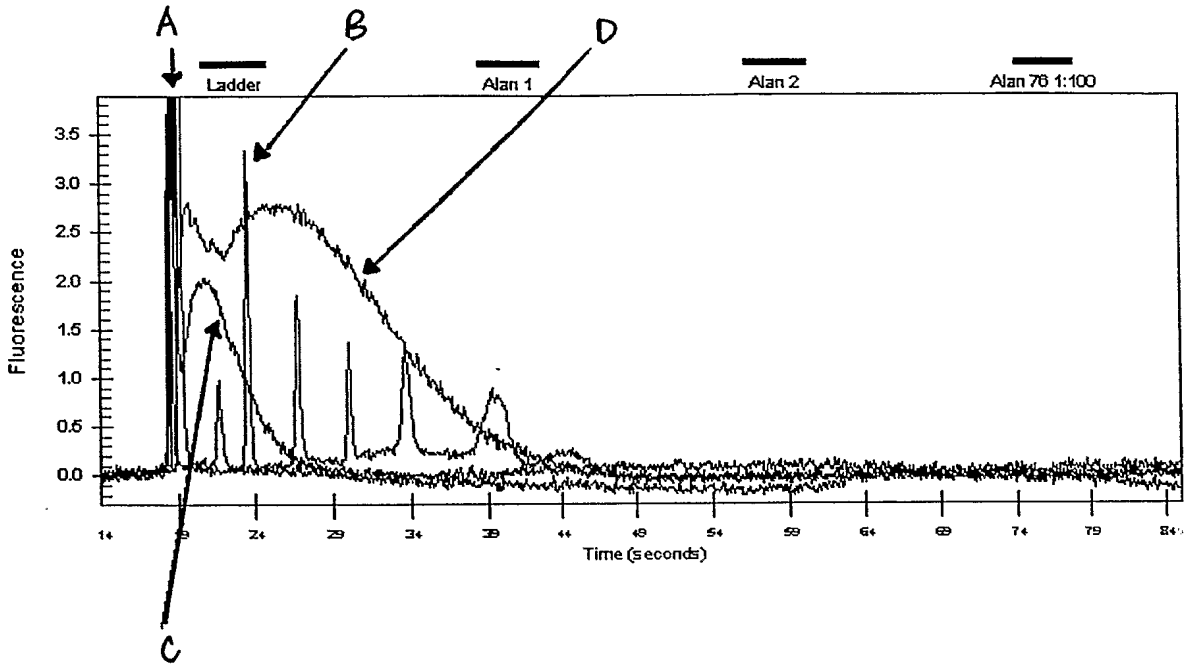


Figure 7


A
B
C
D
Ladder
Alan 1
Alan 2
Alan 76 1:100
Fluorescence
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Time (seconds)

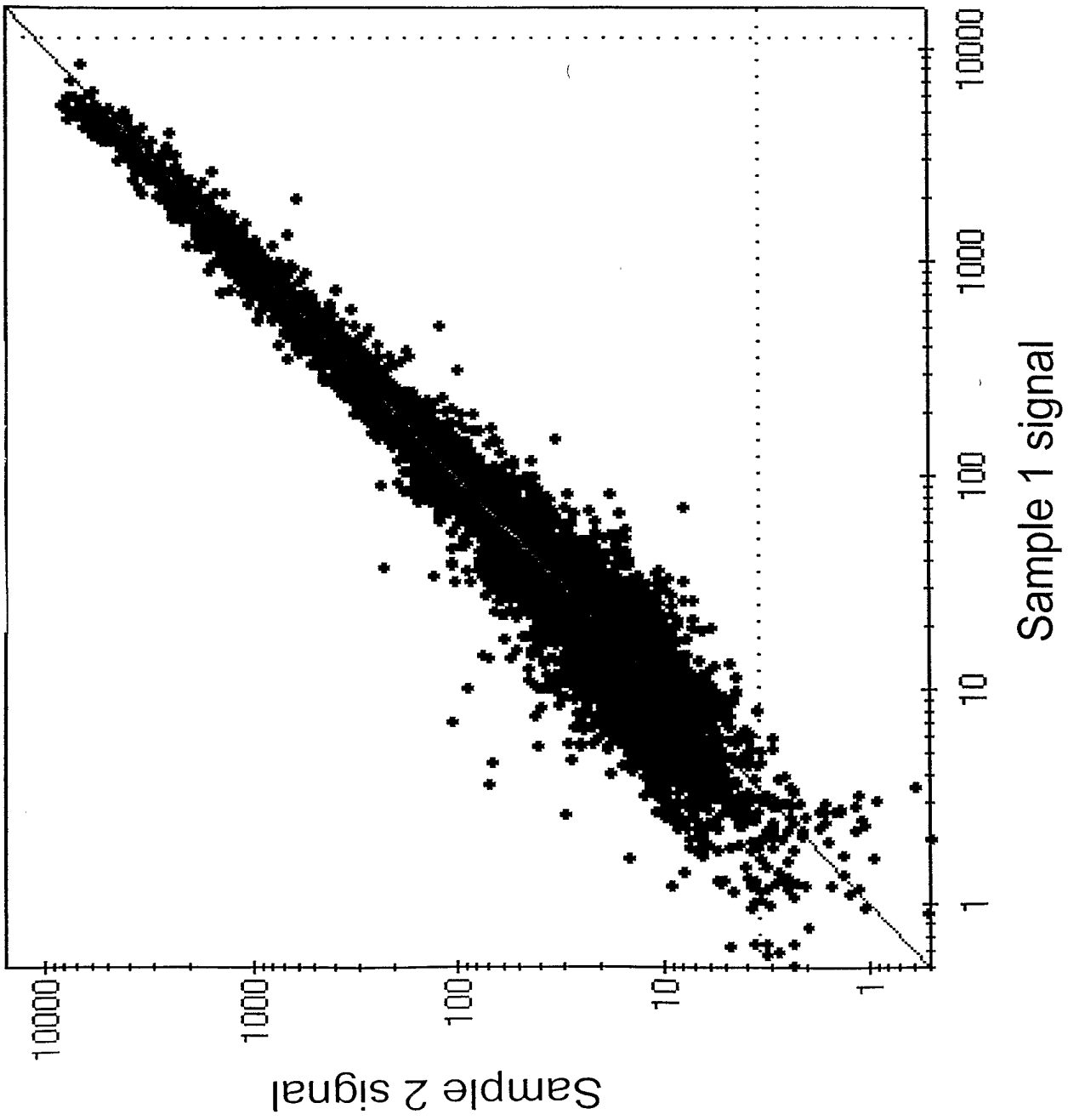

Sample 2 signal
10000
1000
100
10
1
Sample 1 signal
1
10
100
1000
10000

Figure 8

Espacenet

Bibliographic data: WO0070039 (A1) — 2000-11-23

METHOD FOR SUBTRACTING cDNAs BY SUPPRESSING THE SYNTHESIS OF SPECIFICALLY TARGETED mRNAs

Inventor(s): PELLETIER JERRY [CA] ± (PELLETIER, JERRY)

Applicant(s): UNIV MCGILL [CA]; PELLETIER JERRY [CA] ± (MCGILL UNIVERSITY, ; PELLETIER, JERRY)

Classification: - **international:** ***C12N15/10;*** (IPC1-7): C12N15/10; C12Q1/68
- **cooperative:** **C12N15/1096** more

Application number: WO2000CA00586 20000517 Global Dossier

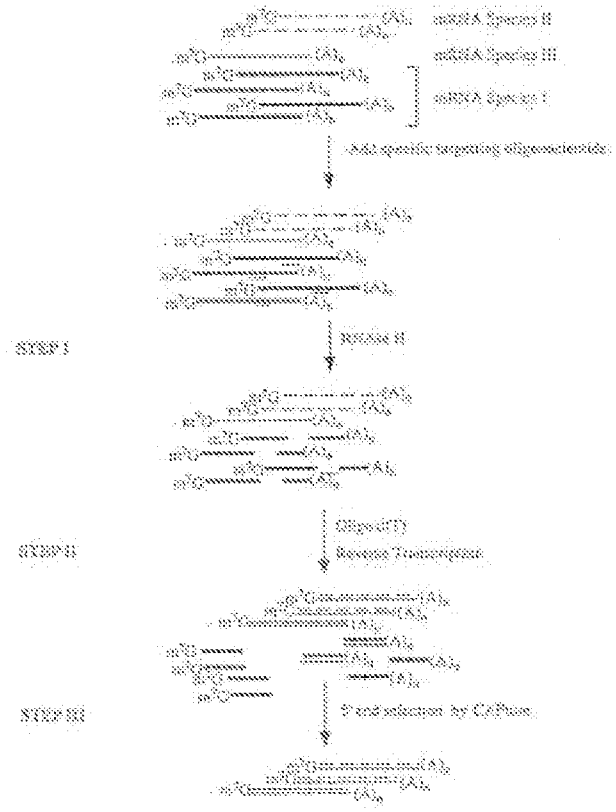
Priority number (s): US19990134548P 19990517

Also published as: AU4739600 (A)

Abstract of WO0070039 (A1)

The present invention relates to genetic engineering and more particularly to a method to specifically modify a mRNA population. The present invention can therefore be applied to the field of cDNA synthesis and cDNA cloning. More specifically, the invention relates to methods of customizing an mRNA population by selectively eliminating chosen mRNAs. The invention further relates to customized cDNA libraries generated from such a customized mRNA population. The invention also relates to a method of specifically targeting and cleaving and/or degrading mRNAs in order to customize the mRNA population. In addition, the invention relates to a method for subtracting cDNAs by suppressing the synthesis of specifically targeted mRNAs.

IDT Exhibit 1020.8068



IDT Exhibit 1020.8069

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/10, C12Q 1/68 — A1

(11) International Publication Number: **WO 00/70039**

(43) International Publication Date: 23 November 2000 (23.11.00)

(21) International Application Number: PCT/CA00/00586

(22) International Filing Date: 17 May 2000 (17.05.00)

(30) Priority Data:
60/134,548 17 May 1999 (17.05.99) US

(71) Applicant *(for all designated States except US)*: MCGILL UNIVERSITY [CA/CA]; 845 Sherbrooke Street West, Montreal, Quebec H3A 1B1 (CA).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: PELLETIER, Jerry [CA/CA]; 8 Lakeview, Baie D'Urfée, Quebec H9X 3B1 (CA).

(74) Agents: DUBUC, Jean, H. et al.; Goudreau Gage Dubuc, The Stock Exchange Tower, 800 Place Victoria, Suite 3400, P.O. Box 242, Montreal, Quebec H4Z 1E9 (CA).

(81) Designated States: AE, AG, 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

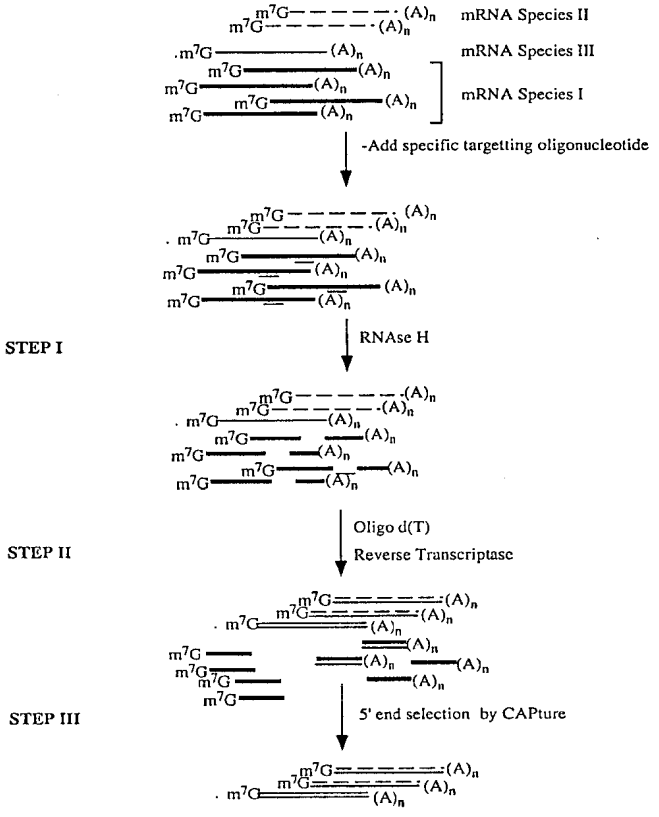
With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHOD FOR SUBTRACTING cDNAs BY SUPPRESSING THE SYNTHESIS OF SPECIFICALLY TARGETED mRNAs

(57) Abstract

The present invention relates to genetic engineering and more particularly to a method to specifically modify a mRNA population. The present invention can therefore be applied to the field of cDNA synthesis and cDNA cloning. More specifically, the invention relates to methods of customizing an mRNA population by selectively eliminating chosen mRNAs. The invention further relates to customized cDNA libraries generated from such a customized mRNA population. The invention also relates to a method of specifically targeting and cleaving and/or degrading mRNAs in order to customize the mRNA population. In addition, the invention relates to a method for subtracting cDNAs by suppressing the synthesis of specifically targeted mRNAs.



IDT Exhibit 1020.8070

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 1020.8071

TITLE OF THE INVENTION

METHOD FOR SUBTRACTING cDNAs BY SUPPRESSING THE SYNTHESIS OF SPECIFICALLY TARGETED mRNAS

FIELD OF THE INVENTION

The present invention relates to genetic engineering and more particularly to a method to specifically modify a mRNA population. The present invention can therefore be applied to the field of cDNA synthesis and cDNA cloning. More specifically, the invention relates to methods of customizing an mRNA population by selectively eliminating chosen mRNAs. The invention further relates to customized cDNA libraries generated from such a customized mRNA population. The invention also relates to a method of specifically targeting and cleaving and/or degrading mRNAs in order to customize the mRNA population. In addition, the invention relates to a method for subtracting cDNAs by suppressing the synthesis of specifically targeted mRNAs.

BACKGROUND OF THE INVENTION

The isolation and rapid mapping of complementary DNAs (cDNAs) is central to characterizing the information that is of significant biological relevance in the genome of an organism. A full length cDNA allows one to predict transcription initiation start sites, translation initiation start sites, deduce certain protein characteristics based on primary amino acid sequences, predict transcription termination sites, and visually inspect the 5' and 3' untranslated regions for elements which may be involved in post-transcriptional regulation of gene expression. The analysis of several complete cDNAs of a given gene enables one to gather information on alternative splicing, alternative promoter usage, and alternative polyadenylation signals - all events known to be important in gene expression regulation. In addition, the comparison of genomic and cDNA sequence is essential to determine exon-intron structure and document the

occurrence of RNA editing - a post-transcriptional regulatory mechanism on which only little information is available.

The cloning of mRNA into cDNA for the purposes of functional studies is a complex, interrelated series of enzyme-catalyzed reactions involving the *in vitro* synthesis of a DNA copy of mRNA, its subsequent conversion to duplex cDNA, and insertion into an appropriate prokaryotic vector. The procedure usually involves the following series of steps:

I) Isolation of high quality mRNA from the tissue or cell line of interest.

II) Annealing of a DNA oligonucleotide, either a random hexamer, an oligo d(T) primer or specific oligo, to the mRNA. When full-length cDNAs are required, oligo d(T) is utilized, since this is expected to anneal to the 3' poly (A) tail of the mRNA.

III) Reverse transcriptase is then utilized to prime from the DNA primer and copy the RNA template into cDNA.

IV) Second strand synthesis is performed utilizing RNAse H, DNA polymerase I, and DNA ligase.

V) The ends of the cDNAs are polished, prepared for cloning, and the cDNAs are introduced into an appropriate vector.

Large-scale sequencing of cDNAs randomly picked from libraries has proven to be a very powerful approach to discover expressed sequences that, in turn, once mapped, greatly expedite the process involved in the identification and cloning genes linked to human diseases. However, an impediment to this process is due to the different abundance of mRNAs within a given cDNA library. The mRNAs of a typical cell are distributed into 3 frequency classes (1, 2) that are presumably maintained in representation in cDNA libraries. The frequency distribution of mRNAs within these classes are as follows:

I) a highly abundant class consisting of 10 - 15 mRNAs that altogether represent 10-20% of the total mRNA mass;

II) a middle abundance class consisting of 1000 - 2000 mRNAs making up 30-40% of the total mRNA; and

III) a low abundance class consisting of 15,000 - 20,000 mRNAs covering ~50% of the total mRNA (2).

The distribution of the different mRNAs in these three classes implies that within any given cDNA library, about 20% of the total number of clones in that library come from only ~10-15 abundant mRNA species. Thus, random picking of clones in a cDNA library, often leads to the re-isolation of cDNAs corresponding to the same abundant mRNAs, thereby reducing the frequency at which new genes can be efficiently isolated.

As seen above, reassociation-kinetic analysis indicates that the population of mRNAs of a typical cell can be subdivided into three frequency classes (ref. 2). For gene identification endeavors, a procedure which "normalizes" the relative abundance of genes in a given library can be an important issue in order to provide more "depth" to cDNA libraries.

To date, a small number of normalization protocols in which removal of abundant cDNAs from libraries is performed during their construction have been prepared. For example, current approaches involve a complex series of steps utilizing reassociation kinetics between 2 copies of the libraries or between cDNA and RNA. Since hybridization of RNA or DNA is concentration dependent, rarer species anneal less rapidly and abundant species anneal more rapidly. By controlling the time of annealing, one can achieve a point where abundant mRNA or cDNA are annealed to their complement while less abundant cDNAs are not (refs. 3,4). The major drawback of the current methods of normalization may be 1) the complexity of the procedure, and 2) artifacts which arise during the procedure. For example, the length of some cDNAs may be noticeably shortened in some normalized libraries (ref. 4), so clearly these approaches are not compatible with full-length (FL) cDNA library generation. 3) In addition, the number of additional manipulations are time-consuming and labor intensive.

There thus remains a need to develop a method or procedure, which when implemented during cDNA library construction, reduces the abundance of a target gene or genes in the cDNA library. There also remains the need to devise such a method which could be used to generate cDNA libraries in which clones corresponding to high abundance mRNAs are under-represented, and hence the rate of new gene discovery in such libraries would be greatly enhanced and facilitated. Furthermore, there remains a need to provide a method which could be used to suppress the synthesis of any cDNA of any abundance class. More broadly, there remains a need to provide a simple and versatile method enabling a customization of an mRNA population that is not strictly based on mRNA abundance.

The present invention seeks to meet these and other needs.

The present description refers to a number of documents, the content of which is herein incorporated by reference in their entirety.

SUMMARY OF THE INVENTION

The invention therefore seeks to overcome the drawbacks of the methods and procedures of the prior art.

In general, the invention relates to a method which enables a versatile customization of an mRNA population. In a preferred embodiment, the invention relates to a method of customizing an mRNA population which comprises a targeting of a chosen mRNA or mRNAs with at least one oligonucleotide, under conditions which favor the formation of a duplex between the targeted mRNA or mRNAs and at least one oligonucleotide. The method then comprises a separation of the duplexed mRNA from non-duplexed mRNA, and/or the use of the properties of duplexed versus non-duplexed mRNA (e.g. RT processivity inhibition, differential substrate for cleavage/degradation) thereby yielding a customized mRNA population.

In general, the invention also relates to methods based on the use of specific oligonucleotides to customize an mRNA population.

When such a method targets high abundance mRNAs, and is coupled to a method of generating cDNA libraries, the frequency of re-isolation of cDNAs can then be reduced and, as a corollary, favor the isolation of novel cDNAs.

In general, the invention also concerns methods to subtract targeted cDNAs by suppressing their synthesis.

In one particular embodiment, the present invention also relates to a simple method for normalizing mRNAs which does not compromise the generation of a full length cDNA library therefrom.

The present invention also concerns methods of controlling the time of annealing in order to remove abundant mRNAs or cDNAs from libraries while maintaining less abundant mRNAs or cDNAs in the library.

In another embodiment, the invention relates to a procedure of suppressing the synthesis of target transcripts during the first strand cDNA synthesis, based on the ability of a duplex specific RNAse to specifically target and degrade the RNA moiety of a RNA/DNA duplex. In a preferred embodiment the RNAse is RNAse H.

In one favored embodiment, the mRNA population is customized using an annealing of a complementary DNA oligonucleotide to a given RNA to obtain a RNA/DNA duplex, treating the duplex with RNAse H, thereby resulting in a specific cleavage of only the target RNA at the site of hybridization.

In accordance with the present invention, there is therefore provided a method of customizing an mRNA population comprising selecting an oligonucleotide which is complementary to an mRNA molecule in an mRNA sample, incubating the oligonucleotide with the mRNA sample to obtain an mRNA oligonucleotide mixture under conditions which enable a formation of a duplex between the oligonucleotide and the mRNA molecule and submitting the mRNA-oligonucleotide mixture to a purification step which enables a separation

of duplexed mRNA from non-duplexed mRNA, thereby yielding a customized mRNA population.

There is also provided a method for suppressing the synthesis of target transcripts during first strand cDNA synthesis, comprising the steps of selecting an oligonucleotide which is complementary to an mRNA molecule in an mRNA sample, incubating the oligonucleotide with the mRNA sample to obtain an mRNA oligonucleotide mixture under conditions which enable a formation of a duplex between the oligonucleotide and the mRNA molecule; and submitting the mRNA-oligonucleotide mixture to a duplex specific RNAse which specifically targets the RNA moiety of the duplex; and performing first strand cDNA synthesis on the RNAse-treated duplexes of the previous step.

The present invention also provides a method of subtracting cDNAs by suppressing their synthesis comprising the steps of annealing a set of complementary DNA oligonucleotides to a given mRNA sample under conditions enabling a formation of mRNA-oligonucleotide duplexes; and performing first strand cDNA synthesis on non-duplexed mRNAs.

There is also provided a method of targeting a chosen mRNA or mRNA population for degradation using oligos which are specific for the chosen mRNA or mRNA population, to generate duplexes of oligos and mRNA, submitting the duplex to a duplex-directed degradation or cleavage which leaves non-duplexed mRNA intact and purifying the non-duplexed mRNAs.

The methods of the present invention can provide a certain number of advantages. For example, the invention enables the targeting of any chosen mRNA population, provided that sequence data from the chosen mRNA population is available. The use of oligos to target an mRNA or mRNA population is a versatile and simple way to apply the customizing method of the present invention to any chosen mRNA or mRNA population. Thus, the present invention is not limited to selective elimination of only abundant transcripts.

The well-known principles of screening libraries with oligos can be adapted to the herein disclosed methods of customization of mRNA

populations. Non-limiting examples of mRNAs which could be targeted by the method of the present invention include mRNAs encoding specific protein signatures (e.g. kinase domains, zinc fingers, RNA binding motifs), housekeeping mRNAs, tissue specific mRNAs and mRNAs expressed at specific times during the cell cycle. Thus, the present invention provides a means of selecting any characteristic of a targeted mRNA which can be used to reduce and/or eliminate same from an mRNA population, provided that oligos which are specific and complementary thereto are available. In combination with a cDNA synthesis method, the method of customizing an mRNA population in accordance with the present invention therefore enables the customizing of the cDNA library.

Having targeted a chosen mRNA and enabled the formation of a duplex, the duplexed mRNAs can then be separated from non-duplexed mRNAs, in order to yield the customized mRNA population. While the present invention is described in more detail using a purification through the cap structure to selectively purify such flagged mRNAs, the invention is not so limited. Indeed a number of methods can be used to separate the oligonucleotide-mRNA duplexes from non-duplexed mRNA, in order to yield the customized mRNA population. Non-limiting examples of such methods include hydroxyapatite, which selectively binds double-stranded duplexes, streptavidin/biotin methods, based on the use of biotinylated oligonucleotides, duplex-dependent mRNA cleavage and/or degradation (enzymatic or non-enzymatic directed cleavage/degradation). In addition, a number of methods known to a skilled artisan can be used to flag mRNAs so as to be able to exercise a separation of desired mRNAs or duplexes comprising same.

Of course, it will be understood by the skilled artisan that a targeting of specific mRNAs or RNAs can be done by a number of methods and that the oligos exemplified herein are but a specific embodiment of targeting of specific mRNAs.

One preferred utility of the method of customizing an mRNA population is in the generation of customized cDNA libraries. Indeed, the method of customizing of an mRNA population in accordance with the present invention can be incorporated in an "improved" method to make cDNA libraries.

As described above, one envisaged application of the invention is during first strand cDNA synthesis to remove abundant genes from libraries. This results in a subtraction process and will facilitate the rate of gene discovery. More generally, the cDNA of any gene can be specifically inhibited utilizing an approach described herein.

While the targeting and specific degradation of mRNA has been demonstrated with an RNAse H-dependent (duplex-dependent) degradation of the RNA template, the present invention should not be so limited. Indeed, the present invention extends to any method involving enzymatic or non-enzymatic duplex-directed cleavage and/or degradation of known RNA templates. Non-limiting examples include oligonucleotide metal complex conjugates containing lanthanide complexes (ref. 8).

Similarly, while the specific cleavage and/or degradation of mRNAs has been demonstrated using conventional DNA oligonucleotides to hybridize to the RNA template, any modified oligonucleotide, as known in the art can be used for this purpose, provided that a cleavage of the RNA template occurs, resulting in a physical separation of the 5' and 3' ends of the targeted mRNA template.

Furthermore, while in the embodiment described herein, the RNAse H and Reverse Transcriptase activities are separated in two different steps (Steps I and II in Fig. 3), the instant invention should not be so limited. The most common reverse transcriptases in use, MMLV Superscript™, does not contain RNase H activity. Hence, in accordance with an embodiment based on an enzymatic-directed cleavage of the mRNA, a separate RNAseH activity needs to be added. However, it is known that other reverse transcriptases (RT), such as AMV or HIV reverse transcriptase, contain both activities (reverse

transcriptase and RNAse H) in one complex. Thus, a "one step method" of mRNA cleavage/degradation and cDNA synthesis could be carried out in accordance with the present invention. In such an embodiment, DNA oligonucleotides are annealed to a given RNA template, followed by a direct incubation with the reverse transcriptases containing the RNAse H activity. In this embodiment, the RT could degrade the complementary RNA strand.

It will also be recognized that although preferred, a cleavage/degradation of the targeted mRNA is not necessary in all conditions to enable the construction of a customized cDNA library. Indeed, an inhibition of the processivity of the RT would be sufficient to enable the generation of a customized library. In this particular embodiment, a mRNA or mRNAs would be targeted with oligonucleotides and the processivity of the RT would be inhibited by the duplex, thereby reducing or abrogating the generation of cDNAs from the targeted mRNA templates. The cDNAs can then be separated by a variety of means favoring the purification of full-length cDNAs according to physical feature of the mRNAs (cap-dependent purification, size-dependent separation, poly A-dependent purification). In a preferred embodiment, the oligonucleotides are blocked at the 3' end (as commonly known) to block the priming of the RT from the oligo.

Since the blocking oligonucleotide could also be used as a primer by the RT, in one embodiment, oligonucleotides which contain a modified 3' end (which does not serve as primer for the RT) are used. Such oligonucleotides which cannot act as primers (6) were used in Fig. 4 (SEQ ID NOS:2 and 3).

In one particular embodiment exemplified below, the targeted mRNAs are degraded by RNAse H and full-length mRNAs are selected and purified using oligo dT priming, reverse transcriptase and CAPture™. It will be recognized therefore that the present invention enables a "flagging" of specific mRNAs (e.g. mRNA/oligo duplexes). These duplexes can serve as targets for cleavage or degradation or can block the processivity of a DNA polymerase (e.g.

R.T.). For cDNA synthesis in which a selection of full length mRNAs is preferred, a method which somehow couples the 3' and 5' ends of an mRNA is preferred (e.g. oligo dT priming and selection of capped mRNA).

The instant invention further extends to PCR analysis of gene products generated by RT of mRNA into cDNA. For example, primers flanking both sides of the duplex could be used.

While mRNA is the preferred targeted nucleic acid in accordance with the methods of the present invention, it should be recognized that RNA in general can be targeted with oligonucleotides in order to customize the RNA population. It will be recognized by the skilled artisan that the methods of the present invention can be adapted to target, for example, uncapped mRNAs (and to generate a cDNA library therefrom).

It shall also be recognized that the present invention could also be adapted to the use of RNA oligonucleotides.

In one embodiment, full-length cDNAs are selected via a cap structure-dependent purification step using a cap column (CAPture™). It shall be recognized that other cap-dependent purification methods can be used. Non-limiting examples thereof include anti-cap antibodies and direct oxydization of the cap structure (refs. 9, 10, 11).

In accordance with one embodiment of the present invention, there is provided a method of customizing an mRNA population comprising selecting an oligonucleotide which is complementary to an mRNA molecule in an mRNA sample; incubating the oligonucleotide with the mRNA sample to obtain an mRNA oligonucleotide mixture under conditions which enable a formation of a duplex between the oligonucleotide and the mRNA molecule; and submitting the mRNA-oligonucleotide mixture to a purification step which enables a separation of duplexed mRNA from non-duplexed mRNA, thereby yielding a customized mRNA population.

In another embodiment of the present invention, the mRNA oligonucleotide mixture enables a flagging of specific mRNAs, such that this

flagging enables a selective purification or targeting of chosen mRNAs from the oligonucleotide mRNA mixture.

In accordance with another embodiment of the present invention, there is provided a method for suppressing the synthesis of target transcripts during first strand cDNA synthesis, comprising: selecting an oligonucleotide which is complementary to an mRNA molecule in an mRNA sample; incubating the oligonucleotide with the mRNA sample to obtain an mRNA oligonucleotide mixture, under conditions which enable a formation of a duplex between the oligonucleotide and the mRNA molecule; submitting the mRNA-oligonucleotide mixture to a duplex-specific RNAse which specifically targets the RNA moiety of the duplex; and performing first strand cDNA synthesis on the RNAse-treated duplexes, thereby suppressing the synthesis of RNAse-cleaved duplexes.

In accordance with yet another embodiment of the present invention, there is provided a method of subtracting cDNAs by suppressing their synthesis comprising the steps of annealing a set of complementary DNA oligonucleotides to a given mRNA sample under conditions enabling a formation of mRNA-oligonucleotide duplexes; and performing first strand cDNA synthesis on non-duplexed mRNAs.

In addition, in accordance with another embodiment of the present invention, there is provided a method of customizing an mRNA population comprising: selecting an oligonucleotide which is complementary to an mRNA molecule in an mRNA sample; incubating the oligonucleotide with the mRNA sample to obtain an mRNA oligonucleotide mixture under conditions which enable a formation of a duplex between the oligonucleotide and the mRNA molecule; and submitting the mRNA-oligonucleotide mixture to a duplex-dependent cleavage and/or degradation, thereby yielding a customized mRNA population.

In accordance with the present invention, there is also provided customized mRNA or RNA populations or customized cDNA libraries obtained using one of the methods of the present invention.

In accordance with a related embodiment of the present invention, the RNA or mRNA oligonucleotide mixture, instead of being submitted to the duplex-dependent cleavage and/or degradation, is submitted to a purification step enabling the separation of duplexed RNA or duplexed mRNA from non-duplexed RNA or mRNA.

In order to provide a clear and consistent understanding of terms used in the present description, a number of definitions are provided hereinbelow.

DEFINITIONS

Nucleotide sequences are presented herein by single strand, in the 5' to 3' direction, from left to right, using the one letter nucleotide symbols as commonly used in the art and in accordance with therecommendations of the IUPAC-IUB Biochemical Nomenclature Commission.

Unless defined otherwise, the scientific and technological terms and nomenclature used herein have the same meaning as commonly understood by a person of ordinary skill to which this invention pertains. Generally, the procedures for cell cultures, infection, molecular biology methods and the like are common methods used in the art. Such standard techniques can be found in reference manuals such as for example Sambrook et al. (1989, Molecular Cloning - A Laboratory Manual, Cold Spring Harbor Laboratories) and Ausubel et al. (1994, Current Protocols in Molecular Biology, Wiley, New York).

The present description refers to a number of routinely used recombinant DNA (rDNA) technology terms. Nevertheless, definitions of selected examples of such rDNA terms are provided for clarity and consistency. For certainty, it is emphasized that the present invention finds utility with nucleic acids in general. Non-limiting examples of nucleic acids which can be used in

accordance with the teachings of the present invention include that from eukaryotic cells such as that of animal cells, plant cells, or microorganisms as well as that from prokaryotic cells and viruses.

As used herein, the term "processivity" of a polymerase refers to its property to continue to act on a substrate instead of dissociating therefrom

As used herein, "nucleic acid molecule", refers to a polymer of nucleotides. Non-limiting examples thereof include DNA (e.g. genomic DNA, cDNA) and RNA molecules (e.g. mRNA). The nucleic acid molecule can be obtained by cloning techniques or synthesized. DNA can be double-stranded or single-stranded (coding strand or non-coding strand [anthozoans]).

The term "recombinant DNA" as known in the art refers to a DNA molecule resulting from the joining of DNA segments. This is often referred to as genetic engineering.

The terminology "amplification pair" refers herein to a pair of oligonucleotides (oligos) of the present invention, which are selected to be used together in amplifying a selected nucleic acid sequence by one of a number of types of amplification processes, preferably a polymerase chain reaction. Other types of amplification processes include ligase chain reaction, strand displacement amplification, or nucleic acid sequence-based amplification, as explained in greater detail below. As commonly known in the art, the oligonucleotides are designed to bind to a complementary sequence under selected conditions.

The nucleic acid (i.e. DNA or RNA) for practising the present invention may be obtained according to well known methods.

Oligonucleotide probes or primers of the present invention may be of any suitable length, depending on the particular assay format and the particular needs and targeted genomes employed. In general, the oligonucleotide probes or primers are at least 10 nucleotides in length, preferably between 12 and 24 molecules, and they may be adapted to be especially suited to a chosen nucleic acid amplification system. As commonly known in the art, the

oligonucleotide probes and primers can be designed by taking into consideration the melting point of hybrizidation thereof with its targeted sequence (see below and in Sambrook et al., 1989, Molecular Cloning - A Laboratory Manual, 2nd Edition, CSH Laboratories; Ausubel et al., 1989, in Current Protocols in Molecular Biology, John Wiley & Sons Inc., N.Y.).

The term "oligonucleotide" or "DNA" molecule or sequence refers to a molecule comprised of the deoxyribonucleotides adenine (A), guanine (G), thymine (T) and/or cytosine (C), in a double-stranded form. The term "oligonucleotide" or "DNA" can be found in linear DNA molecules or fragments, viruses, plasmids, vectors, chromosomes or synthetically derived DNA. As used herein, particular double-stranded DNA sequences may be described according to the normal convention of giving only the sequence in the 5' to 3' direction. Of course, as well-known to the skilled artisan (and as exemplified hereinbelow), oligonucleotides or DNA can also be found as single-stranded molecules.

As used herein, "oligonucleotides" or "oligos" define a molecule having two or more nucleotides (ribo or deoxyribonucleotides). In essence, "oligonucleotides" define at least dimers of nucleotides. The size of the oligonucleotide will be dictated by the particular situation and ultimately on the particular use thereof and adapted accordingly by the person of ordinary skill. An oligonucleotide can be synthesized chemically or derived by cloning according to well known methods. In accordance with preferred embodiments of the present invention, the oligos are single-stranded oligonucleotides.

Probes and oligonucleotides of the invention can be utilized with naturally occurring sugar-phosphate backbones as well as modified backbones including phosphorothioates, dithionates, alkyl phosphonates and α-nucleotides and the like. Modified sugar-phosphate backbones are generally taught by Miller, 1988, Ann. Reports Med. Chem. 23:295 and Moran et al., 1987, Nucleic acid molecule. Acids Res., 14:5019. Probes of the invention can be constructed of either ribonucleic acid (RNA) or deoxyribonucleic acid (DNA), and preferably of DNA. General teachings on the synthesis of oligonucleotides and

substituents and modifications thereof can be found for example in US 5,438,131. The selection of the best suited synthesis pathway of an oligonucleotide and of the appropriate modifications, and substituentsto be used, may be selected accordingly by the person of ordinary skill to which the instant invention pertains.

As used herein, a "primer" defines an oligonucleotide which is capable of annealing to a target sequence, thereby creating a double stranded region or duplex which can serve as an initiation point for DNA synthesis under suitable conditions.

"Nucleic acid hybridization" refers generally to the hybridization of two single-stranded nucleic acid molecules having complementary base sequences, which under appropriate conditions will form a thermodynamically favored double-stranded structure. Examples of hybridization conditions can be found in the two laboratory manuals referred above (Sambrook et al., 1989, *supra* and Ausubel et al., 1997, *supra*) and are commonly known in the art. In the case of a hybridization toa nitrocellulose filter, as for example in the well known Southern blotting procedure, a nitrocellulose filter can be incubated overnight at 65°C with a labeled probe in a solution containing 50% formamide, high salt (5 x SSC or 5 x SSPE), 5 x Denhardt's solution, 1% SDS, and 100 µg/ml denatured carrier DNA (e.g. salmon sperm DNA). The non-specifically binding probe can then be washed off the filter by several washes in 0.2 x SSC/0.1% SDS at a temperature which is selected in view of the desired stringency: room temperature (low stringency), 42°C (moderate stringency) or 65°C (high stringency). The selected temperature is based on the melting temperature (Tm) of the DNA hybrid.Of course, RNA-DNA hybrids can also be formed and detected. In such cases. the conditions of hybridization and washing can be adapted according to well known methods by the person of ordinary skill. Stringent conditions will be preferably used (Sambrook et al.,1989, *supra*).

Although the present invention is not specifically dependent on the use of a label, such a label might be beneficial in certain embodiments. Probes or oligonucleotides can be labeled according to numerous well known methods (Sambrook et al., 1989, supra). Non-limiting examples of labels include ^{3}H, ^{14}C, ^{32}P, ^{33}P and ^{35}S. Non-limiting examples of detectable markers include ligands, fluorophores, chemiluminescent agents, enzymes, and antibodies. Other detectable markers for use with oligos, which can enable a detection and/or purification and/or separation in accordance with the methods of the invention, include biotin/streptavidin. It will become evident to the person of ordinary skill that the choice of a particular label and/or "bait" dictates the manner in which it is bound to the oligonucleotide and the manner in which it is separated through the "bait" by the "prey".

As commonly known, radioactive nucleotides can be incorporated into oligonucleotides of the invention by several methods. Non-limiting examples thereof include kinasing the 5' ends of the probes using γ-^{32}P ATP and polynucleotide kinase, using the Klenow fragment of Pol I of *E. coli* in the presence of radioactive dNTP (i.e. uniformly labeled DNA probe using random oligonucleotide primers in low-melt gels), using the SP6/T7 system to transcribe a DNA segment in the presence of one or more radioactive NTP, and the like.

Similarly, the oligos of the present invention can incorporate a radiolabel by T4 DNA polynucleotide kinase or can be synthesized with fluorescently labelled bases.

Amplification of a selected, or target, nucleic acid sequence may be carried out by a number of suitable methods. See generally Kwoh et al., 1990, Am. Biotechnol. Lab. 8:14-25. Numerous amplification techniques have been described and can be readily adapted to suit particular needs of a person of ordinary skill. Non-limiting examples of amplification techniques include polymerase chain reaction (PCR), ligase chain reaction (LCR), strand displacement amplification (SDA), transcription-based amplification, the Qβ

replicase system and NASBA (Kwoh et al., 1989, Proc. Natl. Acad. Sci. USA 86, 1173-1177; Lizardi et al., 1988, BioTechnology 6:1197-1202; Malek et al., 1994, Methods Mol. Biol., 28:253-260; and Sambrook et al., 1989, *supra*). Preferably, amplification will be carried out using PCR.

Polymerase chain reaction (PCR) is carried out in accordance with known techniques. See, e.g., U.S. Pat. Nos. 4,683,195; 4,683,202; 4,800,159; and 4,965,188 (the disclosures of all three U.S. Patent are incorporated herein by reference). In general, PCR involves, a treatment of a nucleic acid sample (e.g., in the presence of a heat stable DNA polymerase) under hybridizing conditions, with one oligonucleotide primer for each strand of the specific sequence to be detected. An extension product of each primer which is synthesized is complementary to each of the two nucleic acid strands, with the primers sufficiently complementary to each strand of the specific sequence to hybridize therewith. The extension product synthesized from each primer can also serve as a template for further synthesis of extension products using the same primers. Following a sufficient number of rounds of synthesis of extension products, the sample is analysed to assess whether the sequence or sequences to be detected are present. Detection of the amplified sequence may be carried out by visualization following EtBr staining of the DNA following gel electrophoresis, or using a detectable label in accordance with known techniques, and the like. For a review on PCR techniques (see PCR Protocols, A Guide to Methods and Amplifications, Michael et al. Eds, Acad. Press, 1990).

As used herein, the term "gene" is well known in the art and relates to a nucleic acid sequence defining a single protein or polypeptide. A "structural gene" defines a DNA sequence which is transcribed into RNA and translated into a protein having a specific amino acid sequence thereby giving rise the a specific polypeptide or protein. It will be readily recognized by the person of ordinary skill, that the nucleic acid sequence of the present invention can be incorporated into anyone of numerous established kit formats which are well known in the art.

The term "vector" is commonly known in the art and defines a plasmid DNA, phage DNA, viral DNA and the like, which can serve as a DNA vehicle into which DNA of the present invention can be cloned. Numerous types of vectors exist and are well known in the art.

The term "allele" defines an alternative form of a gene which occupies a given locus on a chromosome.

As commonly known, a "mutation" is a detectable change in the genetic material which can be transmitted to a daughter cell. As well known, a mutation can be, for example, a detectable change in one or more deoxyribonucleotide. For example, nucleotides can be added, deleted, substituted for, inverted, or transposed to a new position. Spontaneous mutations and experimentally induced mutations exist. The result of a mutations of nucleic acid molecule is a mutant nucleic acid molecule. A mutant polypeptide can be encoded from this mutant nucleic acid molecule.

As used herein, the term "purified" refers to a molecule having been separated from a cellular component. Thus, for example, a "purified protein" has been purified to a level not found in nature. A "substantially pure" molecule is a molecule that is lacking in substantially all other cellular components. In accordance with preferred embodiments of the present invention, the term "purified" is used to relate to a non-duplexed nucleic acid molecule having been separated from a duplexed nucleic acid molecule, or vice versa.

The present invention also relates to a kit comprising an oligonucleotide or kit thereof of the present invention specifically designed to target (or "flag") a chosen mRNA or mRNAs. For example, a compartmentalized kit in accordance with the present invention includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allow the efficient transfer of reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated and the agents

or solutions of each container can be added in a quantitative fashion from one compartment to another. In a particular embodiment, such containers will include for example, a container which will accept the test sample (e.g. mRNA), a container which contains the oligonucleotides used in the assay and containers which contain reagents used to separate oligo-mRNA duplexes from non-duplexed mRNA. In another embodiment, such containers will include, for example, a container which will accept the test sample (e.g. mRNA), a container which contains the oligonucleotides used in the assay (to "flag" a chosen mRNA/mRNAs), RNAse H (to degrade duplexes), containers which contain oligo d(T) and reverse transcriptase (to synthesize cDNA from the treated mRNA sample), containers which contain reagents enabling the selection of full length cDNAs.

BRIEF DESCRIPTION OF THE DRAWINGS

Having thus generally described the invention, reference will now be made to the accompanying drawings, showing by way of illustration a preferred embodiment thereof, and in which:

Figure 1 shows an example of the steps involved in generating cDNA libraries from mRNA. Although a number of strategies can be used for cDNA library generation, of which only one is shown above, all libraries require as a first step, a primer which Reverse Transcriptase (RT) can prime from. In general, representation of mRNA abundance is maintained overall during construction, such that abundant mRNA species end up-being the most abundant cDNAs in the libraries, this is an impediment to novel gene discovery since random isolation of cDNAs from libraries will often lead to re-isolation of more abundantly expressed genes.

Figure 2 is a schematic diagram illustrating the specificity of cleavage of targeted RNA/DNA duplex by RNAse H.

Figure 3 is a schematic outline which shows an incorporation of a subtraction of cDNAs by suppression of cDNA synthesis with an affinity

selection step to remove cDNAs which correspond to RNAse H-targeted mRNAs using affinity selection of mRNA/cDNA duplexes with 5' m7G structures (5,7). The designation of mRNAs as species I-III relate to their classification with respect to their frequency of distribution (see text for further details).

Figure 4 shows an embodiment of the subtraction by suppression of synthesis according to one aspect of the present invention using the WT1 and D75 mRNAs. **A)** is a schematic diagram illustrating WT1 and D75 mRNAs, showing the sizes of the expected cDNA products resulting from priming with oligo d(T). For WT1 one product of ~1500 nucleotides (nts) is expected, whereas for D75 two products are expected, one resulting from priming from the poly (A) tail (~2300 nts) and one resulting from internal priming from an A-rich strech (2000 nts). The region on the WT1 targeted by the oligonucleotides is shown and is expected to generate a cDNA product of ~800 nts. **B)** shows the sequence of oligonucleotides used in the targeting reactions, it must be noted that for oligos having the nucleic acid sequences defined in SEQ ID NOS:2 and 3, the ribose of the last residue is in the L configuration and this is denoted by the symbol "L". **C)** shows complementary DNA products (cDNAs) from RNA templates targeted for subtraction by RNAse H. The nature of the templates ae illustrated above the panel. The presence or absence of RNAse H during pretreatment is also indicated above the panel. When targeting, oligonucleotides were pre-incubated with the test transcripts, this is indicatedon the bottom of the panel. The arrow indicates the cDNA product produced from 3' truncated WT1 template generated by RNAse H digestion of WT1.

Other objects, advantages and features of the present invention will become more apparent upon reading of the following non-restrictive description of preferred embodiments with reference to the accompanying drawing which is exemplary and should not be interpreted as limiting the scope of the present invention.

IDT Exhibit 1020.8091

DESCRIPTION OF THE PREFERRED EMBODIMENT

In accordance with one preferred embodiment of the present invention, an approach in which suppression of the synthesis of targeted transcripts occurs during the first strand reaction (see Fig. 1 for summary of first strand synthesis) is favored. Such an embodiment is based on the fact that RNAse H is an enzyme which specifically targets and degrades the RNA moiety of a RNA/DNA duplex. Therefore, annealing of a complementary DNA oligonucleotide to a given RNA, followed by treatment with RNAse H, results in a specific cleavage of only the target RNA at the site of hybridization (Fig. 2). This has the effect of physically separating the RNA targets into two portions, one half containing the 3' end, and another half containing the 5' m7G cap structure. As outlined in Fig. 3, if cDNAs are generated from these twoportions, utilizing for example reverse transcriptase and an oligo d(T) primer, a cDNA molecule originating from the targeted mRNA (mRNA species I; in Figs. 2 and 3) would extend from the poly (A) tail to the site of cleavage. However, a cDNA molecule originating from mRNA species II (which is not targeted by an oligonucleotide) would extend from the poly (A) tail to the m7G cap structure at the 5' end of the molecule. Following cDNA synthesis (Step II in Fig. 3), there are essentially two types of RNA/DNA hybrids in the mixture - one in which the RNA moiety of the hybrid contains a m7G cap structure at the 5' end, and one in which the RNA moiety of the duplex contains a 5' phosphate (i.e. truncated cDNAs). The products of this reaction can then be separated using a 5' affinity capture method specific for mRNA having m7G ends (5). This affinity procedure utilizes an immobilized protein (eIF-4E, a 24kDa cap-binding protein) which specifically recognizes m7G groups to isolate mRNA with m7G ends. When the mixture of reaction products from Step II is subjected to 5' m7G selection using eIF-4E, cDNAs whose mRNAs have not been targeted for degradation by RNAse H will be retained (since these contain a cap structure), whereas cDNAs whose mRNAs have been targeted for degradation will not be retained (since the RNA moiety of the incomplete cDNAs have a 5' phosphate group) (Fig 3). Since the

purified material from step III can now be used for second strand synthesis (see Fig. 1), the instant procedure eliminates cDNAs corresponding to the specific RNAse H targeted transcript, from the resulting cDNA library. By hybridizing the mRNA to be used as a template for cDNA library construction with a cocktail of oligonucleotides complementary to known targets of abundant transcripts (non-limiting examples include α-tubulin, elongation factor α1 and α-actin), the cDNA synthesis of these known abundant genes can be prematurely terminated during first strand synthesis. These truncated cDNAs are then removed following 5' affinity CAPture™ with GST/meIF-4E (5) (Step III in Fig. 3). The suppression procedure can even be made isoform-specific by directing the blocking (e.g. targeting of specific motifs or differentially spliced exons) oligonucleotide to the unique region of a particular transcript.

The present invention is illustrated in further detail by the following non-limiting example.

EXAMPLE 1

Validation of one embodiment of the method of subtracting cDNAs using WT1 or D75 mRNAs.

In the experiment presented in Fig. 4, two mRNA species, WT1 and D75, were pretreated with 4 complementary DNA oligonucleotides specific to WT1 (I-IV) and one control oligonucleotide (SP6) which does not hybridize to either WT1 or D75 (Fig. 4B). Following hybridization of the oligonucleotides to WT1 or D75, the RNAs were incubated with RNAse H. Following RNAseH treatment, cDNA reactions were performed with MMLV Superscript II™ and oligo d(T) (Fig. 4C). As shown in lane 1, a cDNA product originating from WT1 has the expected size of 1,500 nts. As well, cDNA synthesized from D75 mRNA generates products of 2,000 and 2,300 nts, due to priming from the poly (A) tail and mispriming from an internal A-rich tract by the oligo d(T) primer (lane 2). Pretreatment of either WT1 (lane 3) or D75 (lane 9) with RNAse H, in the absence of a targeting DNA oligonucleotide does not

change the profile of the cDNA products (this shows the duplex-dependent activity of RNAseH). Pretreatment of WT1 with 4 different complementary oligonucleotides (shown in Fig. 4B) and RNAse H, produced mainly two effects - a novel cDNA product corresponding to ~800 nts appears (lanes 4-7) and there is an apparent reduction in the amount of full-length WT1 cDNA generated. In this experiment, 4 oligonucleotides containing different 3' ends were utilized. In oligo #1 (SEQ ID NO:1), the last nucleotide is a cytosine residue, which is not complementary to the targeted WT1 sequence (containing a terminal cytosine residue instead of a guanosine residue). In oligo # 2 (SEQ ID NO:2), the ribose of the last nucleotide is in the L configuration instead of the D configuration. In oligo #3 (SEQ ID NO:3), the ribose of the last nucleotide is in the Lconfiguration and the linkage of the nucleotide to the rest of the oligo isinverted to give a 3'-3' linkage. Oligo # 4 (SEQ ID NO:4) is perfectly complementary to the WT1 target sequence.

Oligonucleotides with different 3' ends were used in order to show a) that modified nucleotides can be used in accordance with one embodiment of the present invention and, b) to determine if carryover of the targeting primer to the RT reaction would result in cDNA molecules priming not only from the oligo d(T) primer, but also from the targeting oligossince oligos #2 and #3 cannot act as primers for the reverse transcriptase (RT) enzyme.

Figure 4 shows that indeed, oligonucleotides with different 3' ends can be used to target a chosen mRNA for degradation. In addition, since no novel cDNA products specific to oligos #1 and 4 (which can prime cDNA synthesis) are observed in Figure 4 (compare lanes 4 and 7 to lanes5 and 6), it appears no carryover of the targeting primer to the RT reaction enables new cDNA primings.

Suppression of cDNA synthesis is specific to the WT1 oligos which target a known region of the WT1 gene since an SP6 oligonucleotide (SEQ ID NO:5), which is not complementary to any region of WT1, did notresult in suppression of cDNA synthesis (compare lane 8 to lanes 4-7). Furthermore,

the results presented herein are specific to the WT1 transcript since targeting D75 with these oligos did not alter the cDNA profile obtained with this transcript (lanes 10-14).

Taken together, these results demonstrate the feasibilibity of blocking RT processivity by first degrading the mRNA template. In addition, it validates the duplex-dependent method of customizing an mRNA population.

Although the present invention has been described hereinabove by way of preferred embodiments thereof, it can be modified, without departing from the spirit and nature of the subject invention as defined in the appended claims.

REFERENCES.

1. Davidson et al., 1979, Science 204:1052-1059.
2. Bishop et al., 1974, Nature 250:199-204.
3. Bento Soares et al., 1994, Proc. Natl. Acad. Sci. USA 91:9228-9232.
4. de Fatima Bonaldo et al., 1996, Genome Res. 6:791-806.
5. Edery et al., 1995, Mol. Cell. Biol. 15:3363-3371.
6. Borkow et al., 1997, Int. J. Biochem. Cell. Biol. 29:1285-1295.
7. Edery et al., USP 5,219,989.
8. Hall et al., 1996, Nucl. Acids Res. 24:3522-3526.
9. de Magistris et al., 1988, Nucleic Acids Res. 16:3141-56.
10. Carninci et al., 1997, DNA Res. 4:61-6.
11. Schwer et al., 1987, Cell 50:163-9.

WHAT IS CLAIMED IS:

1. A method of customizing an mRNA population comprising:

a) selecting an oligonucleotide which is complementary to an mRNA molecule in an mRNA sample;

b) incubating said oligonucleotide with said mRNA sample to obtain an mRNA oligonucleotide mixture under conditions which enable a formation of a duplex between said oligonucleotide and said mRNA molecule; and

c) submitting said mRNA-oligonucleotide mixture to a purification step which enables a separation of duplexed mRNA from non-duplexed mRNA,

thereby yielding a customized mRNA population.

2. A method for suppressing the synthesis of target transcripts during first strand cDNA synthesis, comprising:

a) selecting an oligonucleotide which is complementary to an mRNA molecule in an mRNA sample;

b) incubating said oligonucleotide with said mRNA sample to obtain an mRNA oligonucleotide mixture under conditions which enable a formation of a duplex between said oligonucleotide and said mRNA molecule; and

c) submitting said mRNA-oligonucleotide mixture to a duplex-specific RNAse which specifically targets the RNA moiety of said duplex; and

d) performing first strand cDNA synthesis on the RNAse-treated duplexes of c),

thereby suppressing the synthesis of RNAse-cleaved duplexes.

IDT Exhibit 1020.8097

3. The method of claim 2, wherein said RNAse is RNAse H.

4. An improved method of cDNA synthesis, the improvement consisting into a use of a method according to claim 1 or 2.

5. A method of subtracting cDNAs by suppressing their synthesis comprising the steps of :

a) annealing a set of complementary DNA oligonucleotides to a given mRNA sample under conditions enabling a formation of mRNA-oligonucleotide duplexes; and

b) performing first strand cDNA synthesis on non-duplexed mRNAs.

6. The method of claim 5, further comprising the following step:

a) a specific elimination of said mRNA-oligonucleotides duplexes from said mRNA sample.

7. The method of claim 5, further comprising the following step:

a) a specific cleavage and/or degradation of the mRNA of said mRNA-oligonucleotides duplexes from said sample.

8. A method of targeting a chosen mRNA or mRNA population for degradation using oligos which are specific for said chosen mRNA or mRNA population to generate duplexes of oligos and mRNA, submitting said duplex to a duplex-directed degradation or cleavage which leaves non-duplexed mRNA intact and purifying said non-duplexed mRNAs.

IDT Exhibit 1020.8098

9. The method of claim 2, wherein first strand synthesis is performed on RNAse-treated duplexes of c).

10. A method of customizing an mRNA population comprising:

a) selecting an oligonucleotide which is complementary to an mRNA molecule in an mRNA sample;

b) incubating said oligonucleotide with said mRNA sample to obtain an mRNA oligonucleotide mixture under conditions which enable a formation of a duplex between said oligonucleotide and said mRNA molecule; and

c) submitting said mRNA-oligonucleotide mixture to a duplex-dependent cleavage and/or degradation,

thereby yielding a customized mRNA population.

11. A method of customizing an RNA population comprising

a) selecting an oligonucleotide which is complementary to an RNA molecule in an RNA sample;

b) incubating said oligonucleotide with said RNA sample to obtain an RNA oligonucleotide mixture under conditions which enable a formation of a duplex between said oligonucleotide and said RNA molecule; and

c) submitting said RNA-oligonucleotide mixture to a purification step which enables a separation of duplexed RNA from non-duplexed RNA,

thereby yielding a customized RNA population.

12. A method of customizing an RNA population comprising

a) selecting an oligonucleotide which is complementary to an RNA molecule in an RNA sample;

IDT Exhibit 1020.8099

b) incubating said oligonucleotide with said RNA sample to obtain an RNA oligonucleotide mixture under conditions which enable a formation of a duplex between said oligonucleotide and said RNA molecule; and

c) submitting said RNA-oligonucleotide mixture to a duplex-dependent cleavage and/or degradation;

thereby yielding a customized RNA population.

IDT Exhibit 1020.8100

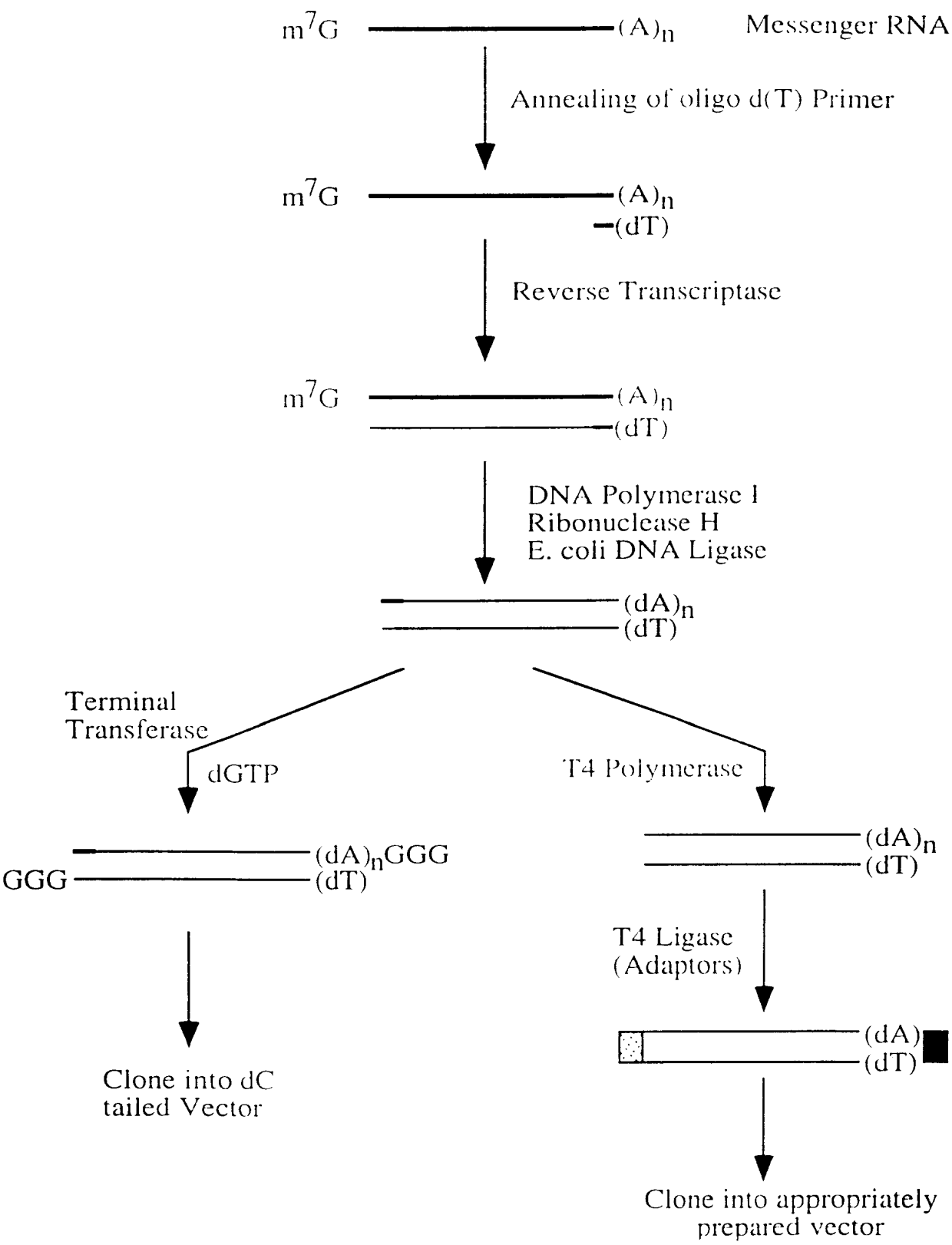


Fig. 1

IDT Exhibit 1020.8101

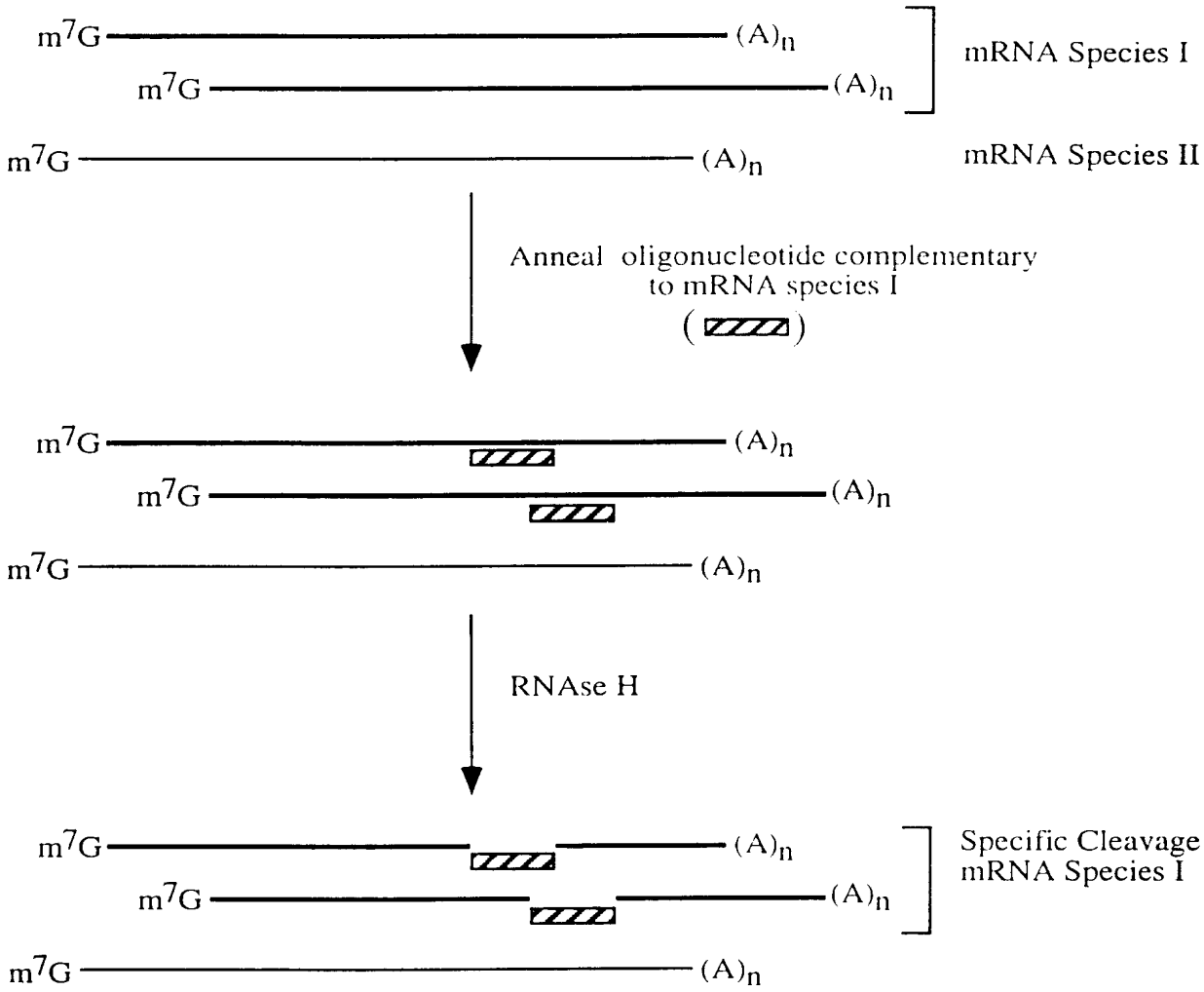


Fig. 2

IDT Exhibit 1020.8102

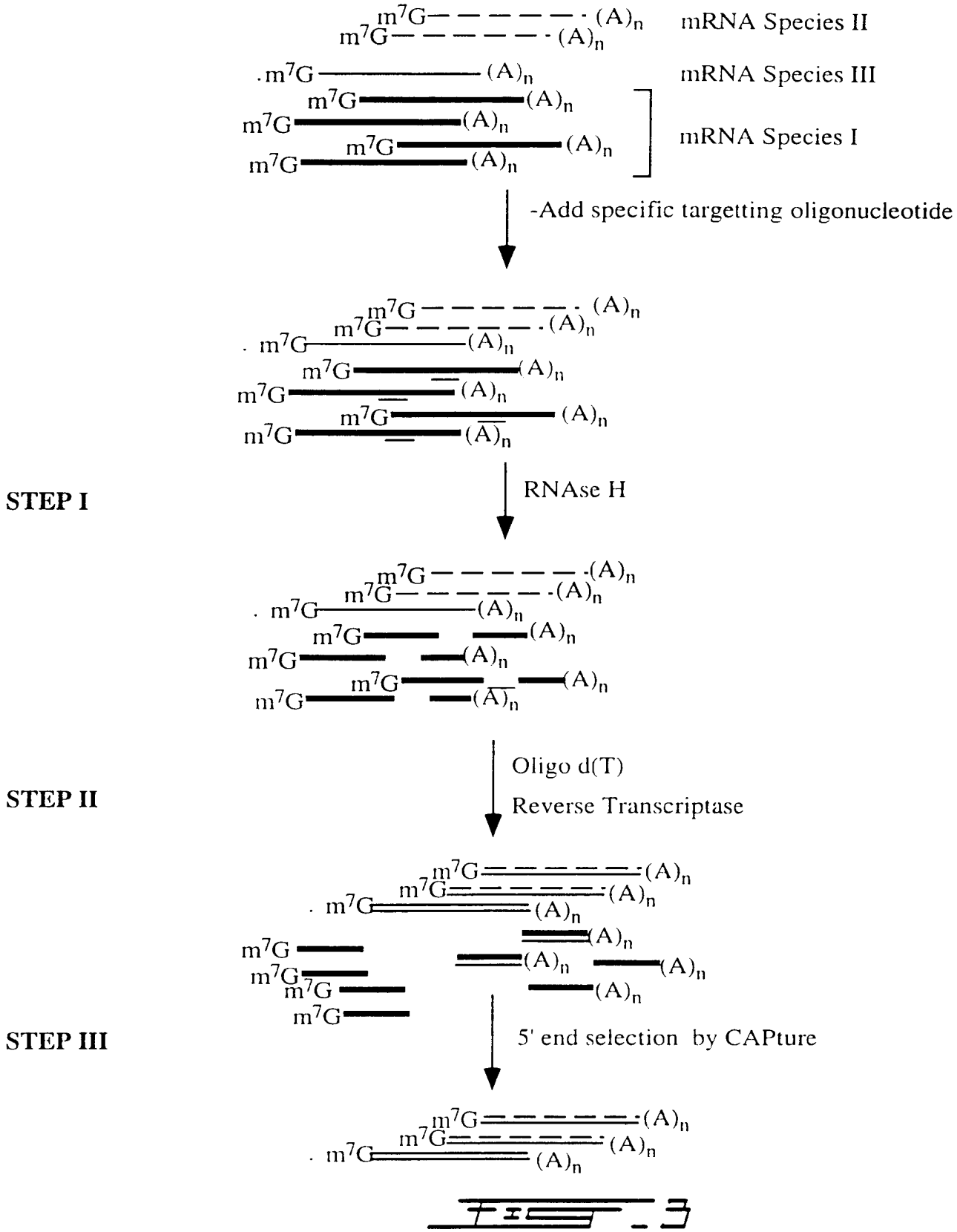
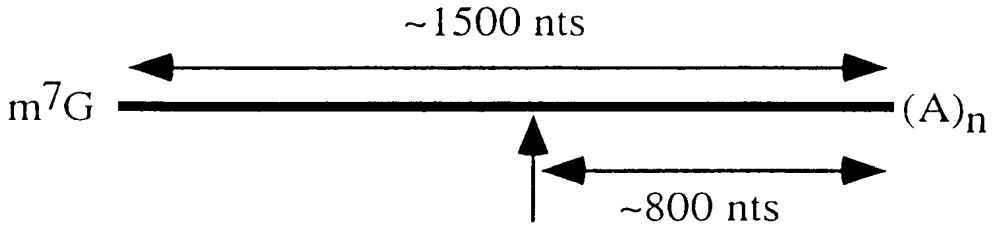


Fig. 3

IDT Exhibit 1020.8103



Area targetted by oligos

D75 mRNA

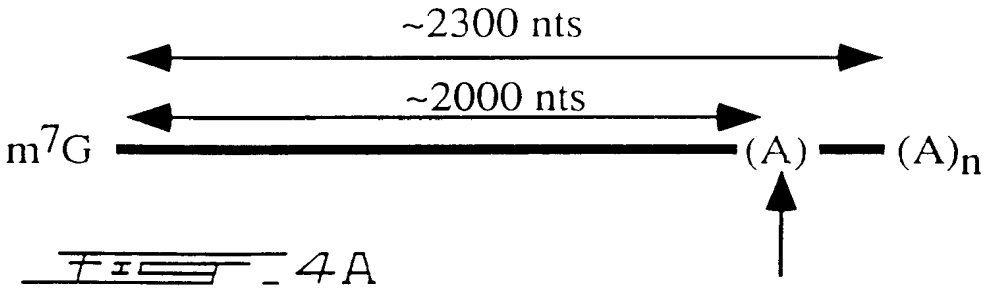


Fig 4A

Internal A-rich tract

Oligo #I	5' GGCAGCCATACACCGGA-C 3'
Oligo #II	5' GGCAGCCATACACCGGA-(L)-C 3'
Oligo #III	5' GGCAGCCATACACCGG 3'-3'-(L)-C 5'
Oligo #IV	5' GGCAGCCATACACCGGA 3'
SP6 Oligo	5' GCATTTAGGTGACACTATAG 3'

Fig 4B

IDT Exhibit 1020.8104

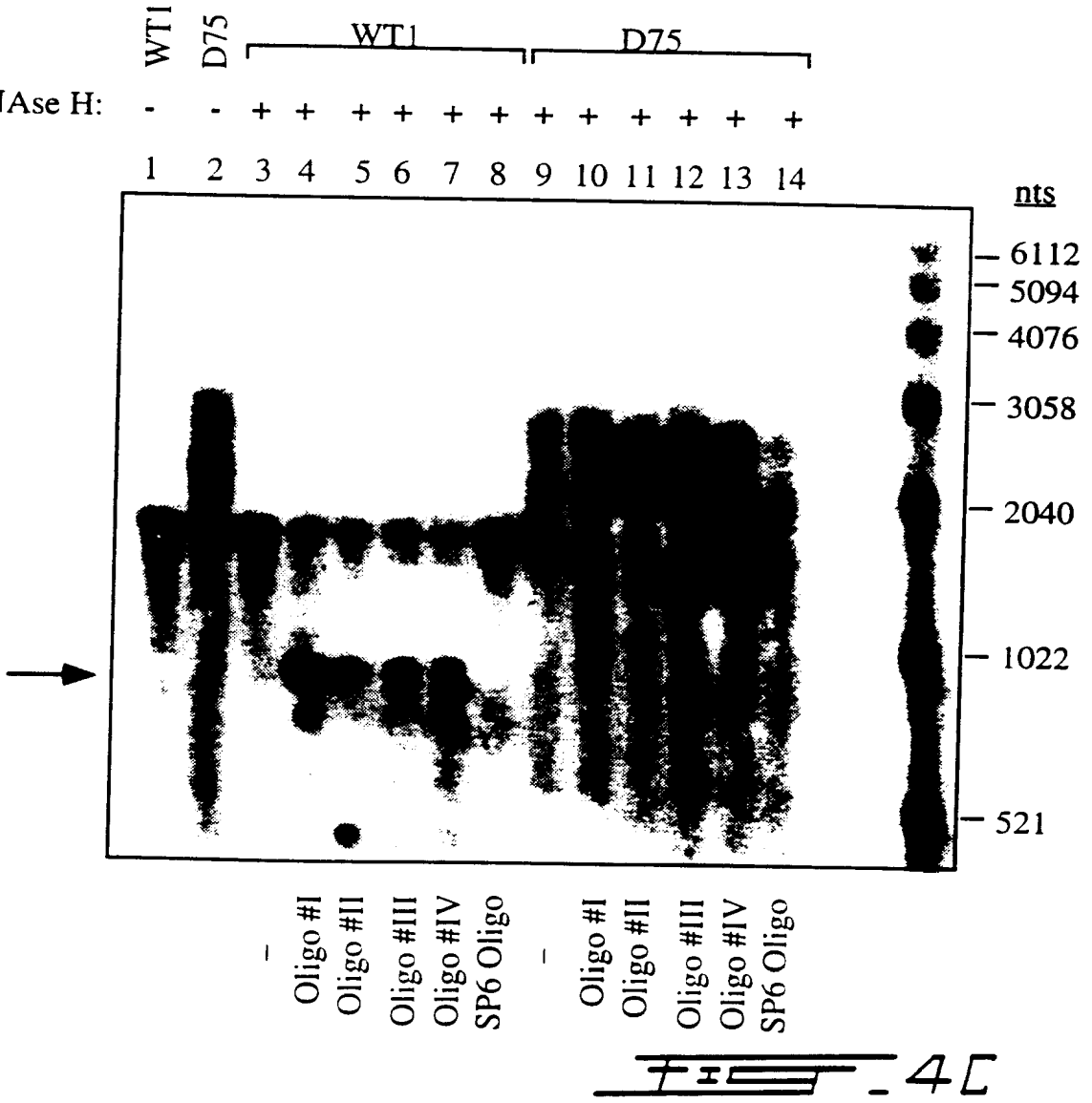


Fig. 4C

IDT Exhibit 1020.8105

SEQUENCE LISTING:

SEQ ID NO:1	5'GGCAGCCATACACCGGA-C$^{3'}$
SEQ ID NO:2	5'GGCAGCCATACACCGGA-(L)-C$^{3'}$
SEQ ID NO:3	5'GGCAGCCATACACCGG$^{3'-3'}$-(L)-C$^{5'}$
SEQ ID NO:4	5'GGCAGCCATACACCGGA$^{3'}$
SEQ ID NO:5	5'GCATTTAGGTGACACTATAG$^{3'}$

INTERNATIONAL SEARCH REPORT

Int tional Application No
PCT/CA 00/00586

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/10 C12Q1/68

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C12Q

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

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
WPI Data, PAJ, CAB Data, STRAND, EPO-Internal, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MACDONALD CLINTON ET AL: "RNase H-oligonucleotide-directed mRNA purification (ROMP) of apoII mRNA." NUCLEIC ACIDS RESEARCH, vol. 21, no. 3, 1993, pages 765-766, XP002147535 ISSN: 0305-1048 the whole document --- -/--	1,11

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

[X] Patent family members are listed in annex.

° Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
15 September 2000	29/09/2000

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

1

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

IDT Exhibit 1020.8107

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	OSTERMEIER CHRISTIAN ET AL: "Improved cloning of antibody variable regions from hybridomas by an antisense-directed RNase H digestion of the P3-X63-Ag8.653 derived pseudogene mRNA." NUCLEIC ACIDS RESEARCH, vol. 24, no. 10, 1996, pages 1979-1980, XP002147536 ISSN: 0305-1048 the whole document	2-4, 8-10,12
A	EP 0 373 914 A (UNIV MCGILL) 20 June 1990 (1990-06-20) the whole document	
A	WO 89 05358 A (UNIV IOWA RES FOUND) 15 June 1989 (1989-06-15) the whole document	
A	EDERY I ET AL: "AN EFFICIENT STRATEGY TO ISOLATE FULL-LENGTH CDNAS BASED ON AN MRNA CAP RETENTION PROCEDURE (CAPTURE)" MOLECULAR AND CELLULAR BIOLOGY,US,WASHINGTON, DC, vol. 15, June 1995 (1995-06), pages 3363-3371, XP002915963 ISSN: 0270-7306 cited in the application the whole document	
A	TOSHIFUMI HATTA ET AL: "MECHANISMS OF THE INHIBITION OF REVERSE TRANSCRIPTION BY UNMODIFIEDAND MODIFIED ANTISENSE OLIGONUCLEOTIDES" FEBS LETTERS,NL,ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, vol. 330, no. 2, 1 September 1993 (1993-09-01), pages 161-164, XP000574999 ISSN: 0014-5793 the whole document	
A	BOIZIAU C ET AL: "MECHANISMS OF THE INHIBITION OF REVERSE TRANSCRIPTION BY ANTISENSE OLIGONUCLEOTIDES" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES, vol. 89, no. 2, 1992, pages 768-772, XP002147537 1992 ISSN: 0027-8424 the whole document	

-/--

1

IDT Exhibit 1020.8108

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	OYAMA F ET AL: "INTRINSIC PROPERTIES OF REVERSE TRANSCRIPTASE IN REVERSE TRANSCRIPTION ASSOCIATED RNASE H IS ESSENTIALLY REGARDED AS AN ENDONUCLEASE" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 264, no. 31, 1989, pages 18808-18817, XP002147538 ISSN: 0021-9258 the whole document ---	
A	US 5 891 637 A (RUPPERT SIEGFRIED J W) 6 April 1999 (1999-04-06) the whole document ---	
A	WO 94 03472 A (GEN PROBE INC) 17 February 1994 (1994-02-17) the whole document ---	
A	BONALDO FATIMA DE M ET AL: "NORMALIZATION AND SUBTRACTION: TWO APPROACHES TO FACILITATE GENE DISCOVERY" GENOME RESEARCH,US,COLD SPRING HARBOR LABORATORY PRESS, vol. 6, no. 9, 1 September 1996 (1996-09-01), pages 791-806, XP002039972 ISSN: 1088-9051 the whole document -----	

1

IDT Exhibit 1020.8109

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int. tional Application No
PCT/CA 00/00586

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0373914 A	20-06-1990	CA 2005252 A	13-06-1990
		DE 68916769 D	18-08-1994
		DE 68916769 T	10-11-1994
		US 5219989 A	15-06-1993
WO 8905358 A	15-06-1989	AT 151467 T	15-04-1997
		AU 620364 B	20-02-1992
		AU 2782989 A	05-07-1989
		CA 1339935 A	30-06-1998
		DE 3855864 D	15-05-1997
		DE 3855864 T	25-09-1997
		EP 0348458 A	03-01-1990
		JP 2502516 T	16-08-1990
		JP 3019994 B	15-03-2000
		US 5403711 A	04-04-1995
		US 5962425 A	05-10-1999
		US 5491133 A	13-02-1996
US 5891637 A	06-04-1999	NONE	
WO 9403472 A	17-02-1994	AU 670116 B	04-07-1996
		AU 4792093 A	03-03-1994
		AU 700253 B	24-12-1998
		AU 6802796 A	19-12-1996
		CA 2141430 A	17-02-1994
		EP 0587298 A	16-03-1994
		EP 0978570 A	09-02-2000
		JP 7509368 T	19-10-1995
		NO 950381 A	03-04-1995
		US 5766849 A	16-06-1998
		US 5908744 A	01-06-1999

IDT Exhibit 1020.8110

PTO/SB/08a (01-22)
Approved for use through 05/31/2024. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

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

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	17381888
Filing Date	2021-07-21
First Named Inventor	Douglas A. Amorese
Art Unit	N/A
Examiner Name	Not Yet Assigned
Attorney Docket Number	NUGN-001/04US 34079/550

U.S.PATENTS Remove

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Patent citation information please click the Add button. Add

U.S.PATENT APPLICATION PUBLICATIONS Remove

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Published Application citation information please click the Add button. Add

FOREIGN PATENT DOCUMENTS Remove

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2]	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1	2013/190441	WO	A2	2013-12-27	Univ Hong Kong Chinese		
	2	2013/191775	WO	A2	2013-12-27	Nugen Technologies, Inc		
	3	2014/039556	WO	A1	2014-03-13	Guardant Health, Inc		

IDT Exhibit 1020.8111

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17381888
	Filing Date	2021-07-21
	First Named Inventor	Douglas A. Amorese
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	NUGN-001/04US 34079/550

	4	2014/082032	WO	A1	2014-05-30	Univ Toledo		
	5	2014/144092	WO	A1	2014-09-18	Nugen Technologies, Inc		
	6	2014/150931	WO	A1	2014-09-25	Nugen Technologies, Inc		
	7	2015/031691	WO	A1	2015-03-05	Cellular Research, Inc		
	8	2015/073711	WO	A1	2015-05-21	Nugen Technologies, Inc		
	9	2015/131107	WO	A1	2015-09-03	Nugen Technologies, Inc		
	10	89/09284	WO	A1	1989-10-05	Univ Iowa Res Found		
	11	92/07951	WO	A1	1992-05-14	Kow, Yoke		
	12	93/18052	WO	A1	1993-09-16	Isis Pharmaceuticals, Inc		
	13	94/16090	WO	A1	1994-07-21	Molecular Tool, Inc		
	14	96/40998	WO	A1	1996-12-19	Pioneer Hi-bred International, Inc		

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17381888
	Filing Date	2021-07-21
	First Named Inventor	Douglas A. Amorese
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	NUGN-001/04US 34079/550

	15	97/12061	WO	A1	1997-04-03	Epicentre Technologies Corp		
	16	97/25416	WO	A2	1997-07-17	Nyfotek As		
	17	98/044151	WO	A1	1998-10-08	Glaxo Group Ltd		
	18	98/06736	WO	A1	1998-02-19	Life Technologies, Inc		
	19	98/38296	WO	A1	1998-09-03	Gesher Israel Advanced Biotecs		
	20	99/10540	WO	A1	1999-03-04	Lopez, Osvaldo		
	21	99/11819	WO	A1	1999-03-11	Perkin Elmer Corp		
	22	99/42618	WO	A1	1999-08-26	Dade Behring Inc		

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

NON-PATENT LITERATURE DOCUMENTS Remove

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]
	1		

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

IDT Exhibit 1020.8113

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	17381888
Filing Date	2021-07-21
First Named Inventor	Douglas A. Amorese
Art Unit	N/A
Examiner Name	Not Yet Assigned
Attorney Docket Number	NUGN-001/04US 34079/550

EXAMINER SIGNATURE			
Examiner Signature		Date Considered	

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

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

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	17381888
Filing Date	2021-07-21
First Named Inventor	Douglas A. Amorese
Art Unit	N/A
Examiner Name	Not Yet Assigned
Attorney Docket Number	NUGN-001/04US 34079/550

CERTIFICATION STATEMENT

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

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

OR

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

See attached certification statement.

The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

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

Signature	/Thomas C. Meyers/	Date (YYYY-MM-DD)	2022-08-04
Name/Print	Thomas C. Meyers	Registration Number	36,989

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

Privacy Act Statement

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

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

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

IDT Exhibit 1020.8116

Espacenet

Bibliographic data: WO0052191 (A1) — 2000-09-08

cDNA SYNTHESIS IMPROVEMENTS

Inventor(s): LI WU BO; JESSEE JOEL A; GRUBER CHRISTIAN E ± (LI, WU, BO, ; JESSEE, JOEL, A, ; GRUBER, CHRISTIAN, E)

Applicant(s): LIFE TECHNOLOGIES INC [US] ± (LIFE TECHNOLOGIES, INC)

Classification: **- international:** ***C12N15/10;*** (IPC1-7): C12P19/34
- cooperative: **C07K16/40**; **C12N15/1096**; **C12P19/34**

Application number: WO2000US05138 20000301 Global Dossier

Priority number (s): US19990122395P 19990302

Also published as: AU3384900 (A) AU779060 (B2) CA2364483 (A1) EP1165823 (A1) EP1165823 (A4) more

Abstract of WO0052191 (A1)

The present invention generally relates to methods of making cDNA molecules and cDNA libraries. The invention also relates to cDNA molecules and cDNA libraries produced according to these methods, as well as to vectors and host cells containing such cDNA molecules and libraries. The invention also relates to kits for making the cDNA molecules and libraries of the invention.

IDT Exhibit 1020.8117

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : **C12P 19/34** — **A1**

(11) International Publication Number: **WO 00/52191**

(43) International Publication Date: 8 September 2000 (08.09.00)

(21) International Application Number: PCT/US00/05138

(22) International Filing Date: 1 March 2000 (01.03.00)

(30) Priority Data:
60/122,395 2 March 1999 (02.03.99) US

(71) Applicant: LIFE TECHNOLOGIES, INC. [US/US]; 9800 Medical Center Drive, Rockville, MA 20850 (US).

(72) Inventors: LI, Wu, Bo; 11909 Riding Loop Terrace, North Potomac, MD 20878 (US). JESSEE, Joel, A.; 4139 Old National Pike, Mount Airy, MD 21771 (US). GRUBER, Christian, E.; 2105 Charlton House Way, Frederick, MD 21701 (US).

(74) Agents: ESMOND, Robert, W. et al.; Sterne, Kessler, Goldstein & Fox P.L.L.C., Suite 600, 1100 New York Avenue, Washington, DC 20005–3934 (US).

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

(54) Title: cDNA SYNTHESIS IMPROVEMENTS

(57) Abstract

The present invention generally relates to methods of making cDNA molecules and cDNA libraries. The invention also relates to cDNA molecules and cDNA libraries produced according to these methods, as well as to vectors and host cells containing such cDNA molecules and libraries. The invention also relates to kits for making the cDNA molecules and libraries of the invention.

IDT Exhibit 1020.8118

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 1020.8119

cDNA Synthesis Improvements

Background of the Invention

The present invention relates to the field of molecular and cellular biology. The invention generally relates to methods of synthesizing cDNA. More specifically, the present invention relates to methods of increasing the average cDNA insert size and more particularly, to increasing the percentage of full-length cDNA present within cDNA libraries. Thus, the present invention provides improved cDNA libraries useful in gene discovery.

In examining the structure and physiology of an organism, tissue or cell, it is often desirable to determine its genetic content. The genetic framework of an organism is encoded in the double-stranded sequence of nucleotide bases in the deoxyribonucleic acid (DNA) which is contained in the somatic and germ cells of the organism. The genetic content of a particular segment of DNA, or gene, is only manifested upon production of the protein which the gene encodes. In order to produce a protein, a complementary copy of one strand of the DNA double helix (the "coding" strand) is produced by polymerase enzymes, resulting in a specific sequence of ribonucleic acid (RNA). This particular type of RNA, since it contains the genetic message from the DNA for production of a protein, is called messenger RNA (mRNA).

Within a given cell, tissue or organism, there exist many mRNA species, each encoding a separate and specific protein. This fact provides a powerful tool to investigators interested in studying genetic expression in a

tissue or cell. mRNA molecules may be isolated and further manipulated by various molecular biological techniques, thereby allowing the elucidation of the full functional genetic content of a cell, tissue or organism.

A common approach to the study of gene expression is the production of complementary DNA (cDNA) clones. In this technique, the mRNA molecules from an organism are isolated from an extract of the cells or tissues of the organism. This isolation often employs solid chromatography matrices, such as cellulose or agarose, to which oligomers of thymidine (T) have been complexed. Since the 3' termini on most eukaryotic mRNA molecules contain a string of adenosine (A) bases, and since A binds to T, the mRNA molecules can be rapidly purified from other molecules and substances in the tissue or cell extract. From these purified mRNA molecules, cDNA copies may be made using the enzyme reverse transcriptase (RT) or DNA polymerases having RT activity, which results in the production of single-stranded cDNA molecules. The single-stranded cDNAs may then be converted into a complete double-stranded DNA copy (*i.e.*, a double-stranded cDNA) of the original mRNA (and thus of the original double-stranded DNA sequence, encoding this mRNA, contained in the genome of the organism) by the action of a DNA polymerase. The protein-specific double-stranded cDNAs can then be inserted into a vector, which is then introduced into a host bacterial, yeast, animal or plant cell, a process referred to as transformation or transfection. The host cells are then grown in culture media, resulting in a population of host cells containing (or in many cases, expressing) the gene of interest or portions of the gene of interest.

This entire process, from isolation of mRNA to insertion of the cDNA into a vector (e.g., plasmid, viral vector, cosmid, etc.) to growth of host cell populations containing the isolated gene or gene portions, is termed "cDNA cloning." If cDNAs are prepared from a number of different mRNAs, the resulting set of cDNAs is called a "cDNA library," an appropriate term since the set of cDNAs represents a "population" of genes or portions of genes

comprising the functional genetic information present in the source cell, tissue or organism. Genotypic analysis of these cDNA libraries can yield much information on the structure and function of the organisms from which they were derived.

The ability to increase the total amount of cDNA produced, and more particularly to produce a cDNA libraries having an increase in the average size of the cDNA molecules and/or to produce cDNA libraries having an increase in the percentage of full-length cDNA molecules would provide a significant advance in cDNA library construction. Specifically, such advances would greatly improve the probability of finding full-length genes of interest.

Ideally, synthesis of a cDNA molecule initiates at or near the 3' termini of the mRNA molecules. Priming of cDNA synthesis at the 3' termini at the poly A tail using an oligo(dT) primer ensures that the 3' message of the mRNAs will be represented in the cDNA molecules produced. Priming which occurs within the mRNA molecules (internal priming) results in synthesis of cDNA molecules which do not contain the full-length message for the genes of interest. That is, internal priming results in truncated cDNA molecules which contain only a portion of the gene or genes of interest. Typically, internal priming causes a loss of the 3' sequences from the message population. Thus, internal priming lowers the total amount of cDNA produced, decreases the average insert size of cDNA molecules for a cDNA library and/or decreases the percentage of full-length cDNA molecules in a given cDNA library. Sequencing analysis has indicated that many eukaryotic mRNAs have internal poly adenylation stretches which may serve as a priming site when an oligo(dT) primer is used for first strand cDNA synthesis with reverse transcriptase. Moreover, research has shown that some mRNAs can have as many as 16 internal priming sites (Lovett, M., *et al.*, The construction of full-length cDNA libraries by conventional methods and a novel double capture technique, University of Texas Southwestern Medical Center, Dallas, TX, presented at the 48th Annual Meeting held by The American Society of Human

Genetics, October 27-31, 1998, Denver, Colorado). Thus, internal priming of the primer to such internal poly A sequences may adversely affect cDNA synthesis.

The present invention alleviates, prevents, reduces or substantially reduces internal priming thereby providing improvements in cDNA and cDNA library construction. Accordingly, the present invention greatly facilitates gene discovery by providing cDNA libraries containing a greater percentage of full-length genes.

The present invention therefore relates to synthesizing a cDNA molecule or molecules from an mRNA template or population of mRNA templates under conditions sufficient to increase the total amount of cDNA produced, increase the length of the cDNA molecules produced, and/or increase the amount or percentage of full-length cDNA molecules produced. In accordance with the invention, any conditions which inhibit, prevent, reduce or substantially reduce internal priming may be used. Such conditions preferably include but are not limited to optimizing primer concentrations, optimizing reaction temperatures and/or optimizing primer length or specificity. Such result may also be accomplished in accordance with the invention by optimizing the reverse transcription reaction, preferably by inhibiting or preventing reverse transcription until optimum or desired reaction conditions are achieved.

Conventional methods for constructing cDNA libraries use a molar ratio of oligo(dT) primer/mRNA template of 15:1 for first strand cDNA synthesis. The use of such excess amounts of oligo(dT) primer allows internal priming of one or more primers to one or more of the mRNA templates in the reaction. According to a preferred aspect of the present invention, the amount of oligo(dT) primer is reduced for synthesis of first strand cDNA to inhibit, prevent, reduce or substantially reduce internal priming. Preferred molar ratios of primer to template range from about 12:1; 10:1; 9:1; 8:1; 7:1; 6:1; 5:1; 4:1; 3:1; 2:1; 1:1; 1:2; 1:3; 1:4; 1:5; 1:6; 1:7; 1:8; 1:9; 1:10 and 1:12.

Preferably, molar ratios of primer (e.g., oligo(dT)) to template (e.g., mRNA) range from about 5:1 to about 1:20, although lower molar ratios of primer to template may be used in accordance with the invention. Specifically, molar ratios of primer to template may be below about 1:10; 1:15; 1:20; 1:25; 1:50; 1:75; and 1:100. Preferably, ranges of molar ratios are below about 5:1; 4:1; 3:1; 2:1; 1:1; 1:2; 1:3; 1:4; and 1:5. Most preferably, ratios of primer to template range from about 10:1 to 1:10; 5:1 to 1:10; 4:1 to 1:10; 3:1 to 1:10; 2.5:1 to 1:10; 2:1 to 1:10; 1.5:1 to 1:10; and 1:1 to 1:10. The optimum ratios of primer to template may vary depending on the primer, mRNA, reverse transcription enzyme and reaction conditions (annealing temperature, buffering salts, etc.). The desired primer to template ratios can be readily determined by one skilled in the art.

In conventional methods of cDNA library construction, annealing or hybridizing primer to template is not carried out at a temperature which prevents, inhibits, reduces or substantially reduces internal priming. Typically, the mixture (e.g., mRNA and oligo(dT) primer) is chilled on ice after denaturation or heating. This process typically causes annealing or hybridization of the primer to internal sites. According to a preferred aspect of the present invention, the temperature during the annealing or hybridization between the primer and the template is maintained so that internal priming is inhibited, prevented, reduced or substantially reduced. In accordance with the invention, such a result is accomplished by carrying out primer annealing or hybridization at higher temperatures. Such conditions may also reduce the formation of mRNA secondary structures during cDNA synthesis. Preferably, temperatures for annealing or hybridizing primers to the templates range from about 10°C to about 90°C; more preferably about 10°C to about 80°C; still more preferably about 20°C to about 75°C; more preferably about 25°C to about 75°C; still more preferably about 30°C to about 65°C; still more preferably about 37°C to about 60°C; still more preferably about 40°C to about 60°C; still more preferably about 45°C to about 60°C; still more

preferably about 45°C to about 55°C; and most preferably about 45°C to about 65°C. The temperature used may vary depending on the type and amount of primer and template and depending on the temperature optimum of the reverse transcription enzyme. The optimum temperature or temperature ranges can be readily determined by one skilled in the art.

Conventional methods for cDNA synthesis typically requires the use of oligo(dT) primers of a particular length (12-18 bases or mer). Such primer length, however, lowers specificity of the primer thereby allowing internal priming. Thus, the invention also relates to increasing specificity of the primers to prevent, inhibit, reduce or substantially reduce internal priming. In a preferred aspect, primer specificity is increased by increasing the length of the primer. Thus, for cDNA synthesis, longer oligo(dT) primers may be used in accordance with the invention. Preferably, primer length ranges from about 20 to about 100 bases, about 20 to about 75 bases, about 20 to about 60 bases, and about 20 to about 50 bases; more preferably about 20 to about 45 bases; more preferably about 20 to about 40 bases; and most preferably about 25 to about 35 bases. In a preferred aspect, the length of the primers are greater than 19 bases; more preferably greater than about 20 bases; more preferably greater than about 25 bases; and still more preferably greater than about 30 bases. Such primer lengths refer to the length of the primers which anneal or hybridize to the template. Optimum length and content (nucleotide sequence) of the primers may vary depending on the type of template, the desired reaction conditions, and the reverse transcription enzyme. In accordance with the invention, additional sequences and/or modified nucleotides may be included in the primers of the invention. For example, additional sequences (which do not necessarily anneal or hybridize to the template) may be included in the primers of the invention to assist in cDNA synthesis including sequences comprising one or more restriction endonuclease sites, one or more derivative nucleotides (e.g., hapten containing nucleotides such as biotinylated nucleotides), and the like. The type and length of the primers used in

accordance with the invention can be readily determined by one or more skilled in the art.

Conventional cDNA synthesis methods do not control or vary activity of the reverse transcription enzyme to optimize the reverse transcription reaction. In accordance with the invention, the activity of the reverse transcriptase is preferably controlled to start synthesis at a desired time in the reaction. In a preferred aspect, reverse transcriptase activity is inhibited or prevented until optimum or desired reaction conditions are achieved. Such a result is accomplished in accordance with the invention by the use of inhibitors (such as antibodies or antibody fragments) which inhibit reverse transcriptase activity. Such reverse transcriptase inhibitors prevent or inhibit reverse transcriptase activity at low temperatures such that internal priming is prevented, inhibited, reduced or substantially reduced. In accordance with the invention, such inhibitors preferably prevent reverse transcriptase activity below 35°C, below 40°C, below 45°C, below 50°C, below 55°C, below 60°C, below 65°C, below 70°C, below 75°C, below 80°C, below 85°C and below 90°C. Depending on the thermostability of the enzyme having reverse transcriptase activity, the inhibitor may be designed to inhibit activity of the enzyme at a point at or near the temperature optimum for the enzyme of interest. Preferably, the inhibitor is inactivated at a temperature below or near the temperature optimum of the enzyme used, thereby allowing reverse transcription to take place. Thus, the invention generally relates to the use of reverse transcriptase inhibitors in cDNA synthesis. The type and amount of inhibitor may vary depending on the type and amount of reverse transcription enzyme and depending on the reaction conditions to be used. The type of inhibitor and conditions used with such inhibitor can be readily determined by one of ordinary skill in the art.

In accordance with the invention, any one or a combination of the above improvements to cDNA synthesis may be used. Using any one or a combination of these improvements provides for improved first strand cDNA

synthesis (e.g., more total cDNA, larger cDNA and/or more full-length cDNA). In accordance with the invention, the first strand cDNA molecules may be used as templates to make one or more double stranded nucleic acid molecules (e.g., double strand cDNA molecules) by incubating one or more of the first strand cDNA molecules produced by the methods of the invention under conditions sufficient to make one or more nucleic acid molecules complementary to all or a portion of the first strand cDNA molecules. Conditions for making double stranded nucleic acid molecules preferably include incubation with one or more components consisting of one or more DNA polymerases, one or more nucleotides, one or more buffering salts, and one or more primers. In another aspect of the invention, such conditions are modified to provide an increase in the total amount of double stranded cDNA produced, an increase in the length or size of the double stranded cDNA molecule produced, and/or an increase in percentage full-length double stranded cDNA molecule produced. Preferably, such conditions relate to optimization of ribonuclease (RNase) digestion after first strand cDNA synthesis. During first strand cDNA synthesis, if a full-length cDNA molecule complementary to the mRNA template is not made, a single stranded mRNA containing the cap structure will be present at the 5' end of the mRNA of the mRNA/cDNA hybrid. If a full-length cDNA is produced, a double stranded mRNA/cDNA hybrid is produced with no single stranded mRNA present. Preferably, such digestion conditions are optimized so that the single stranded mRNA of the mRNA/cDNA double stranded molecules formed during first strand cDNA synthesis is subject to RNase digestion. In this manner, cap structure from mRNA/cDNA hybrids which are not full-length are removed while full-length mRNA/cDNA hybrids will retain the cap structure. Thus, cap capture can be used to select for full-length molecules and select against molecules which are not full-length. In a preferred aspect, the conditions are such that the single stranded mRNA of the mRNA/cDNA hybrid is digested or degraded while the mRNA of the double stranded mRNA/cDNA hybrid is not

degraded or not substantially degraded. Thus, such RNase digestion is conducted under conditions such that second strand synthesis is not substantially adversely affected. That is, second strand synthesis in accordance with the invention produces larger double stranded cDNA molecules compared to conventional techniques. Conventional RNase I conditions typically range from 25 u/μg to 40 u/μg mRNA at 37°C and RNase A conditions typically are 1000 ng/μg mRNA at 37°C. Using conventional RNase digestion, the average size of double stranded cDNA molecules produced is about 200 bases. According to the present invention the average size of double stranded cDNA molecules produced is preferably greater than about 300 bases, greater than about 400 bases, greater than about 500 bases, greater than about 600 bases, greater than about 700 bases, greater than about 800 bases, greater than about 900 bases, greater than about 1 kilobase, greater than about 1.5 kilobases, and greater than about 2 kilobases. In one embodiment of the invention, the concentration of the ribonuclease, the type of ribonuclease and reaction conditions are optimized to improve double stranded cDNA synthesis in accordance with the invention. Preferred ribonucleases for use in ribonuclease digestions include ribonuclease A (RNase A) and/or ribonuclease I (RNase I). Generally, lower temperatures (about 4°C to about 50°C) and higher salt concentrations (about 5 mM to about 5 M) will assist in inhibiting or controlling RNase digestion in accordance with the invention. Salts used may include sodium chloride, potassium, chloride, magnesium chloride, sodium acetate etc. Additionally, lowering RNase amounts or concentrations may be used to accomplish the desired result. Such concentrations for RNase A may range from about 0.001 ng/μg mRNA to about 500 ng/μg of mRNA and for RNase I may range from about 0.001 u/μg mRNA to about 500 u/μg mRNA. The incubation temperature, RNase concentration and salt concentration may be readily determined by one skilled in the art. In a preferred aspect, concentration of the RNase A include ranges from 0.1 ng/μg mRNA to 10 ng/μg mRNA in TE buffer (10 mM Tris, pH 7.5,

1 mM EDTA) at 37°C. Alternatively, the concentration of the RNase A can include ranges from 0.1 ng/μg mRNA to 500 ng/μg mRNA in 10 mM Tris, pH 7.5 buffer containing 250 mM NaCl at 25°C for 30 minutes. Preferably, concentration of the RNase I used ranges from 0.1 unit/μg mRNA to 1.0 unit/μg mRNA in 10 mM Tris-HCl (pH 7.5), 5 mM EDTA (pH 8.0), 200 mM sodium acetate at 37°C. Alternatively, the concentration of the RNase I can be used at ranges from 1.0 unit/μg mRNA to 100 units/μg mRNA in the same buffer at 25°C for 30 minutes.

In another aspect, the invention relates to capture or binding of the cap structure (e.g., m^7GpppN) of the mRNA before, during or after first strand cDNA synthesis. Thus, the invention relates to selection of mRNA (before first strand synthesis) or mRNA/cDNA hybrids (after or during first strand synthesis) which have the cap structure in carrying out the methods of the invention. Such selection or capture may be accomplished with any cap binding molecule such as eIF4E, eIF4E peptides, eIF4E peptide fragments (see WO 98/08865) and antibodies or antibody fragments specific for cap structure. In a preferred aspect, selection of the cap structure is accomplished after first strand synthesis. More preferably, such cap capture occurs after ribonuclease digestion in accordance with the methods of the invention. For example, mRNA/cDNA hybrids subjected to ribonuclease digestion are captured and then used for second strand cDNA synthesis according to the invention.

Thus, the present invention is generally directed to methods of synthesizing nucleic acid molecules. The present invention is more specifically directed to methods of making one or more nucleic acid molecules, especially cDNA molecules or cDNA libraries, comprising mixing one or more nucleic acid templates (preferably mRNA, poly A RNA or a population of mRNA molecules) with at least one polypeptide having reverse transcriptase activity, and incubating the mixture under conditions sufficient to make one or more first nucleic acid molecules (e.g., first strand cDNA) complementary to all or a portion of the one or more nucleic acid templates.

In accordance with the invention, such conditions provide for an increased total amount of nucleic acid molecule (cDNA) produced, compared to conventional procedures which do not employ the improved modifications or conditions of the invention. The invention also provides for an increase of length or average size of the nucleic acid molecules (cDNA) produced and/or an increase in the percentage or amount of full-length nucleic acid molecules (cDNA) produced, compared to conventional procedures which do not employ the improved modifications or conditions of the invention. Determining the amount, length and full-length content of the cDNA produced can be determined by conventional techniques well known in the art and as described herein. The percentage or average percentages of full-length cDNA in cDNA libraries produced in accordance with the invention are preferably above about 15%, more preferably above about 20%, more preferably above about 25%, more preferably above about 30%, more preferably above about 40%, more preferably above about 50%, more preferably above about 60%, more preferably above about 70%, more preferably above about 80% and most preferably above about 90%. Such full-length percentages are preferably determined by random selection of a portion of the clones of the cDNA library of interest (e.g., 100 to 1000 clones), sequencing the clones and comparing the sequences to known sequence data bases.

In preferred aspects of the invention, the improved results of the invention are preferably accomplished by one or a combination of modifications to the conditions for nucleic acid or cDNA synthesis. Such conditions preferably include modifications for improving first strand cDNA synthesis and/or improving second strand cDNA synthesis.

In a preferred aspect, the invention specifically relates to methods of making one or more double stranded cDNA molecules comprising incubating one or more mRNA molecules (preferably a population of mRNA molecules) with one or more primers of the invention at temperatures and primer concentrations to prevent, inhibit, reduce or substantially reduce internal

priming prior to or during first strand cDNA synthesis. Such reaction is preferably conducted in the presence of one or more inhibitors of reverse transcriptase activity in accordance with the invention. Ribonuclease digestion is preferably conducted before second strand cDNA synthesis and at ribonuclease concentrations sufficient to increase the length, amount and/or size of double stranded cDNA molecules produced during second strand synthesis. In accordance with the invention, cap capture is preferably accomplished during or after the ribonuclease digestion.

The invention is also directed to nucleic acid molecules and cDNA molecules or populations of cDNA molecules (single or double-stranded) produced according to the above-described methods and to vectors (particularly expression vectors) comprising these nucleic acid molecules and cDNA molecules. The invention also relates to host cells containing such cDNA molecules and/or vectors.

The invention is also directed to kits for use in the methods of the invention. Such kits can be used for making single or double-stranded nucleic acid molecules. The kits of the invention comprise a carrier, such as a box or carton, having therein one or more containers, such as vials, tubes, bottles and the like. Such kits may comprise at least one component selected from the group consisting of primers (preferably primers having higher specificity and most preferably oligo(dT) primers having a length equal to or greater than 20 bases), one or more polypeptides having reverse transcriptase activity (reverse transcriptases and DNA polymerases), one or more inhibitors of reverse transcription (e.g., antibodies and antibody fragments directed against polypeptides having RT activity), one or more cap binding molecules (e.g., antibodies or antibody fragments directed against cap structure), nucleic acid synthesis reaction buffers, one or more nucleotides, one or more vectors, and instructions for carrying out the methods of the invention.

The invention also relates to compositions for use in the invention or made while carrying out the methods of the invention. Such compositions

may comprise at least one primer (e.g., oligo(dT) or derivatives thereof) and at least one template in a sample or reaction mixture in amounts or ratios in accordance with the invention. Such composition may further comprise one or more polypeptides having reverse transcriptase activity, one or more reverse transcription inhibitors (e.g., anti-RT antibodies or fragments thereof), one or more nucleotides, one or more cap binding molecules (e.g., anti-cap antibodies for fragments thereof), one or more buffering salts and the like. Such compositions may also be maintained at a temperature to avoid internal priming in accordance with the invention.

The compositions of the invention may also comprise amounts of ribonuclease in accordance with the invention. Such compositions may further comprise at least one component selected from one or more mRNA/cDNA hybrids, one or more nucleotides, one or more polypeptides having reverse transcriptase activity, one or more buffering salts, one or more cap binding molecules (e.g., anti-cap antibodies or fragments thereof) and the like.

The invention also relates to one or more antibodies (monoclonal and polyclonal) and fragments thereof for use in the methods, compositions and kits of the invention. Such antibodies, include anti-cap and/or anti-RT antibodies and antibody fragments.

Other preferred embodiments of the present invention will be apparent to one of ordinary skill in the art in view of the following drawings and description of the invention.

Brief Description of the Drawings

Figure 1 is an autoradiograph of first strand cDNA synthesized with SuperScript™ II (SS II) RT at 45°C with a 5/6 Kb template with molar ratios of oligo$(dT)_{25\text{-}30}$/mRNA of 1:1, 2.5:1, 5:1, 10:1, and 50:1.

Figure 2 is an autoradiograph of first strand cDNA synthesized with ThermoScript™ II (TS II) RT at 45°C, 50°C and 55°C with a 5/6 Kb template with molar ratios of oligo $(dT)_{25-30}$/mRNA of 1:1, 2.5:1, 5:1, 10:1, and 50:1.

Figure 3 is an autoradiograph of first strand cDNA synthesized with SS II RT using standard reaction temperatures and varying reaction temperatures with a molar ratio of biotinylated-*Not* I-oligo$(dT)_{25}$/mRNA of 0:1, 1:1 and 15:1.

Figure 4 is an autoradiograph of first strand cDNA synthesized with TS II RT using standard reaction conditions in which the primer/template annealing is incubated on ice prior to cDNA synthesis and using conditions according to the invention in which annealing and the synthesis reaction temperatures are maintained above 30°C (preferably above 37°C) with a molar ratio of biotinylated-*Not* I-oligo$(dT)_{25}$/mRNA of 1:1 and 15:1. Maintaining the annealing and reaction temperatures above 30°C (preferably above 37°C) in accordance with the invention may also be referred to as "hot start."

Figure 5 is an autoradiograph of second strand cDNA synthesized using different amounts of RNase A.

Figure 6 is an autoradiograph of second strand cDNA synthesized using different amounts of RNase I.

Detailed Description of the Preferred Embodiments

Definitions:

In order to provide a clearer and consistent understanding of the specification and claims, including the scope to be given such terms, the following definitions are provided.

Internal priming as used herein refers to hybridization or annealing of one or more primers at one or more sites within one or more mRNA molecules other than at the poly A tail located at the 3' termini of the mRNA molecule.

Library as used herein refers to a set of nucleic acid molecules (circular or linear) which is representative of all or a portion or significant portion of the DNA content of an organism (a "genomic library"), or a set of nucleic acid molecules representative of all or a portion or significant portion of the expressed genes (a "cDNA library") in a cell, tissue, organ or organism. Such libraries may or may not be contained in one or more vectors.

Vector as used herein refers to a plasmid, cosmid, phagemid or phage DNA or other DNA molecule which is able to replicate autonomously in a host cell, and which is characterized by one or a small number of restriction endonuclease recognition sites at which such DNA sequences may be cut in a determinable fashion without loss of an essential biological function of the vector, and into which DNA may be inserted in order to bring about its replication and cloning. The vector may further contain one or more markers suitable for use in the identification of cells transformed with the vector. Markers, for example, include but are not limited to tetracycline resistance or ampicillin resistance. Such vectors may also contain one or more recombination sites, one or more termination sites, one or more origins of replication, and the like.

Primer as used herein refers to a single-stranded oligonucleotide that is extended by covalent bonding of nucleotide monomers during amplification or polymerization of a DNA molecule. Preferred primers for use in the invention include oligo(dT) primers or derivatives or variants thereof.

Oligonucleotide as used herein refers to a synthetic or natural molecule comprising a covalently linked sequence of nucleotides which are joined by a phosphodiester bond between the 3' position of the deoxyribose or ribose of one nucleotide and the 5' position of the deoxyribose or ribose of the adjacent nucleotide.

Template as used herein refers to double-stranded or single-stranded nucleic acid molecules which are to be amplified, synthesized or sequenced. In the case of a double-stranded molecules, denaturation of its strands to form

a first and a second strand is preferably performed before these molecules may be amplified, synthesized or sequenced, or the double stranded molecule may be used directly as a template. For single stranded templates, a primer, complementary to a portion of the template is hybridized or annealed under appropriate conditions and one or more polymerases or reverse transcriptases may then synthesize a nucleic acid molecule complementary to all or a portion of said template. The newly synthesized molecules, according to the invention, may be equal or shorter in length than the original template.

Incorporating as used herein means becoming a part of a DNA and/or RNA molecule or primer.

Amplification as used herein refers to any *in vitro* method for increasing the number of copies of a nucleotide sequence with the use of a polymerase. Nucleic acid amplification results in the incorporation of nucleotides into a DNA and/or RNA molecule or primer thereby forming a new molecule complementary to a template. The formed nucleic acid molecule and its template can be used as templates to synthesize additional nucleic acid molecules. As used herein, one amplification reaction may consist of many rounds of replication. DNA amplification reactions include, for example, polymerase chain reactions (PCR). One PCR reaction may consist of 5 to 100 "cycles" of denaturation and synthesis of a DNA molecule.

Nucleotide as used herein refers to a base-sugar-phosphate combination. Nucleotides are monomeric units of a nucleic acid sequence (DNA and RNA). The term nucleotide includes ribonucleoside triphosphate ATP, UTP, CTG, GTP and deoxyribonucleoside triphosphates such as dATP, dCTP, dITP, dUTP, dGTP, dTTP, or derivatives thereof. Such derivatives include, for example, [αS]dATP, 7-deaza-dGTP, 7-deaza-dATP, and biotinylated or haptenylated nucleotides. The term nucleotide as used herein also refers to dideoxyribonucleoside triphosphates (ddNTPs) and their derivatives. Illustrated examples of dideoxyribonucleoside triphosphates include, but are not limited to, ddATP, ddCTP, ddGTP, ddITP, and ddTTP.

According to the present invention, a "nucleotide" may be unlabeled or detectably labeled by well known techniques. Detectable labels include, for example, radioactive isotopes, fluorescent labels, chemiluminescent labels, bioluminescent labels and enzyme labels.

Hybridization or **annealing** as used herein refers to base pairing of two complementary single-stranded nucleic acid molecules (RNA and/or DNA) to give a double-stranded molecule. As used herein, two nucleic acid molecules may be hybridized or annealed, although the base pairing is not completely complementary. Accordingly, mismatched bases do not prevent hybridization or annealing of two nucleic acid molecules provided that appropriate conditions, well known in the art, are used. In the present invention, the term hybridization or annealing preferably refers to hybridization of one or more primers (e.g., oligo(dT) or derivatives thereof) to one or more templates (e.g., mRNA).

Host cell as used herein refers to any prokaryotic or eukaryotic cell that is the recipient of a replicable expression vector or cloning vector. The terms "host" or "host cell" may be used interchangeably herein. For examples of such hosts, see Maniatis *et al.*, "Molecular Cloning: A Laboratory Manual," Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1982). Preferred prokaryotic hosts include, but are not limited to, bacteria of the genus *Escherichia* (e.g., *E. coli), Bacillus, Staphylococcus, Agrobacter* (e.g., *A. tumefaciens), Streptomyces, Pseudomonas, Salmonella, Serratia, Caryophanon*, etc. The most preferred prokaryotic host is *E. coli.* Bacterial hosts of particular interest in the present invention include *E. coli* strains K12, DH10B, DH5α, Stbl2 and HB101, and others available from Life Technologies, Inc. Preferred eukaryotic hosts include, but are not limited to, fungi, fish cells, yeast cells, plant cells and animal cells. Particularly preferred animal cells are insect cells such as *Drosophila* cells, *Spodoptera* Sf9, Sf21 cells and *Trichoplusa* High-Five cells; nematode cells such as *C. elegans* cells;

and mammalian cells such as COS cells, CHO cells, VERO cells, 293 cells, PERC6 cells, BHK cells and human cells.

Expression vector as used herein refers to a vector which is capable of enhancing the expression of a gene or portion of a gene which has been cloned into it, after transformation or transfection into a host cell. The cloned gene is usually placed under the control (i.e., operably linked to) certain control sequences such as promoter sequences. Such promoters include but are not limited to phage lambda P_L promoter, and the *E. coli lac, trp* and *tac* promoters. Other suitable promoters will be known to the skilled artisan.

The nucleic acid templates suitable for reverse transcription according to the invention include any nucleic acid molecule or populations of nucleic acid molecules (preferably one or more RNA molecules (e.g., one or more mRNA molecules or poly A^+ RNA molecules, and more preferably a population of mRNA molecules) or one or more DNA molecules), particularly those derived from a cell or tissue. In a preferred aspect, a population of mRNA molecules (a number of different mRNA molecules) are used to make a cDNA library according to the present invention.

To make the nucleic acid molecule or molecules complementary to the one or more templates, a primer (e.g., an oligo(dT) primer) and one or more nucleotides are used for nucleic acid synthesis typically in the 3' to 5' direction. Nucleic acid molecules suitable for reverse transcription according to this aspect of the invention include any nucleic acid molecule, particularly those derived from a prokaryotic or eukaryotic cell. Such cells may include normal cells, diseased cells, transformed cells, established cells, progenitor cells, precursor cells, fetal cells, embryonic cells, bacterial cells, yeast cells, animal cells (including human cells), avian cells, plant cells and the like, or tissue isolated from a plant (e.g., corn, tomato, tobacco, potato, soy bean, etc.) or an animal (e.g., human, cow, pig, mouse, sheep, horse, monkey, canine, feline, rat, rabbit, bird, fish, insect, etc.). Such nucleic acid molecules may also be isolated from viruses.

The nucleic acid molecules that are used as templates to prepare cDNA molecules according to the methods of the present invention are preferably obtained from natural sources, such as a variety of cells, tissues, organs or organisms. Cells that may be used as sources of nucleic acid molecules may be prokaryotic (bacterial cells, including but not limited to those of species of the genera *Escherichia*, *Bacillus*, *Serratia*, *Salmonella*, *Staphylococcus*, *Streptococcus*, *Clostridium*, *Chlamydia*, *Neisseria*, *Treponema*, *Mycoplasma*, *Borrelia*, *Legionella*, *Pseudomonas*, *Mycobacterium*, *Helicobacter*, *Erwinia*, *Agrobacterium*, *Rhizobium*, *Xanthomonas* and *Streptomyces*) or eukaryotic (including fungi (especially yeasts), plants, protozoans and other parasites, and animals including insects (particularly *Drosophila* spp. cells), nematodes (particularly *Caenorhabditis elegans* cells), and mammals (particularly human cells)).

Mammalian somatic cells that may be used as sources of nucleic acids include blood cells (reticulocytes and leukocytes), endothelial cells, epithelial cells, neuronal cells (from the central or peripheral nervous systems), muscle cells (including myocytes and myoblasts from skeletal, smooth or cardiac muscle), connective tissue cells (including fibroblasts, adipocytes, chondrocytes, chondroblasts, osteocytes and osteoblasts) and other stromal cells (e.g., macrophages, dendritic cells, Schwann cells). Mammalian germ cells (spermatocytes and oocytes) may also be used as sources of nucleic acids for use in the invention, as may the progenitors, precursors and stem cells that give rise to the above somatic and germ cells. Also suitable for use as nucleic acid sources are mammalian tissues or organs such as those derived from brain, kidney, liver, pancreas, blood, bone marrow, muscle, nervous, skin, genitourinary, circulatory, lymphoid, gastrointestinal and connective tissue sources, as well as those derived from a mammalian (including human) embryo or fetus.

Any of the above cells, tissues and organs may be normal, diseased, transformed, established, progenitors, precursors, fetal or embryonic.

Diseased cells may, for example, include those involved in infectious diseases (caused by bacteria, fungi or yeast, viruses (including AIDS, HIV, HTLV, herpes, hepatitis and the like) or parasites), in genetic or biochemical pathologies (e.g., cystic fibrosis, hemophilia, Alzheimer's disease, muscular dystrophy or multiple sclerosis) or in cancerous processes. Transformed or established animal cell lines may include, for example, COS cells, CHO cells, VERO cells, BHK cells, HeLa cells, HepG2 cells, K562 cells, 293 cells, L929 cells, F9 cells, and the like. Other cells, cell lines, tissues, organs and organisms suitable as sources of nucleic acids for use in the present invention will be apparent to one of ordinary skill in the art.

Once the starting cells, tissues, organs or other samples are obtained, nucleic acid molecules (such as mRNA) may be isolated therefrom by methods that are well-known in the art (See, e.g., Maniatis, T., *et al.*, *Cell 15*:687-701 (1978); Okayama, H., and Berg, P., *Mol. Cell. Biol. 2*:161-170 (1982); Gubler, U., and Hoffman, B.J., *Gene 25*:263-269 (1983); and Message Maker™ mRNA Isolation System available from Life Technologies, Inc.). The nucleic acid molecules thus isolated may then be used to prepare cDNA molecules and cDNA libraries in accordance with the present invention. The cDNA molecules and/or cDNA libraries produced in accordance with the invention are preferably contained in one or more vectors. Such vectors may be introduced into one or more host cells by standard transformation or transfection techniques well known in the art. Preferred host cells include prokaryotic host cells such as cells of the genus *Escherichia*, particularly *E. coli*.

Enzymes for use in the compositions, methods and kits of the invention include any enzyme having reverse transcriptase activity. Such enzymes include, but are not limited to, retroviral reverse transcriptase, retrotransposon reverse transcriptase, hepatitis B reverse transcriptase, cauliflower mosaic virus reverse transcriptase, bacterial reverse transcriptase, *Tth* DNA polymerase, *Taq* DNA polymerase (Saiki, R.K., *et al.*, *Science 239*:487-491

(1988); U.S. Patent Nos. 4,889,818 and 4,965,188), *Tne* DNA polymerase (WO 96/10640), *Tma* DNA polymerase (U.S. Patent No. 5,374,553) and mutants, fragments, variants or derivatives thereof (see, e.g., commonly owned, co-pending U.S. Patent Application Nos. 08/706,702 and 08/706,706, both filed September 9, 1996, which are incorporated by reference herein in their entireties). As will be understood by one of ordinary skill in the art, modified reverse transcriptases and DNA polymerase having RT activity may be obtained by recombinant or genetic engineering techniques that are well-known in the art. Mutant reverse transcriptases or polymerases can, for example, be obtained by mutating the gene or genes encoding the reverse transcriptase or polymerase of interest by site-directed or random mutagenesis. Such mutations may include point mutations, deletion mutations and insertional mutations. Preferably, one or more point mutations (e.g., substitution of one or more amino acids with one or more different amino acids) are used to construct mutant reverse transcriptases or polymerases for use in the invention. Fragments of reverse transcriptases or polymerases may also be obtained by deletion mutation by recombinant techniques that are well-known in the art, or by enzymatic digestion of the reverse transcriptase(s) or polymerase(s) of interest using any of a number of well-known proteolytic enzymes.

Preferred enzymes for use in the invention include those that are reduced or substantially reduced in RNase H activity. Such enzymes that are reduced or substantially reduced in RNase H activity may be obtained by mutating the RNase H domain within the reverse transcriptase of interest, preferably by one or more point mutations, one or more deletion mutations, and/or one or more insertion mutations as described above. By an enzyme "substantially reduced in RNase H activity" is meant that the enzyme has less than about 30%, less than about 25%, less than about 20%, more preferably less than about 15%, less than about 10%, less than about 7.5%, or less than about 5%, and most preferably less than about 5% or less than about 2%, of

the RNase H activity of the corresponding wildtype or RNase H^+ enzyme such as wildtype Moloney Murine Leukemia Virus (M-MLV), Avian Myeloblastosis Virus (AMV) or Rous Sarcoma Virus (RSV) reverse transcriptases. The RNase H activity of any enzyme may be determined by a variety of assays, such as those described, for example, in U.S. Patent No. 5,244,797, in Kotewicz, M.L., *et al.*, *Nucl. Acids Res. 16*:265 (1988), in Gerard, G.F., *et al.*, *FOCUS 14(5)*:91 (1992), and in U.S. Patent No. 5,668,005, the disclosures of all of which are fully incorporated herein by reference.

Polypeptides having reverse transcriptase activity for use in the invention may be obtained commercially, for example from Life Technologies, Inc. (Rockville, Maryland), Pharmacia (Piscataway, New Jersey), Sigma (Saint Louis, Missouri) or Boehringer Mannheim Biochemicals (Indianapolis, Indiana). Alternatively, polypeptides having reverse transcriptase activity may be isolated from their natural viral or bacterial sources according to standard procedures for isolating and purifying natural proteins that are well-known to one of ordinary skill in the art (see, e.g., Houts, G.E., *et al.*, *J. Virol. 29*:517 (1979)). In addition, the polypeptides having reverse transcriptase activity may be prepared by recombinant DNA techniques that are familiar to one of ordinary skill in the art (see, e.g., Kotewicz, M.L., *et al.*, *Nucl. Acids Res. 16*:265 (1988); Soltis, D.A., and Skalka, A.M., *Proc. Natl. Acad. Sci. USA 85*:3372-3376 (1988)).

Preferred polypeptides having reverse transcriptase activity for use in the invention include M-MLV reverse transcriptase, RSV reverse transcriptase, AMV reverse transcriptase, Rous Associated Virus (RAV) reverse transcriptase, Myeloblastosis Associated Virus (MAV) reverse transcriptase and Human Immunodeficiency Virus (HIV) reverse transcriptase, and others described in WO 98/47921 and derivatives, variants, fragments or mutants thereof, and combinations thereof. In a further preferred embodiment, the reverse transcriptases are reduced or substantially reduced in RNase H

activity, and are most preferably selected from the group consisting of M-MLV H⁻ reverse transcriptase, RSV H⁻ reverse transcriptase, AMV H⁻ reverse transcriptase, RAV H⁻ reverse transcriptase, MAV H⁻ reverse transcriptase and HIV H⁻ reverse transcriptase, and derivatives, variants, fragments or mutants thereof, and combinations thereof. Reverse transcriptases of particular interest include AMV RT and M-MLV RT, and more preferably AMV RT and M-MLV RT having reduced or substantially reduced RNase H activity (preferably AMV RT $\alpha H^-/BH^+$ and M-MLV RT H^-). The most preferred reverse transcriptases for use in the invention include SuperScript™, SuperScript™ II, ThermoScript™ and ThermoScript™ II available from Life Technologies, Inc. See generally, WO 98/47921, U.S. Patents 5,244,797 and 5,668,005, the entire contents of each of which are herein incorporated by reference.

A variety of DNA polymerases are useful in accordance with the present invention. Such polymerases include, but are not limited to, *Thermus thermophilus* (*Tth*) DNA polymerase, *Thermus aquaticus* (*Taq*) DNA polymerase, *Thermotoga neapolitana* (*Tne*) DNA polymerase, *Thermotoga maritima* (*Tma*) DNA polymerase, *Thermococcus litoralis* (*Tli* or VENT™) DNA polymerase, *Pyrococcus furiosis* (*Pfu*) DNA polymerase, DEEPVENT™ DNA polymerase, *Pyrococcus woosii* (*Pwo*) DNA polymerase, *Bacillus sterothermophilus* (*Bst*) DNA polymerase, *Bacillus caldophilus* (*Bca*) DNA polymerase, *Sulfolobus acidocaldarius* (*Sac*) DNA polymerase, *Thermoplasma acidophilum* (*Tac*) DNA polymerase, *Thermus flavus* (*Tfl/Tub*) DNA polymerase, *Thermus ruber* (*Tru*) DNA polymerase, *Thermus brockianus* (DYNAZYME™) DNA polymerase, *Methanobacterium thermoautotrophicum* (*Mth*) DNA polymerase, *Mycobacterium* spp. DNA polymerase (*Mtb*, *Mlep*), and mutants, variants and derivatives thereof.

DNA polymerases used in accordance with the invention may be any enzyme that can synthesize a DNA molecule from a nucleic acid template, typically in the 5' to 3' direction. Such polymerases may be mesophilic or

thermophilic. Mesophilic polymerases include T4 DNA polymerase, T5 DNA polymerase, T7 DNA polymerase, Klenow fragment DNA polymerase, DNA polymerase III, DNA polymerase I and the like. Thermostable DNA polymerases include *Taq*, *Tne*, *Tma*, *Pfu*, VENT™, DEEPVENT™, *Tth* and mutants, variants and derivatives thereof (U.S. Patent No. 5,436,149; U.S. Patent No. 5,512,462; WO 92/06188; WO 92/06200; WO 96/10640; Barnes, W.M., *Gene 112*:29-35 (1992); Lawyer, F.C., *et al.*, *PCR Meth. Appl. 2*:275-287 (1993); Flaman, J.-M., *et al.*, *Nucl. Acids Res. 22(15)*:3259-3260 (1994)).

DNA polymerases for use in the invention may be obtained commercially, for example from Life Technologies, Inc. (Rockville, Maryland), Perkin-Elmer (Branchburg, New Jersey), New England BioLabs (Beverly, Massachusetts) or Boehringer Mannheim Biochemicals (Indianapolis, Indiana).

The present invention is also directed to nucleic acid molecules produced by the methods of the invention, which may be cDNA molecules, especially full-length cDNA molecules, to vectors (particularly expression vectors) comprising these nucleic acid molecules and cDNA molecules and to host cells comprising these nucleic acid molecules, cDNA molecules, and/or vectors.

Recombinant vectors may be produced according to this aspect of the invention by inserting, using methods that are well-known in the art, one or more of the cDNA molecules or nucleic acid molecules prepared according to the present methods into one or more vectors. The vector used in this aspect of the invention may be, for example, a phage or a plasmid vector, and is preferably a plasmid. Preferred are vectors comprising *cis*-acting control regions to the nucleic acid encoding the polypeptide of interest. Appropriate *trans*-acting factors may be supplied by the host, supplied by a complementing vector or supplied by the vector itself upon introduction into the host.

Expression vectors useful in the present invention include chromosomal-, episomal- and virus-derived vectors, e.g., vectors derived from

bacterial plasmids or bacteriophages, and vectors derived from combinations thereof, such as cosmids and phagemids, and will preferably include at least one selectable marker such as a tetracycline or ampicillin resistance gene for culturing in a bacterial host cell. Prior to insertion into such an expression vector, the cDNA or nucleic acid molecules of the invention should be operatively linked to an appropriate promoter.

Among vectors preferred for use in the present invention include pQE70, pQE60 and pQE-9, available from Qiagen; pBS vectors, Phagescript vectors, Bluescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene; pcDNA3 available from Invitrogen; pGEX, pTrxfus, pTrc99a, pET-5, pET-9, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia; and pSPORT1, pSPORT2, pSV·SPORT1, pCMVSPORT6 and pCMVSPORT available from Life Technologies, Inc. Other suitable vectors will be readily apparent to the skilled artisan.

The invention may be used in conjunction with any methods of cDNA synthesis that are well-known in the art (see, e.g., Gubler, U., and Hoffman, B.J., *Gene 25*:263-269 (1983); Krug, M.S., and Berger, S.L., *Meth. Enzymol. 152*:316-325 (1987); Sambrook, J., *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, pp. 8.60-8.63 (1989); PCT US98/19948; and WO 98/51699) to produce cDNA molecules or libraries. Other methods of cDNA synthesis which may advantageously use the present invention will be readily apparent to one of ordinary skill in the art.

Having obtained cDNA molecules or libraries according to the present methods, these cDNAs may be isolated for further analysis or manipulation. Detailed methodologies for purification of cDNAs are taught in the GENETRAPPER™ manual (Life Technologies), which is incorporated herein by reference in its entirety, although alternative standard techniques that are known in the art (see, e.g., Sambrook, J., *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor, NY: Cold Spring Harbor

Laboratory Press, pp. 8.60-8.63 (1989)) may also be used. The cDNA molecules or libraries produced by the invention may also be further manipulated by standard molecular biology techniques such as two hybrid analysis, cDNA normalization, sequencing and amplification. More particularly, the methods of the invention and the cDNA molecules or libraries produced by such methods may be used in combination with RT-PCR and 5' RACE technology (Life Technologies, Inc.) and differential display.

A variety of inhibitors and binding molecules are suitable for use in the present methods. Included among these inhibitors or binding molecules are antibodies that bind to the above-described polypeptides having reverse transcriptase activity (such as anti-RT antibodies including anti-AMV RT antibodies, anti-M-MLV RT antibodies or anti-RSV RT antibodies) or to cap structure (e.g., anti-cap antibodies), and fragments thereof (such as Fab or $F(ab')_2$ fragments). Such antibodies may be polyclonal or monoclonal, and may be prepared in a variety of species according to methods that are well-known in the art. See, for instance, Sutcliffe, J.G., *et al.*, *Science 219:*660-666 (1983); Wilson *et al., Cell 37*: 767 (1984); and Bittle, F.J., *et al., J. Gen. Virol. 66*:2347-2354 (1985). Antibodies specific for any of the above-described reverse transcriptases or cap structures can be raised against the intact polymerase polypeptide or cap structures or one or more fragments thereof. These polypeptides or cap structures or fragments thereof may be presented together with a carrier protein (e.g., albumin) to an animal system (such as rabbit or mouse) or, if they are long enough (at least about 25 amino acids), without a carrier.

As used herein, the term "antibody" (Ab) may be used interchangeably with the terms "polyclonal antibody" or "monoclonal antibody" (mAb), except in specific contexts as described below. These terms, as used herein, are meant to include intact molecules as well as antibody fragments (such as, for example, Fab and $F(ab')_2$ fragments) which are capable of specifically binding

to a polypeptide having reverse transcriptase activity (such as a DNA polymerase or a reverse transcriptase) or cap structures or portions thereof.

The antibodies used in the methods of the present invention may be polyclonal or monoclonal, and may be prepared by any of a variety of methods (see, e.g., U.S. Patent No. 5,587,287). For example, polyclonal antibodies may be made by immunizing an animal with one or more polypeptides having reverse transcriptase activity or cap structures or portions thereof according to standard techniques (see, e.g., Harlow, E., and Lane, D., *Antibodies: A Laboratory Manual*, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press (1988); Kaufman, P.B., *et al.*, In: *Handbook of Molecular and Cellular Methods in Biology and Medicine*, Boca Raton, Florida: CRC Press, pp. 468-469 (1995)). Alternatively, monoclonal antibodies (or fragments thereof) to be used in the present methods may be prepared using hybridoma technology that is well-known in the art (Köhler *et al., Nature 256*:495 (1975); Köhler *et al., Eur. J. Immunol. 6*:511 (1976); Köhler *et al., Eur. J. Immunol. 6*:292 (1976); Hammerling *et al.*, In: *Monoclonal Antibodies and T-Cell Hybridomas*, New York: Elsevier, pp. 563-681 (1981); Kaufman, P.B., *et al.*, In: *Handbook of Molecular and Cellular Methods in Biology and Medicine*, Boca Raton, Florida: CRC Press, pp. 444-467 (1995)).

It will be appreciated that Fab, $F(ab')_2$ and other fragments of the above-described antibodies may be used in the methods described herein. Such fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce $F(ab')_2$ fragments). Antibody fragments may also be produced through the application of recombinant DNA technology or through synthetic chemistry.

The invention also provides kits for use in accordance with the invention. Such kits comprise a carrier means, such as a box or carton, having in close confinement therein one or more container means, such as vials, tubes, bottles and the like, wherein the kit may comprise (in the same or separate containers) one or more host cells, one or more reverse transcriptases, one or

more reverse transcription inhibitors, one or more cap binding molecules, one or more DNA polymerases, suitable buffers, one or more nucleotides and/or one or more primers (e.g., oligo(dT) for reverse transcription). The kits encompassed by this aspect of the present invention may further comprise additional reagents and compounds necessary for carrying out standard nucleic acid reverse transcription protocols.

It will be readily apparent to one of ordinary skill in the relevant art that other suitable modifications and adaptations to the methods and applications described herein are obvious and may be made without departing from the scope of the invention or any embodiment thereof. Having now described the present invention in detail, the same will be more clearly understood by reference to the following examples, which are included herewith for purposes of illustration only and are not intended to be limiting of the invention.

Examples

Example 1

Comparison of first strand cDNA synthesis with varying ratios of oligo (dT) primer/mRNA

This example compares first strand cDNA synthesis of the MAP4 gene with various ratios of oligo dT primer/starting mRNA. All components are available from Life Technologies, Inc., Rockville, Maryland, unless specified otherwise.

The master mix for Superscript II reverse transcriptase (SS II RT) was prepared as specified in Table 1 below.

Table 1

Component	µl	µl
5X SSII RT buffer	4	28
0.1 M DTT	2	14
10 mM dNTP	1	7
α-^{32}P dCTP	0.5	3.5
Water	1.5	10.5
Total volume	**9**	**63**

The master mix for ThermoScript™ II RT (TS RT) (AMV RT $\alpha H^{-}\beta H^{+}$) (see WO 98/47921) was prepared as specified in Table 2 below.

Table 2

Component	µl	µl
10X TS II buffer*	2	38
0.1 M DTT	2	38
10 mM dNTP	2	38
α-^{32}P dCTP	0.5	9.5
Rnase OUT (40 u/µl)	1	19
Water	1.5	28.5
Total volume	**9**	**171**

*10X TS II buffer comprises 50 mM Tris-HCl (pH 8.4), 750 mM KCl, and 75 mM $MgCl_2$.

The master annealing mix was prepared by adding a 5 Kb MAP4 mRNA, oligo$(dT)_{25\text{-}30}$ and water to 5 tubes in the amounts specified in Table 3 below.

Table 3

Ratio of oligo (dT)/mRNA	1:1	2.5:1	5:1	10:1	50:1
Component	Volume (µl)				
MAP4 mRNA (1 µg/µl)	5	5	5	5	5
Oligo$(dT)_{25\text{-}30}$ (10 ng/µl)	2.6	6.5	13.1	-	-
Oligo$(dT)_{25\text{-}30}$ (100 ng/µl)	-	-	-	2.6	13.1
Water	42.4	38.5	31.9	42.4	31.9
Total volume	**50**	**50**	**50**	**50**	**50**

The mixture was heated at 70°C for 10 minutes and then chilled on ice for 5 minutes.

Synthesis of first strand cDNA was done by adding 9 µl of the appropriate reverse transcriptase master mix, 10 µl of the master annealing mix and 1 µl of either SS II RT (200 units/ul) or TS II RT (15 units/ul) for a total volume of 20 µl as summarized in Table 4 below.

Table 4

Tube	Reverse Transcriptase	Temperature	Ratio of oligo $(dT)_{25-30}$/ mRNA
1	SS II	45°C	1
2			2.5
3			5
4			10
5			50
1	TS II RT	45°C	1
2			2.5
3			5
4			10
5			50
6		50°C	1
7			2.5
8			5
9			10
10			50
11		55°C	1
12			2.5
13			5
14			10
15			50

The reactions were incubated for 1 hour at 45°C for SS II RT and at 45, 50 or 55°C for TS II RT. The tubes were placed on ice to complete the reaction. 18 µl first stand cDNA of the reaction tube was precipitated and re-suspended in 10 µl of water. 5 µl of the first strand cDNA was mixed with 5 µl of standard loading buffer (60 mM NaOH, 4 mM EDTA, 0.1%

bromophenol blue), and loaded onto 1.4% alkaline agarose gel for analysis. These results are shown in Figures 1 and 2.

Figure 1 is an autoradiograph of first strand cDNA synthesized with SS II RT at 45°C. Lane M is the 1 kb DNA ladder. Lanes 1-5 represents reaction conditions with a molar ratio of oligo$(dT)_{25-30}$/mRNA of 1:1, 2.5:1, 5:1, 10:1 and 50:1, respectively. Figure 2 is an autoradiograph of first strand cDNA synthesized with TS II RT. Lane M is the 1 kb DNA ladder. Lanes 1-5 represent reaction conditions at 45°C with a molar ratio of oligo$(dT)_{25-30}$/mRNA of 1:1, 2.5:1, 5:1, 10:1 and 50:1, respectively. Lanes 6-10 represent reaction conditions at 50°C with a molar ratio of oligo $(dT)_{25-30}$/mRNA of 1:1, 2.5:1, 5:1, 10:1 and 50:1, respectively. Lanes 11-15 represent reaction conditions at 55°C with a molar ratio of oligo $(dT)_{25-30}$/mRNA of 1:1, 2.5:1, 5:1, 10:1 and 50:1, respectively. The results show that by reducing the molar ratio of oligo(dT) primer/mRNA (preferably to 1:1) internal priming with reverse transcriptase was almost entirely eliminated.

Example 2

Comparison of first strand cDNA synthesis under standard and hot start conditions

This experiment was designed to compare first strand cDNA synthesis of the MAP4 gene with standard reaction and hot start conditions.

The annealing mix was prepared by mixing 1 µg of MAP4 mRNA and biotinylated *Not* I oligo$(dT)_{25}$ primer ((Biotin)$_4$ GACTAGTTCTAGAT CGCGAGCGG CCGCCCTTTTT TTTTTTTTTTTT TTTTTTTT; see WO 98/51699) in the desired molar ratio of oligo (dT)/mRNA of 0:1, 1:1 or 15:1 in thin-walled PCR tubes and bringing the volume up to 10 µl with water. If several tubes are identical, they may be made in one batch and aliquotted accordingly. The annealing mix was kept on ice.

The master mix for Superscript II reverse transcriptase (SS II RT) was prepared as specified in Table 5 below.

Table 5

Component	µl	µl
5X SSII RT buffer	4	28
0.1 M DTT	2	14
10 mM dNTP	1	7
α-^{32}P dCTP	0.5	3.5
SSII RT (200 u/µl)	1	7
Water	1.5	10.5
Total volume	**10**	**70**

The SS II RT master mix was then divided into two equal aliquots, one for processing with standard reaction temperatures (batch 1) and one for processing with hot start reaction temperatures (batch 2). To allow for condensation, an additional 10% volume of water was added to batch 2. All mixes were kept on ice.

Synthesis of first strand cDNA was begun by briefly spinning tubes containing annealing mix to collect droplets, placing the tubes in a thermocycler and then heating them to 70°C for 10 minutes. After this 10 minute cycle at 70°C, the tubes of annealing mix for batch 1 were immediately removed to ice. The tubes of annealing mix for batch 2 were allowed to cool to 45°C in the thermocycler while the batch 2 master mix was placed in the thermocycler and incubated at 45°C for 5 minutes. After the 5 minute incubation, 11 µl of the master mix for batch 2 was added to each batch 2 annealing tube and mixed with a pipette 2 times. Care was taken not to spin the tubes to avoid lowering the temperature.

10 µl of the master mix for batch 1 was added to each batch 1 annealing tube. The batch 1 tubes were lightly vortexed and briefly centrifuged to collect condensation droplets. The batch 1 tubes were then returned to the thermocylcer and the tubes from both batch 1 and 2 were incubated at 45°C for one hour.

5 µl of the first strand cDNA from each tube was mixed with 5 µl of standard loading buffer (60 mM NaOH, 4 mM EDTA, 0.1% bromophenol

blue) and loaded onto 1.4% alkaline agarose gel for analysis. The results are shown in Figure 3.

Figure 3 is an autoradiograph of first strand cDNA synthesized with SS II RT. Lanes 1, 3 and 5 represents batch 1 reaction conditions with a molar ratio of biotinylated oligo(dT)/mRNA of 0:1, 1:1 and 15:1, respectively. Lanes 2, 4 and 6 represents batch 2 reaction conditions with a molar ratio of biotinylated oligo(dT)/mRNA of 0:1, 1:1 and 15:1, respectively.

First strand cDNA was also synthesized with TS II RT using 15 units of TS II RT per μg mRNA using a biotinylated oligo(dT)/mRNA ratio of 1:1 and 15:1. The same protocol described above was followed, except that the temperature was varied to 50°C. The results are shown in Figure 4. Figure 4 is an autoradiograph of first strand cDNA synthesized with TS II RT. Lane M is the 1 kb DNA ladder. Lanes 1 and 3 represent reactions conditions using standard reaction temperatures at a 1:1 ratio and 15:1 ratio, respectively. Lanes 2 and 4 represent hot start reactions conditions at a 1:1 ratio and 15:1 ratio, respectively, as described above.

The results indicated that by dropping the reaction temperature to the reverse transcriptase reaction temperature after denaturation of the primer and mRNA mixture, the reaction was started directly and internal priming was avoided entirely.

Example 3

Synthesis of double strand cDNA by controlling the reaction temperature and the concentration of salt and RNase

This example describes the synthesis of double stranded cDNA by controlling the reaction temperature and the concentration of salt and different ribonuclease (RNases) during the treatment of the cDNA/mRNA hybrids after first strand cDNA synthesis.

First strand cDNA was synthesized as described above in Example 2 and digested with either RNase I or RNase A as further described below.

RNase I digestion of first strand cDNA was done by re-suspending the first strand cDNA in 180 µl of water and 20 µl of 10X RNase I buffer (100 mM Tris-HCl (pH 7.5), 50 mM EDTA, 2 M sodium acetate). 2.5 units of RNase I (1 unit/µg mRNA) were added and the mixture was mixed well. The RNase I digestion mixture was incubated at 25°C for 30 minutes and extracted with phenol/chloroform once. The supernatant was precipitated with 1 µl of glycogen, 100 µl of ammonium acetate and 800 µl of ethanol.

RNase A digestion of first strand cDNA was done by re-suspending the first strand cDNA in 200 µl of digestion buffer (10 mM Tris-HCl (pH 7.5), 250 mM NaCl). 12.5 ng of RNase A (5 ng/µg mRNA) were added and the mixture was mixed well. The RNase A digestion mixture was incubated at 25°C for 30 minutes and extracted with phenol/chloroform once. The supernatant was precipitated with 1 µl of glycogen, 100 µl of ammonium acetate and 800 µl of ethanol.

Example 4

Enrichment of the full-length cDNA clones with cap-binding proteins

This example describes enrichment of full-length cDNA clones with the cap-binding protein eIF4E.

cDNA was prepared by precipitating the RNase I treated first strand cDNA described in Example 3 above and washing with 70% ethanol. The resulting pellet was dried at room temperature for 5 minutes, and re-suspended in 210 µl of 10 mM KPO_4, 100 mM KCl, 2 mM EDTA, 6 mM DTT and 5% glycerol. The cDNA was stored on ice.

eIF4E glutathione sepharose 4B beads were prepared by first mixing glutathione sepharose 4B beads (Pharmacia, Sweden) well. To prepare eIF4E beads, a recombinant host cell expressing GST tagged eIF4E protein (the eIF4E gene was cloned into a GST fusion vector to create a N-terminal GST-eIF4E fusion gene) was grown and the fusion protein was purified by standard

techniques. Thus, the invention also relates to recombinant host cells expressing eIF4E protein (particularly as fusion proteins), to vectors comprising the genes expressing such proteins or fusion proteins and to the recombinant proteins or fusion proteins produced. In the present invention any tag can be used (e.g., His Tag, GST tag, HA tag, *Trx* tag, etc.). Such tags may be positioned at the carboxy and/or N-terminal region of the eIF4E gene.

The GST-eIF4E fusion protein was complexed with sepharose 4B beads by glutathione coupling using gluthionine sepharose 4B beads (Pharmacia Biotech) following the manufacturers protocols. 200 µl of the beads were transferred to a 1.5 ml microcentrifuge tube, centrifuged for 1 second, and 75 µl of supernatant was removed. The beads were washed twice with 1 ml of reaction buffer (10 mM KPO_4, 100 mM KCl, 2 mM EDTA, 6mM DTT and 5% glycerol), and re-suspended in 258 µl of reaction buffer, followed by the addition of 42 µl (18 pmoles/µl) of eIF4E protein (600 pmoles/100µl beads). The mixture was mixed on a head to head roller at 4°C for 30 minutes. The mixture was then centrifuged for 1 second, and the supernatant was removed. The beads were washed twice with 1 ml of reaction buffer and once with 1 ml of 25 µg/ml yeast tRNA in reaction buffer. 20 µl of reaction buffer and 5 µg of yeast tRNA were then added to the beads. 200 µl of RNase I treated cDNA was added to the beads, and the content was mixed on a roller at room temperature for 1 hour. After 1 hour, the mixture was centrifuged for 1 second, and the supernatant was removed. The beads were washed twice with 1 ml of reaction buffer and once with 1 ml of 500 µM GDP in reaction buffer. The cDNA was eluted twice with 250 µl of 500 µM GDP in reaction buffer. The eluted solutions were pooled and centrifuged for 1 minute to remove the beads. The eluted cDNA was extracted twice with an equal volume of phenol/chloroform. The cDNA was divided into two tubes and precipitated with 1 µl of glycogen, 0.5 volume of 7.5 M ammonium acetate and 2.5 volume of ethanol.

Example 5

Evaluation of the cDNA library

To evaluate the quality of the cDNA libraries constructed with the above-described full-length methods, the MAP4 gene (5-6 kb) and other genes was selected as the target genes. MAP4 and other cDNA clones were isolated from libraries constructed by standard methods well-known in the art (see SuperScript™ Plasmid Manual, Life Technologies, Inc.) and the above-described full-length methods with 3' and 5' GeneTrapper cDNA Positive Selection System (Life Technologies, Inc., Rockville, Maryland). The positive clones were size analyzed by PCR. Tables 6 and 7 below summarizes the results of the enrichment of full-length cDNA clones in human fibroblast cDNA libraries constructed with methods well-known in the art (control) and the full-length methods described above (full-length method).

Table 6

Gene	% full-length with 5' GeneTrapper		% full-length with 3' GeneTrapper	
	control*	full-length method	control*	full-length method
MAP4 (5-6 kb)	12.8	90.3	6.25	37.5

The control library was constructed with SS II RT using known methods.

Table 7

Gene name	Full-length of gene (Kb)	% Full-length by 5' GeneTrapper
MAP4 (Microtubule-associated protein 4)	5/6	90.3
β-Adaptin*	3.8/5.7	90.0
TR (Transferrin receptor)	5.0	45.0
PTK (Protein tyrosine Kinase)	3.0	84.4
RPA (DNA Replication protein A)	1.4	98.0

* There are two members of the genes, 3.8 kb and 5.7 kb in the family.

These results show that the full-length methods described above yielded >90% full-length cDNA clones with the 5' GeneTrapper system, compared to <13% using standard methods. Furthermore, the above-described full-length methods yielded >37% full-length clones with the 3' GeneTrapper system, as compared to <7% using standard methods.

Example 6

First strand cDNA synthesis, RNase I digestion and eIF-4E capture

All conditions and parameters described above in Examples 2, 3 (RNase I) and 4 were followed, except for the following: 4 reactions of 10 μg of human fibroblast cytoplasmic mRNA were used per reaction (see WO 98/45311); the biotinylated primer-adapter $(Biotin)_4$-GACTAGTTCTAGATCGCGAGCGGCCGCCC$(T)_{25}$ was used at a 1:1 primer/mRNA molar ratio; TS II RT was used at 50°C; and SS II RT was used at 45°C. Table 8 below summarizes the first strand cDNA and eIF-4E capture results.

Example 7

Second strand cDNA synthesis

Second strand cDNA was synthesized by first dissolving each of the four reaction pellets obtained in Example 6 above in 104 μl of DEPC-treated water and then adding the following reagents to each reaction:

4 μl of 5X First Strand Buffer*

30 μl of 5X Second Strand Buffer*

2 μl of 0.1 M DTT

4 μl of 10 mM dNTPs

1 μl of *E. coli* DNA ligase (10 units/μl)

1 μl of *E. coli* RNAse H (2 units/μl)

4 µl of *E. coli* DNA polymerase (10 units/µl)

see SuperScript Plasmid System manual (Life Technologies, Inc., Rockville, Maryland)

These reactions mixtures were then incubated for 2 hours at 16°C. 2 µl of T4 DNA polymerase (5 units/µl) was added and incubation at 16°C was continued for 5 more minutes.

Example 8

Streptavidin Bead Preparation

During the last 30 minutes of the 2 hour second strand reaction described in Example 7 above, streptavidin paramagnetic beads were prepared as follows.

Streptavidin paramagnetic beads (Seradyn) were gently mixed by pipetting until the beads were completely re-suspended. 150 µl of the mixed beads were transferred to the bottom of a microcentrifuge tube for each reaction. The tubes were inserted into a Magna-Sep Magnetic Particle

Separator (Life Technologies, Inc., Rockville, Maryland) (the magnet) and let sit for 2 minutes. While the tubes were in the magnet, the supernatant was removed by pipetting and 100 µl of TE buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA) was immediately added to the beads.

The tubes were then removed from the magnet and the beads were gently re-suspended by finger tapping or vortexing at the lowest setting. The tubes were re-inserted into the magnet. After 2 minutes, the supernatant was removed, the beads were re-suspended in 160 µl of binding buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 1 M NaCl) and the tubes were placed into a microcentrifuge tube rack.

Example 9

Capture of the double-stranded cDNA library

After incubating the second strand reaction with T4 DNA polymerase as described in Example 7 above, the reaction mixtures were placed on ice and 10 µl of 0.5 M EDTA was added. Then the cDNA library was captured according to the following procedure (see generally WO 98/51699).

The paramagnetic beads prepared according to Example 8 were transferred to the second strand reaction mixture tubes and gently mixed by pipetting and the suspension was incubated for 60 minutes at room temperature. The tubes were then inserted into the magnet. After 2 minutes, the supernatant was removed and discarded.

100 µl of wash buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 500 mM NaCl) was added to the beads, the beads were re-suspended by finger tapping or gently vortexing at the lowest setting and the tubes were re-inserted into the magnet for 2 minutes. The supernatant was removed and discarded. This washing step was repeated one more time and then 100 µl of wash buffer was added to the beads. The tubes were then again inserted into the magnet for 5 minutes.

Example 10

Not *I Digestion*

After the 5 minute incubation described in the last step of Example 9, the supernatant was removed and discarded from the paramagnetic beads and 41 µl of autoclaved, distilled water, 5 µl of REact 3 buffer, 4 µl of *Not* I was added and the beads were mixed well by pipetting. The reaction was then incubated for 2 hours at 37°C. The tubes were then inserted into the magnet for 2 minutes and the supernatant containing the cDNA library was transferred to fresh tubes.

50 µl of phenol:chloroform:isoamyl alcohol (25:24:1) was added to the supernatant, the solution was vortexed thoroughly, and then centrifuged at room temperature for 5 minutes at 14,000 x g. 45 µl of the upper, aqueous layer was carefully removed and transferred to fresh microcentrifuge tubes. 23 µl of 7.5 M ammonium acetate, 1 µl of glycogen (20 µg) and 172 µl of ethanol (-20°C) was added. The solution was mixed well and stored on dry ice (or -70°C freezer) for 15 min.

The ethanol solution was then centrifuged at 4°C for 30 minutes at 14,000 x g. The supernatant was carefully removed from the small pellets. 100 µl of 70% ethanol was added and the tubes were centrifuged at room temperature for 2 minutes at 14,000 x g. The ethanol was removed and the pellets were dried in a speed-vac for 2 minutes or until dry. The pellets were then dissolved in 20 µl of TE buffer (10 mM Tris-HCl (pH 7.5), 0.1 mM EDTA). The final yield of cDNA was determined by the Cerenkov counts (see Table 8 below).

Table 8

Reverse Transcriptase	standard (S) or varied (V) temperature	% Incorporation (ng of cDNA)	Amount of cDNA after eIF-4E capture
TS II RT	S	27% (2,720 ng)	512 ng
TS II RT	V (hot start)	26% (2,640 ng)	473 ng
SS II RT	S	46% (4,560 ng)	306 ng
SS II RT	V (hot start)	47% (4,730 ng)	363 ng

Example 11

Ligation of cDNA to the vector and introduction into E. coli

From 10 to 30 ng of the un-fractionated or size fractionated (≥1.5 kb by low melting gel electrophoresis) cDNA was ligated into a vector pCMVSPORT 6 (Life Technologies, Inc.). This ligation was introduced into *E. coli* by electroporation as described in the SuperScript Plasmid System

manual (Life Technologies, Inc., Rockville, Maryland), except that the cloning vector was pre-digested with *Not* I and Eco RV.

Sequence analysis of randomly selected clones from the cDNA library constructed (304 clones) were analyzed by 5' and 3' sequencing to determine the total percentage of full-length random clones in the cDNA library. Sequences were compared for homology with GeneBank sequences. The results are summarized in Table 9 below. Based on the results, approximately 68% of the random clones were full-length (including known full-length clones and unknown full-length clones). Thus, approximately 17% unknown full-length clones were obtained from the human fibroblast cytoplasmic mRNA library.

Table 9

	Number of Clones	Percentage
Total Sequences	304	73.3%
Sequences with Homology	223	51%
Full-Length Clones	114	17%
Potentially Full-Length	39	17%
Partial Clones	70	31%

Example 12

RNAse Assay

First strand cDNA was treated with RNase A at 1000 ng/μg mRNA in TE buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA) and RNase I 25 to 40 u/μg mRNA in TEN (10 mM Tris-HCl (pH 7.5), 5 mM EDTA (pH 8.0), 200 mM Sodium Acetate) at 37°C essentially as described in Example 3. However, this treatment with large amounts of RNase at elevated temperatures resulted in libraries containing very small average cDNA insert size (about 200 bp). Therefore, a second strand cDNA assay was developed to determine the optimal amount of RNase needed.

First strand cDNA (radioactively labeled and non-radioactively labeled) was synthesized using HeLa mRNA at 500 ng of RNA/reaction. The first strand cDNA was precipitated with ethanol and dissolved in DEPC-treated water. The cold first strand cDNA was added to RNase buffer with different amounts of RNase. After incubation for 30 minutes at 25°C, the treated cDNA was extracted with phenol:chloroform and precipitated with ethanol. The treated cDNA was dissolved in DEPC-treated water, a second strand cDNA reaction was performed with ^{32}P-dCTP plus and minus RNase H. The reaction was extracted with phenol:chloroform and precipitated with ethanol. Equal amounts of cpm was electrophoresed into a 1.4% alkaline-agarose gel. The results are shown in Figure 5 and 6.

Figure 5 is an autoradiograph of second strand cDNA synthesized using different amounts of RNase A. Lane M is the 1 kb DNA ladder. Lane 1 represents untreated first strand cDNA. Lane 2 represents untreated second strand cDNA. Lanes 3, 5, 7 and 9 represent second strand cDNA synthesized without RNase H and with RNase A concentrations of 0, 1.25 ng, 2.5 ng and 5 ng, respectively. Lanes 4, 6, 8 and 10 represent second strand cDNA synthesized with RNase H and with RNase A at concentrations of 0, 1.25 ng, 2.5 ng and 5 ng, respectively.

Figure 6 is an autoradiograph of second strand cDNA synthesized using different amounts of RNase I. Lane M is the 1 kb DNA ladder. Lane 1 represents untreated first strand cDNA. Lane 2 represents untreated second strand cDNA. Lanes 3, 5, 7 and 9 represent second strand cDNA synthesized without RNase H and with RNase I concentrations of 0, 0.5 u, 1.25 u and 2.5 u, respectively. Lanes 4, 6, 8 and 10 represent second strand cDNA synthesized with RNase H and with RNase I at concentrations of 0, 0.5 u, 1.25 u and 2.5 u, respectively.

These gel analysis demonstrated that a concentration of 1.25 ng of RNAse A (see Figure 5) or 0.5 units of RNAse I (see Figure 6) may be optimal to use with 500 ng of starting mRNA.

Example 13

Preparation of antibodies against cap structure

The antibody to cap was generated using m7guanosnine-KLH as the antigen. 1200 hybridomas were plated and only 120 colonies were generated. Of these only 6 colonies were positive for cap. After further analysis, 3 were determined to have the affinity required. The first screen ELISA consists of binding m7guanosine-BSA to an ELISA plate, block with BSA, bind hybridoma supernatants, react with secondary antibody and determine positives via a colorimetric reaction with BCIP/NPT. The secondary screen included incubating appropriate dilutions of the hybridoma supernatants with either 0.1 mM m7GTP, 0.1 mM cap analog $M^7G^{5'}ppp^{5'}G$, 0.5 mM m7guanosine or 0.5 mM GTP. The pretreated supernatant was then used in the standard ELISA procedure. The GTP did not compete with the m7guanosine-BSA whereas the m7 versions all competed efficiently.

Having now fully described the present invention in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious to one of ordinary skill in the art that the same can be performed by modifying or changing the invention within a wide and equivalent range of conditions, formulations and other parameters without affecting the scope of the invention or any specific embodiment thereof, and that such modifications or changes are intended to be encompassed within the scope of the appended claims.

All publications, patents and patent applications mentioned in this specification are indicative of the level of skill of those skilled in the art to which this invention pertains, and are herein incorporated by reference to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference.

What Is Claimed Is:

1. A method for synthesizing one or more cDNA molecules or population of cDNA molecules comprising:

mixing at least one mRNA or poly A RNA template or population of such templates with at least one polypeptide having reverse transcriptase activity; and

incubating said mixture under conditions sufficient to increase the amount or percentage of full-length cDNA molecules synthesized.

2. The method of claim 1, wherein said conditions reduce or substantially reduce internal priming.

3. The method of claim 1, wherein said polypeptide is a reverse transcriptase selected from the group consisting of M-MLV RT, RSV RT, AMV RT, RAV RT, MAV RT, and HIV RT, and derivatives, fragments, mutations and variants thereof.

4. The method of claim 3, wherein said reverse transcriptase is reduced or substantially reduced in RNase H activity.

5. The method of claim 2, wherein said conditions comprise annealing or hybridizing one or more primers to said templates at elevated temperatures.

6. The method of claim 5, wherein said elevated temperature ranges from about 20°C to about 90°C.

7. The method of claim 2, wherein said conditions comprise lowering the amount of primer relative to the amount of said template.

8. The method of claim 7, wherein the ratio of said primer to said template ranges from about 5:1 to about 1:20.

9. The method of claim 1, wherein said conditions comprise the use of an inhibitor of the polypeptide having reverse transcriptase activity.

10. The method of claim 9, wherein said inhibitor is an antibody or antibody fragment.

11. The method of claim 10, wherein said antibody or antibody fragment is polyclonal or monoclonal.

12. The method of claim 2, wherein said conditions comprise the use of a primer having a high specificity.

13. The method of claim 2, wherein said conditions comprise increasing the length of said primer.

14. The method of claim 13, wherein the length of said primer which hybridizes to said template ranges from about 20 bases to about 60 bases.

15. The method of claim 1, wherein said method further comprises incubating at least one of said cDNA molecules under conditions sufficient to make at least one second nucleic acid molecule complementary to all or a portion of said at least one cDNA molecule, thereby producing one or more double stranded cDNA molecules.

16. The method of claim 15, wherein said conditions for making said second nucleic acid molecule increases the amount or percentage of full-length double stranded cDNA molecules.

17. The method of claim 16, wherein said conditions comprise optimizing ribonuclease digestion.

18. The method of claim 17, wherein said conditions allow digestion of single stranded mRNA contained in mRNA/cDNA hybrids formed after first strand cDNA synthesis.

19. The method of claim 18, wherein said conditions prevent, inhibit, reduce or substantially reduce digestion of mRNA in the double stranded mRNA/cDNA hybrid.

20. The method of claim 17, wherein said ribonuclease is selected from the group consisting of RNase A and RNase I, or combinations thereof.

21. A cDNA molecule or population of cDNA molecules made according to the method of claim 1.

22. A cDNA molecule or population of cDNA molecules made according to he method of claim 15.

23. A vector comprising the nucleic acid molecule of claim 21.

24. A vector comprising the nucleic acid molecule of claim 22.

25. A host cell comprising the vector of claim 23.

26. A host cell comprising the vector of claim 24.

27. A host cell comprising the molecule of claim 21.

28. A host cell comprising the molecule of claim 22.

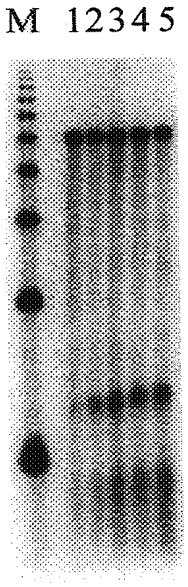
29. A kit for making an increased amount or percentage of full-length cDNA comprising at least one component selected from the group consisting of one or more primers, one or more reverse transcription inhibitors, one or more reverse transcription enzymes, one or more nucleotides, one or more cap binding molecules, one or more reverse transcription buffers and instructions for making full-length cDNA.

30. A composition for making an increased amount or percentage of full-length cDNA.

31. An antibody or fragment thereof which specifically binds to a polypeptide having reverse transcriptase activity.

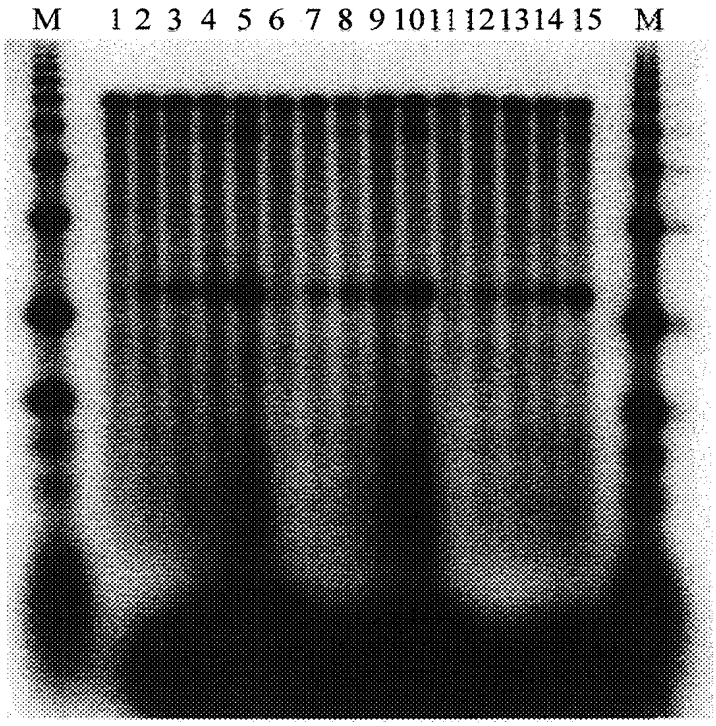
32. An antibody or fragment thereof which specifically binds to a mRNA cap structure.

Figure 1



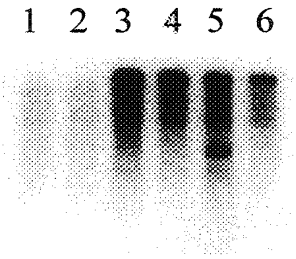
IDT Exhibit 1020.8167

Figure 2



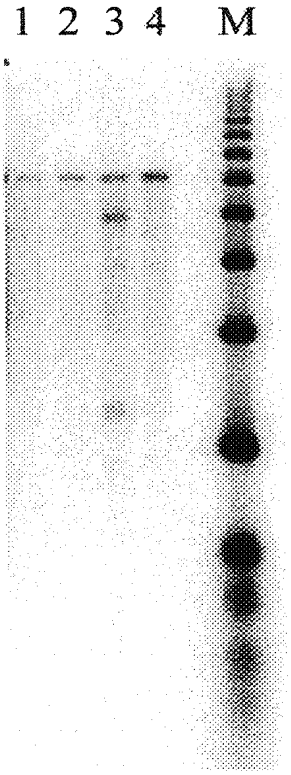
IDT Exhibit 1020.8168

Figure 3



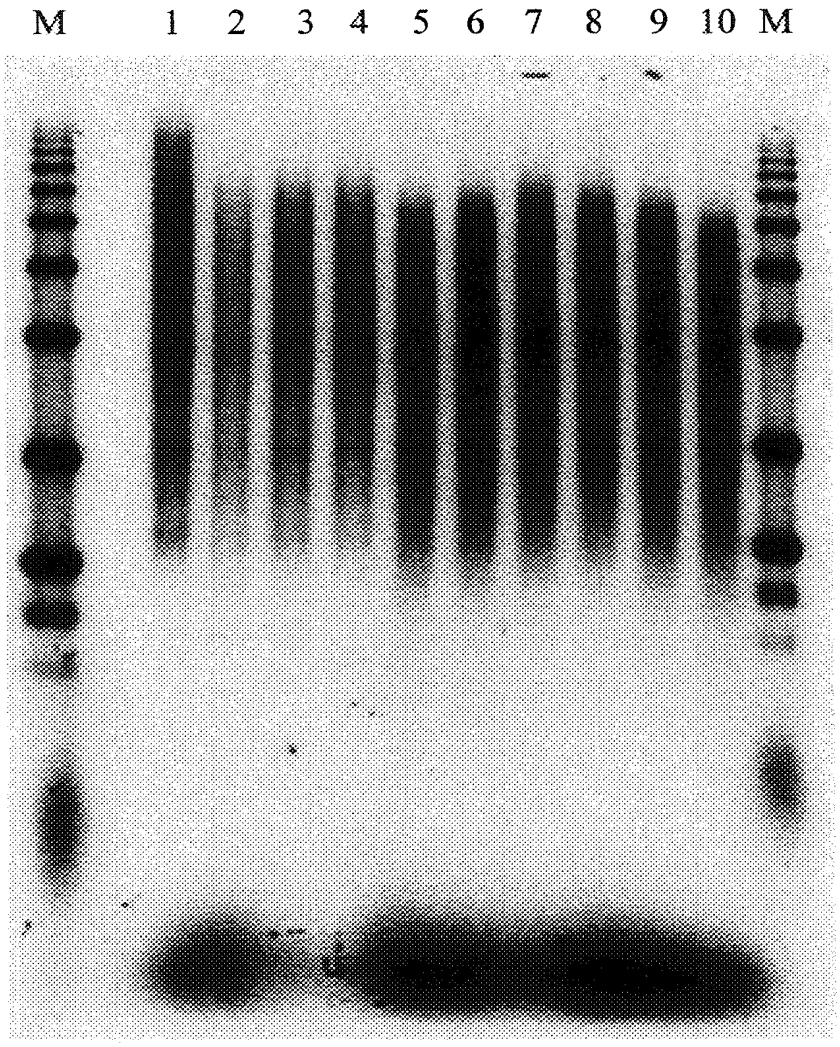
IDT Exhibit 1020.8169

Figure 4



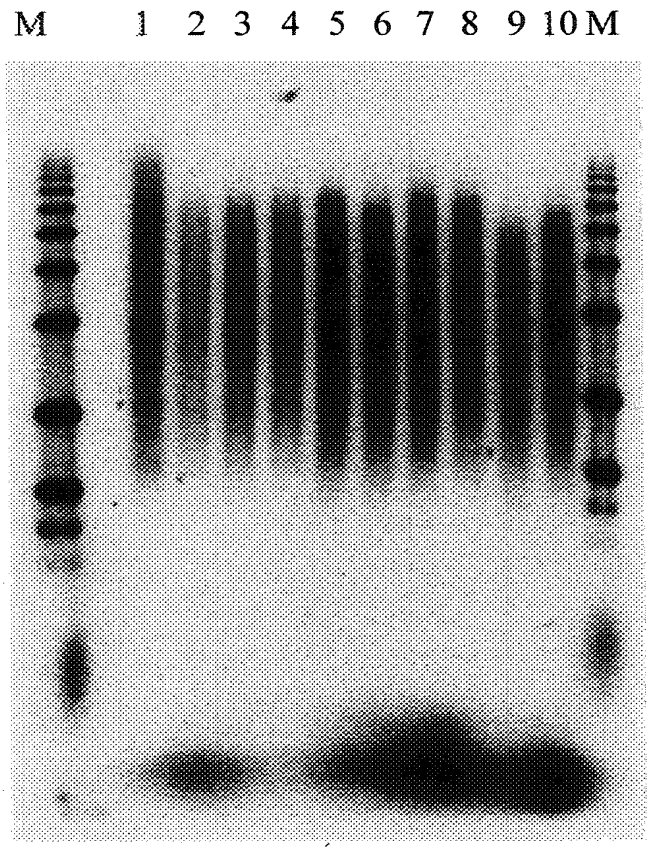
IDT Exhibit 1020.8170

Figure 5



IDT Exhibit 1020.8171

Figure 6



IDT Exhibit 1020.8172

SEQUENCE LISTING

<110> Life Technologies, Inc.

<120> cDNA Synthesis Improvements

<130> 0942.487PC01

<140>
<141>

<150> US 60/122,395
<151> 1999-03-02

<160> 1

<170> PatentIn Ver. 2.0

<210> 1
<211> 54
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 1
gactagttct agatcgcgag cggccgccct ttttttttt ttttttttt tttt 54

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/05138

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) :C12P 19/34
US CL : 435/91.1
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/91.1,435/91.4, 435/91.3, 435/91.5

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

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

biosis caplus japio medline embase

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	SAMBROOK et al. Molecular Cloning, A Laboratory Mannual, Second Edition, 1989, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.	1-28, 31-32
A	MORRIS et al. A novel strategy for directional cloning of random primed cDNA. In Novations, Novagen, Inc. June 1995, No. 3, pages 1-8.	1-22, 29-30
A	MIERENDORF et al. A high efficiency cDNA Cloning vector for protein expression, purification and substractive Probe Synthesis. In Novations, Nova gen, Inc. June 1995, No. 3, pages 1-3.	23-24

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search
21 APRIL 2000

Date of mailing of the international search report
25 MAY 2000

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
JAYAVANTHI SATISH
Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)★

IDT Exhibit 1020.8174

Espacenet

Bibliographic data: WO02060318 (A2) — 2002-08-08

DETECTION OF METHYLATED CPG RICH SEQUENCES DIAGNOSTIC FOR MALIGNANT CELLS

Inventor(s): PLASS CHRISTOPH [US] ± (PLASS, CHRISTOPH)

Applicant(s): UNIV OHIO STATE RES FOUND [US]; PLASS CHRISTOPH [US] ± (THE OHIO STATE UNIVERSITY RESEARCH FOUNDATION, ; PLASS, CHRISTOPH)

Classification: - international: ***C12N15/09; C12Q1/68; G01N27/447; G01N33/53; G01N33/574;*** (IPC1-7): A61B
- cooperative: **C12Q1/6827; C12Q1/6886;** C12Q2600/154; C12Q2600/158 more

Application number: WO2002US03077 20020131 Global Dossier

Priority number (s): US20010775398 20010131

Also published as: WO02060318 (A3) CA2435770 (A1) EP1373567 (A2) EP1373567 (A4) JP2004524837 (A) more

Abstract of WO02060318 (A2)

The present invention provides methods for determining the methylation status of CpG-containing dinucleotides on a genome-wide scale using infrequent cleaving, methylation sensitive restriction endonucleases and two-dimensional gel electrophoretic display of the resulting DNA fragments. Such methods can be used to diagnose cancer, classify tumors and provide prognoses for cancer patients. The present invention also provides isolated polynucleotides and oligonucleotides comprising CpG dinucleotides that are differentially methylated in malignant cells as compared to normal, non-malignant cells. Such polynucleotides and oligonucleotides are useful for diagnosis of cancer. The present invention also provides methods for identifying new DNA clones within a library that contain specific CpG dinucleotides that are differently methylated in cancer cells as compared to normal cells.

IDT Exhibit 1020.8175

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 August 2002 (08.08.2002)

PCT

(10) International Publication Number
WO 02/060318 A2

(51) International Patent Classification[7]: A61B

(21) International Application Number: PCT/US02/03077

(22) International Filing Date: 31 January 2002 (31.01.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/775,398 31 January 2001 (31.01.2001) US

(71) Applicant *(for all designated States except US)*: **THE OHIO STATE UNIVERSITY RESEARCH FOUNDATION** [US/US]; 1960 Kenny Road, Columbus, OH 43210-1036 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **PLASS, Christoph** [DE/US]; 90 Winthrop Road, Columbus, OH 43214 (US).

(74) Agent: **DOCHERTY, Pamela, A.**; Calfee, Halter & Griswold LLP, 1400 McDonald Investment Center, 800 Superior Avenue, Cleveland, OH 44114 (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 02/060318 A2

(54) Title: DETECTION OF METHYLATED CPG RICH SEQUENCES DIAGNOSTIC FOR MALIGNANT CELLS

(57) Abstract: The present invention provides methods for determining the methylation status of CpG-containing dinucleotides on a genome-wide scale using infrequent cleaving, methylation sensitive restriction endonucleases and two-dimensional gel electrophoretic display of the resulting DNA fragments. Such methods can be used to diagnose cancer, classify tumors and provide prognoses for cancer patients. The present invention also provides isolated polynucleotides and oligonucleotides comprising CpG dinucleotides that are differentially methylated in malignant cells as compared to normal, non-malignant cells. Such polynucleotides and oligonucleotides are useful for diagnosis of cancer. The present invention also provides methods for identifying new DNA clones within a library that contain specific CpG dinucleotides that are differently methylated in cancer cells as compared to normal cells.

IDT Exhibit 1020.8176

Detection of Methylated CpG Rich Sequences Diagnostic for Malignant Cells

This invention was conducted, at least in part, with government support under National Institutes of Health Grants No: P30 CA16058 and CA80912 awarded by the National Cancer Institute. The U.S. government has certain rights in the invention.

Background of the Invention

Diagnosis of cancer, classification of tumors, and cancer-patient prognosis all depend on detection of properties inherent to cancer, or malignant cells, that are absent in normal, nonmalignant cells. Since cancer is largely a genetic disease, resulting from and associated with changes in the DNA of cells, one important method of diagnosis is through detection of related changes within the DNA of cancer cells. Such changes can be of two types. The first type of change is a genetic change that occurs when the sequence of nucleotide bases within the DNA is changed. Base changes, deletions and insertions in the DNA are examples of such genetic changes. The second type of change in the DNA is an epigenetic change. Epigenetic changes do not result in nucleotide sequence changes, but rather, result in modification of nucleotide bases. The most common type of epigenetic change is DNA methylation.

In mammalian cells, DNA methylation consists exclusively of addition of a methyl group to the 5-carbon position of cytosine nucleotide bases. In the process, cytosine is changed to 5-methylcytosine. Cellular enzymes carry out the methylation events. Only cytosines located 5' to guanosines in CpG dinucleotides are methylated by the enzymes in mammalian cells. Such CpG dinucleotides are not distributed randomly throughout the genome. Instead, there are regions of mammalian genomes which contain many CpG dinucleotides, while other areas of the genome contain few CpG dinucleotides. Such CpG-rich areas of the genome are called "CpG islands." Most often, CpG islands are located in the transcriptional promoter regions of genes.

Not all CpG islands are methylated. However, the methylation status of CpG islands (i.e., whether the CpG dinucleotides within a particular CpG island are methylated or not) is relatively constant in cells. Nevertheless, the pattern of CpG island methylation can change and, when it does, often a new, relatively stable methylation pattern is established. Such changes in methylation of CpG islands can be either increases or decreases in methylation.

IDT Exhibit 1020.8177

Methylation of CpG islands in the promoter region of a few specific genes has been observed in some types of human cancer. However, at present it is still uncertain whether the methylation status of multiple CpG islands in the genomic DNA of subjects suspected of having cancer can be used as a diagnostic tool for determining whether or not tissue obtained from such subjects contain malignant cells.

Summary Of The Invention

The present invention relates to methods for identifying CpG islands which are diagnostic of one or more cancers in a subject. The method employs restriction landmark genomic scanning (RLGS) techniques and comprises separately digesting genomic DNA which has been obtained from malignant cells derived from a particular tumor tissue and genomic DNA which has been obtained from control cells derived from healthy tissue with an infrequently cutting restriction enzyme that is not capable of cleaving methylated recognition sites to provide a first set of DNA restriction fragments from the tumor tissue, referred to hereinafter as "malignant cell restriction fragments", and a first set of DNA restriction fragments from the healthy tissue, referred to hereinafter as "control cell restriction fragments"; attaching a detectable label to the ends of the malignant and control cell restriction fragments; digesting the labeled malignant and control cell restriction fragments with a second restriction enzyme; separating each set of restriction fragments on a gel; digesting the restriction fragments in each of the gels with a third more frequently cutting restriction enzyme; electrophoresing each set of restriction fragments in a direction perpendicular to the first direction to provide a first pattern of detectable malignant cell restriction fragments and a second pattern of detectable control cell restriction fragments; and comparing the second pattern to the first pattern to identify control cell restriction fragments, hereinafter referred to as "diagnostic fragments", which are absent, or exhibit an decreased intensity of label in the first pattern. Such fragments comprise CpG islands that are methylated in the malignant cells. Such patterns are useful for characterizing tissue which is suspected of containing malignant cells. Preferably, each of the diagnostic fragments is then isolated and sequenced, at least in part. In one preferred embodiment, the first restriction enzyme is NotI. In another preferred embodiment, the first restriction enzyme is AscI. Advantageously, the present method permits the detection of numerous methylation sites within the entire genome. In accordance with the present method, applicants have determined that particular CpG islands are preferentially methylated in DNA obtained from tumor tissues of subjects diagnosed as having breast cancer, glioma, acute myeloid leukemia,

IDT Exhibit 1020.8178

primitive neuroectodermal tumors of childhood, colon cancer, head and neck cancer, testicular cancer, and lung cancer.

The present invention also provides isolated polynucleotides, referred to hereinafter as "CpG diagnostic polynucleotides", and isolated oligonucleotides referred to hereinafter as "CpG diagnostic oligonucleotides", which are useful for characterizing tissue samples obtained from a subject suspected of having gliomas, acute myeloid leukemia, primitive neuroectodermal tumors of childhood, or cancer of the breast, colon, head and neck, testicle or lung. The CpG diagnostic polynucleotides and oligonucleotidess both comprise a sequence which contains CpG islands that have been shown to be preferentially methylated in DNA that has been obtained from malignant cells of subjects diagnosed as having breast cancer, glioma, acute myeloid leukemia, primitive neuroectodermal tumor of childhood, colon cancer, head and neck cancer, testicular cancer or lung cancer. The CpG diagnostic polynucleotides are from 35 to 3000, preferably, 35 to 100 nucleotides in length, and comprise from 15 to 34, preferably 18 to 25 of the consecutive nucleotides contained with the sequences depicted in the accompanying DNA sequence listing, or sequences which are complementary thereto. The CpG diagnostic polynucleotides comprise two or, preferably, more CpG dinucleotides or dinucleotides which are complementary thereto. The CpG diagnostic oligonucleotides are from 15 to 34 nucleotides in length and comprise from 15 to 34 consecutive nucleotides contained within the sequences depicted in the sequence listing, or sequences which are complementary thereto. The CpG oligonucleotides comprises two or more CpG dinucleotides, or dinucleotides which are complementary thereto.

The present invention also relates to methods which employ the CpG diagnostic polynucleotides and oligonucleotides of the present invention to characterize tissue from patients suspected of having cancer. Such methods are based on the methylation status of CpG islands that have been shown to be preferentially methylated in DNA that has been obtained from tumor tissues of subjects diagnosed as having breast cancer, glioma, acute myeloid leukemia, primitive neuroectodermal tumor of childhood, colon cancer, head and neck cancer, testicular cancer and lung cancer. In one method, DNA which is isolated from suspected tumor tissue from a subject is digested into smaller fragments and reacted with a CpG diagnostic polynucleotide under stringent hybridization conditions. The reaction products are then assayed to determine the size or the sequence of the DNA fragment with which the CpG diagnostic polynucleotide has hybridized. The size or the sequence of the DNA fragment to which the CpG diagnostic polynucleotide has hybridized, hereinafter referred to as the "target DNA fragment", indicates whether the target DNA fragment

IDT Exhibit 1020.8179

comprises methylated or non-methylated CpG islands. The presence of methylated CpG islands in the target DNA fragment indicates that the DNA has been obtained from a tumor or neoplasm for which the diagnostic CpG polynucleotide serves as a diagnostic marker.

In another method the DNA from the suspected tumor tissue is treated with a chemical compound which converts non-methylated cytosines to a different nucleotide base. An example of such a compound is sodium bisulfite which converts non-methylated cytosines to uracil. The DNA is then reacted with at CpG diagnostic oligonucleotides under conditions which permit the CpG diagnostic oligonucleotide to hybridize with a complementary sequence in the DNA, referred to hereinafter as the "target sequence". The DNA is also reacted with a modified CpG diagnostic oligonucleotide. The modified CpG diagnostic oligonucleotide comprises a sequence that is complementary to a modified target sequence, i.e., a sequence in which the non-methylated cytosines in the target sequence are converted to a different nucleotide base, e.g. uracil, when treated with a chemical compound. The reaction products are then assayed to determine whether the DNA contains sequences which have hybridized with the CpG diagnostic oligonucleotide or with the modified CpG diagnostic oligonucleotide. Hybridization of the sample DNA with the CpG diagnostic oligonucleotide, as opposed to the modified CpG diagnostic oligonucleotide, indicates that the cytosines in the target sequence are methylated and that the DNA sample has been obtained from a tumor or neoplasm for which the CpG oligonucleotide has been shown to serve as a diagnostic marker.

The present invention also relates to a method of identifying genes whose expression is increased or decreased in cancer cells.

Brief Description of the Figures

Fig. 1. Methylation detection in restriction landmark genomic scanning (RLGS) profiles. A, Diagram of the RLGS procedure showing the quantitative nature of methylation detection on NotI fragments displayed on RLGS profiles. Methylation detection in RLGS profiles depends on the methylation sensitivity of the endonuclease activity of NotI. Differences in digestion are assessed by radiolabelling the DNA at cleaved NotI sites. Following further endonuclease digestion, two-dimensional electrophoretic separation and autoradiography, the intensity of a DNA fragment on the resultant RLGS profile quantitatively reflects the copy number and methylation status of the NotI fragment. A priori, this allows NotI fragments containing single-copy CpG islands to be distinguished from the abundant NotI fragments present in repeat elements and rDNA sequences. B, A portion of an RLGS profile from

IDT Exhibit 1020.8180

normal peripheral blood lymphocyte DNA displaying nearly 2,000 single-copy NotI fragments and 15-20 high copy-number fragments. First-dimension separation of labeled NotI/EcoRV fragments extends from right to left horizontally. Following in-gel digestion with HinfI, the fragments are separated vertically downward into a polyacrylamide gel and autoradiographed. To allow uniform comparisons of RLGS profiles from different samples and different laboratories, each fragment is given a three-variable designation (Y coordinate, X coordinate, fragment number). The central region of the RLGS profile used for all comparisons described in this invention has 28 sections (1-5 vertically and B-G horizontally; the 4G and 5G sections were excluded due to high density and lower resolution of fragments). C, Enlarged view of profile section 2D, showing the numbers assigned to each NotI fragment. D, Analysis of the GC content and CpG ratio {(number of CpGs)/(number of guanines)(number of cytosines)}(number of nucleotides analyzed) of 210 non-redundant NotI/EcoRV clones containing the NotI/HinfI fragments seen in B and in other portions of the RLGS profile. Of 210 clones, 184 clones were randomly chosen and 26 corresponded to fragments which were frequently lost from tumor profiles. CpG islands have a GC content of greater than 50% and a CpG value of 0.6 or greater, relative to bulk DNA (average CG content of 40% and CpG ratio of 0.2). Nucleotide sequences were determined with greater than 99% accuracy overall. An average of 377nt/clone were analyzed (not indicative of actual CpG island size). The average NotI/EcoRV clone size was approximately 2 kb.

Fig. 2. Fragment loss from RLGS profiles is due to methylation. Top, portions of the RLGS profiles obtained from normal tissue and from two tumors having NotI fragments with either decreased intensity or no change in intensity. Bottom, Southern-blot analysis of EcoRV (NotI: -) and EvoRV/NotI (NotI: +) restriction digested DNAs from a larger number of samples, including the samples at top. In samples without methylation in the NotI site, the probe detects a smaller fragment on double digestion with NotI and EcoRV. The quantitation from multiple Southern blots using a phosphorimager allowed the determination of a lower limit of reliable detection in RLGS profiles of 30% decreased intensity of the diploid NotI/EcoRV fragments. Presence (+) or absence (-) of the corresponding NotI fragment is indicated. N, normal tissue DNA; T, tumor tissue DNA. A, CpG-island locus 3C1 methylation in low-grade gliomas. B, CpG island locus 2C40 methylation in leukemias. C, CpG-island locus 3E24 methylation in PNETs of childhood. *, EcoRV fragment of approximately 13 kb with homology to the probe. BLAST searches using the NotI-EcoRV clone sequence identified a homologous BAC clone sequence lacking an internal NotI site,

IDT Exhibit 1020.8181

which accounts for the 13-kb fragment on the Southern blot.

Fig. 3. Heterogeneity in CpG-island methylation across tumors. RLGS profiles were generated from 98 primary human tumors and compared with profiles of either matched normal DNA (58 of 98 cases) or to multiple profiles of tissue type-matched normal DNA from unrelated individuals. Loss or decreased intensity of single-copy fragments in the tumors, relative to several neighboring unaltered NotI fragments, were detected by visual inspection of overlaid autoradiographs and confirmed in many cases by independent profiles of the same DNA samples. For each tumor type, the dot plots display the total number of methylated CpG islands (of 1,184 CpG islands analyzed) observed in each tumor. Under the assumption that the tumors are drawn from a homogeneous distribution, with all tumors having the same frequency of methylation, the loss distributions should be approximately Poisson. The colored curve represents the expected distribution. BRE, breast tumors; CLN, colon tumors; GLI, gliomas; HN, head and neck tumors; LEU, acute myeloid leukemias; PNET, primitive neuroectodermal tumors of childhood; TST, testicular tumors.

Fig. 4. Subsets of CpG islands are preferentially methylated. For each tumor type, the histograms display the number of tumors in which the particular CpG islands were methylated. Most of the 1,184 CpG islands were not methylated in any of the tumors (histogram bar at 0 is not shown), but several CpG islands were methylated in multiple tumors. The black line shows the expected distribution under the null hypothesis that the CpG islands have equal frequencies of methylation. Most of the tumor types show significant preferential methylation.

Detailed Description Of The Invention

In one aspect, the present invention relates to methods for identifying clones within a DNA library that can be used for cancer diagnosis and tumor classification, based on the methylation status of CpG dinucleotides contained within or closely adjacent to the specific clones. Such method employs methylation-sensitive restriction endonucleases (MSREs) and restriction landmark genomic scanning (RLGS) gels to identify new, differentially-methylated CpG islands within malignant cells obtained from patients diagnosed as having cancer. In accordance with the present invention, Applicants have identified 93 clones which can be used to determine whether a tumor biopsy from a patient contains benign or malignant cells.

IDT Exhibit 1020.8182

To carry out such method, tissue (referred to hereinafter as "tumor tissue") which contains a tumor or neoplasm is obtained from a patient known to have a cancer. In some cases, the tumor tissue is obtained from a particular type of solid tumor which has bee surgically removed from the patient. In some cases, the tumor tissue is obtained from the hematopoietic system, such as for example, bone marrow or blood, of the patient. The tumor tissue will have been determined to be from either a benign or malignant tumor or neoplasm.

Separately, tissue (referred to hereinafter as "healthy tissue") which does not contain a tumor or neoplasm is obtained from a subject. The healthy tissue, may be obtained by surgically removing normal tissue from the patient or by surgically removing normal tissue from a healthy control subject who does not have cancer. The healthy tissue may also come from the hematopoietic system, such as for example, bone marrow or blood, of a healthy control subject. The healthy tissue will have been determined to be non-tumorigenic or non-neoplastic.

DNA is then isolated from both the tumor tissue and healthy tissue. If the tumor tissue is a solid tissue sample, such procedure may first comprise separating the individual cells contained within the tissue from each other. For example, if the tissue samples were frozen after surgical removal from a patient, cells may be separated from one another by grinding the frozen tissue with a mortar and pestle. DNA is then isolated from the individual cells using procedures well known to those skilled in the art. Commonly, such DNA isolation procedures comprise lysis of the individual cells using detergents, for example. After cell lysis, proteins are commonly removed from the DNA using various proteases. The DNA is then commonly extracted with phenol, precipitated in alcohol and dissolved in an aqueous solution.

In the procedures which follow, the DNA obtained from the tumor tissue is treated separately from the DNA obtained from healthy tissue (i.e., the two DNAs are not mixed). The DNAs are separately analyzed using a method called restriction landmark genomic scanning (RLGS). The purpose is to analyze both DNAs separately. The two analyses are then compared in order to identify CpG islands that distinguish cancer cells from normal cells.

Both DNA samples are treated with restriction enzymes and the free ends that result from the restriction enzyme cleavage are labeled. However, since the isolated DNA is in linear pieces, there are free ends that exist before the DNA is cleaved with the restriction enzymes. To prevent these ends from being labeled, the ends, preferably, are blocked before restriction enzyme treatment. Such blocking can be done by addition of dideoxynucleotides

IDT Exhibit 1020.8183

and sulfur-substituted nucleotides to the free ends before treatment with restriction enzymes. Subsequently, when the DNA is cleaved by restriction enzymes and labeled, only the ends resulting from the restriction enzyme cleavage will be labeled.

After the reaction to block free ends, the DNA samples are cleaved with a first restriction enzyme that can be characterized as an infrequently cleaving, methylation-sensitive restriction enzyme. Examples of suitable first restriction enzymes are NotI, AscI, BssHII and EagI. As used herein the term "infrequently cleaving" refers to a restriction enzyme that is expected to cleave genomic DNA at intervals greater than 10 kilobases. For example, NotI is an infrequently cleaving restriction enzyme. NotI recognizes a nucleotide sequence of 8 base pairs (bp) in the genome (i.e., 5'GCGGCCGC3') and cleaves the DNA at this site. There are an estimated 4000-5000 of such NotI recognition sequences within the human genome. It is estimated that such recognition sequences are spaced at approximately 1 megabase (Mb) intervals within the genome. In contrast, a frequently cleaving restriction enzyme is expected to cleave the human genome at from 5-10 kb intervals. Such an enzyme will have approximately 100-times more cleavage sites within the human genome than infrequently-cleaving enzymes. Such frequently cleaving enzymes usually recognize a nucleotide sequence of less than 8 bp in the genome and cleave the DNA at that site. However, not all restriction enzymes that have nucleotide recognition sequences of less than 8 bp are frequently cleaving enzymes. BssHII and EagI both have 6 bp recognition sequences but the recognition sequences for these two enzymes are spaced at intervals within the genome that are greater than 10 kb. "Methylation sensitive" as used herein refers to any enzyme that is unable to cleave DNA at its normal restriction site if one or more nucleotides within the recognition sequence is methylated. For example, the restriction enzyme NotI will cleave the 5'GCGGCCGC3' recognition sequence if the sequence does not contain a 5-methylcytosine. However, the NotI enzyme will not cleave this sequence if any of the cytosines have been methylated to become 5-methylcytosine.

Following digestion of the DNA with the first restriction enzyme, the ends of the DNA fragments are labeled. This can be done, for example, by attachment of nucleotides carrying a detectable label, such as a radiolabel, to the ends of the DNA sample. Typically, attachment is accomplished by filling in the recessed DNA ends left by cleavage with the first restriction enzyme such that the ends become blunt (i.e., non-recessed). Such end-filling reaction may employ deoxynucleoside triphosphates having a radiolabeled phosphate at the α phosphate position. Such labeled phosphate is preferably ^{32}P.

The labeled fragments from each sample are then cleaved with a second restriction

IDT Exhibit 1020.8184

enzyme. Such second restriction enzyme preferably cleaves human DNA at average intervals of between 5-10 kb. Such enzymes normally have a 6 bp recognition sequence. Preferably, the second restriction enzyme is not methylation sensitive. Examples of suitable second restriction enzymes are PstI, PvuI, EcoRV or BamHI. Cleavage of the DNA fragments with the second restriction enzyme provides a second set of fragments, labeled at the ends left by cleavage with the first enzyme. Many of such second fragments are smaller than the fragments resulting from cleavage with the first restriction enzyme.

The DNA fragments are then separated from one another. Preferably this separation is based on size. Preferably this separation is performed by first-dimension electrophoresis through an agarose tube-shaped gel of approximately 60 cm in length.

After electrophoresis through the tube-shaped gel, the DNA is digested within the gel with a third restriction enzyme. Such third restriction enzymes preferably have recognition sequences of 4 or 6 bp. Such third restriction enzymes also have the property of being able to cleave DNA which is embedded within agarose. One such enzyme is HinfI.

After cleavage by the third restriction enzyme, the DNA is again separated based on size, preferably by electrophoresis through a polyacrylamide gel. Subsequently, the separated DNA fragments are detected based on the labeled ends of the DNA fragments. In those cases where the fragments are radiolabeld, detection is by autoradiography of the two-dimensional gel. Such autoradiography provides a pattern of DNA fragments or "spots." Such pattern is called an RLGS profile.

Each fragment on the RLGS profile obtained from using the DNA from healthy tissues is uniquely identified by its location on the autoradiograph (Y coordinate, X coordinate, fragment number). For each fragment location on the RLGS profile obtained from healthy tissue DNA, the identical location is observed on the RLGS profile obtained from tumor tissue DNA.

In a fragment by fragment comparison of RLGS profiles obtained from tumor tissue DNA with healthy tissue DNA, three different patterns are possible. First, for a given fragment on the healthy tissue RLGS profile, there may be a corresponding fragment at the same location, and of the same intensity, on the tumor tissue RLGS profile. This indicates that the first restriction enzyme cleaved both DNAs at the same sequences (Fig. 1A). This indicates that there were no differences in methylation of the NotI nucleotide recognition sequence of that fragment between the tumor tissue DNA and the healthy tissue DNA.

Second, for a given fragment on the healthy tissue RLGS profile, there may be no fragment at the same location on the tumor tissue RLGS profile. Such a pattern indicates that

IDT Exhibit 1020.8185

the first restriction enzyme did not cleave the tumor tissue DNA at the recognition sequence required to produce that specific fragment, but did cleave at such sequence within the healthy tissue DNA (Fig. 1A). This indicates that there was methylation within the NotI recognition sequence in the tumor tissue DNA but not in the healthy tissue DNA.

Third, for a given fragment on the healthy tissue RLGS profile, there may be a corresponding fragment at the same location on the tumor tissue RLGS profile, but the intensity of the fragment may be of decreased intensity. Such a pattern indicates that the first restriction enzyme cleaved one of two copies (i.e., the genome is diploid) of the tumor tissue DNA at the recognition sequence required to produce that specific fragment (Fig. 1A). In healthy tissue DNA, the first restriction enzyme cleaved both copies of the recognition sequence. This indicates that there was methylation within one of two NotI recogniton sequences in the tumor tissue DNA.

Through comparisons of RLGS profiles obtained from healthy tissue DNA with profiles obtained from a large number of different tumor tissue DNAs, loss of specific fragments in multiple tumors can be associated with a specific type of cancer. Loss of such fragments from RLGS profiles, therefore, can be diagnostic for cancer in a subject. For example, loss of a specific fragment (i.e., methylation of the first restriction enzyme site at the end of said fragment) in a high percentage of tumor tissue DNAs from women known to have breast cancer can be diagnostic for breast cancer in subjects suspected of having the disease. To perform such a diagnostic analysis, DNA isolated from a patient suspected of having breast cancer would be analyzed by RLGS, as described above, to determine whether there was loss of one or more fragments in RLGS profiles that are known to be lost at high frequency in women known to have breast cancer. Similarly, loss of other specific fragments can be diagnostic for other cancers, such as for example, colon cancer, head and neck cancer, lung cancer, testicular cancer, neuroectodermal cancer, gliomas, acute myeloid leukemias, and others.

Loss of a specific fragment in RLGS profiles from multiple tumors can also be diagnostic of several types of cancer, rather than a single type of cancer. For example, loss of a specific fragment can occur in a high percentage of tumor tissue DNAs obtained from individuals with either breast, colon or lung cancer. Loss of such a spot from RLGS profiles using DNA obtained from a patient suspected of having cancer would be diagnostic for either breast, colon or lung cancer in that patient.

Isolated Polynucleotides and Oligonucleotides Diagnostic for Cancer

Individual DNA clones that contain the DNA present in each spot or fragment that

IDT Exhibit 1020.8186

makes up an RLGS profile can be obtained. This is done by constructing a DNA library of healthy tissue DNA that has been cleaved with the same first and second enzymes used to perform the RLGS gel analysis. Such DNA library will contain individual clones, each clone comprising DNA that is present in a single spot of the RLGS profile. The totality of clones within the library is representative of the combined DNA spots in the RLGS profile.

Individual clones within the library can be identified that contain the DNA of each spot on the RLGS profile. This can be done by taking DNA from one or a few individual clones of the DNA library and mixing it with healthy tissue DNA, before RLGS analysis is begun. When this mixture of DNAs is used to produce an RLGS profile, the intensity of the spots that contain the same DNA as the individual clones added to the mixture will be increased. By performing multiple analyses of this type, each spot on an RLGS profile can be matched up with a DNA clone within the library. The result of such an analysis is an ordered human genomic library of restriction fragments containing the same subset of genomic fragments as those displayed on RLGS profiles. In such ordered genomic libraries, an individual library clone corresponding to any spot or fragment in an RLGS profile can be rapidly located.

To design diagnostic CpG polynucleotides and oligonucleotides, the sequence of the DNA within each clone (referred to hereinafter as a "diagnostic clone") that corresponds to a spot that is absent or exhibits decreased intensity on the RLGS profile of the DNA from malignant tumor tissue is sequenced using standard techniques. Once sequence information is obtained, regions comprising multiple CpG dinucleotides are located. Such regions serve as the target sequence for the CpG polynucleotides and oligonucleotides.

The CpG polynucleotides are from 35 to 3000 , preferably from 35 to 1500 nucleotides in length and comprise two or, preferably, more CpG dinucleotides or dinucleotides which are complementary thereto. The CpG diagnostic oligonucleotides are from 15 to 34 nucleotides, preferably from 18 to 25 nucleotides, in length and comprise at least two CpG dinucleotides or dinucleotides which are complementary thereto. The CpG diagnostic polynucleotides and oligonucleotides each comprise a sequence which is substantially complementary to target sequences containing CpG islands that are known to be preferentially methylated in the DNA from one or more types of cancer cells. "Substantially complementary" means that there is enough complementarity between the CpG diagnostic polynucleotides or oligonucleotides and the target sequence so that hybridization occurs between the CpG diagnostic polynucleotides and oligonucleotides under stringent conditions, preferably under highly stringent conditions. Such assays include hybridization assays, such as for example Southern

IDT Exhibit 1020.8187

analysis, where the sample DNA is reacted with the CpG diagnostic polynucleotide under stringent hybridization conditions.

The term "stringent conditions, as used herein, is the "stringency" which occurs within a range from about Tm-5 (5 below the melting temperature of the probe) to about 20 C below Tm. "Highly Stringent hybridization conditions" refers to an overnight incubation at 42 degree C in a solution comprising 50% formamide, 5x SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 g/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.2x SSC at about 65 degree C. As recognized in the art, stringency conditions can be attained by varying a number of factors such as the length and nature, i.e., DNA or RNA, of the probe; the length and nature of the target sequence, the concentration of the salts and other components, such as formamide, dextran sulfate, and polyethylene glycol, of the hybridization solution. All of these factors may be varied to generate conditions of stringency which are equivalent to the conditions listed above.

Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages of formamide result in lower stringency); salt conditions, or temperature. For example, moderately high stringency conditions include an overnight incubation at 37 degree C in a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 0.2 M NaH_2PO_4; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 ug/ml salmon sperm blocking DNA; followed by washes at 50 degree C with 1XSSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC).

Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reag ents may require modification of the hybridization conditions described above, due to problems with compatibility.

Such assays also include polymerase chain reactions (PCR) where the sample DNA and the diagnostic CpG oligonucleotides are reacted, preferably under conditions which result in the synthesis of a single PCR product. Computer programs, such as for example, the "Primer3" program that can be accessed at "http://genome.wi.mit.edu/cgi-

IDT Exhibit 1020.8188

bin/primer/primer_3www.cgi" can be used to determine the size and sequence of the CpG diagnostic oligonucleotides. Optimum conditions are determined empirically.

The CpG diagnostic polynucleotides and oligonucleotides are made using standard techniques. For example, these polynucleotides and oligonucleotides may be made using commercially available synthesizers.

Diagnostic Methods

In another aspect, the present invention relates to methods which use the CpG diagnostic polynucleotides and oligonucleotides to characterize tissue samples from a subject suspected of having cancer, referred to hereinafter as test sample DNA. To do this, DNA is isolated from the cells of the tissue sample of the patient. Preferably, DNA that serves as a control is also obtained from healthy tissue of the test subject or a control subject as described previously. The diagnostic methods comprise reacting the test sample DNA with the diagnostic CpG polynucleotide or oligonucleotide and assaying the products that are formed as the result of the reaction. In some cases, the sample DNA is digested into smaller fragments prior to reaction with the CpG diagnostic polynucleotides or oligonucleotides. In some cases, a portion of the test sample DNA is first reacted with a chemical compound, such as for example sodium bisulfite, which converts methylated cytosines to a different nucleotide base.

Southern Blot Analysis

One such method for diagnosing cancer in a patient involves cleavage of the test sample DNA with a methylation sensitive enzyme, then Southern blot analysis of said cleaved DNA using a CpG diagnostic polyncleotide or oligonucleotide as a probe. For example, the DNA from the patient and the control, healthy tissue DNA are separately cleaved with a methylation-sensitive restriction endonuclease, such nuclease being the same first restriction enzyme used to identify the diagnostic spot in the RLGS profile that corresponds to the CpG diagnostic polynucleotide or oligonucleotide. After cleavage, the test sample and control DNAs are electrophoretically separated by size in different lanes of the same agarose gel and blotted to a membrane that can be used in hybridization, such as for example, nitrocellulose or nylon. The membrane is then used in a hybridization reaction with a labeled CpG diagnostic polynucleotide or oligonucleotide. The labeled CpG diagnostic polynucleotide or oligoneucleotide will hybridize to complementary DNA sequences on the membrane. After hybridization, the location on the membrane where the probe hybridized to the control and patient DNAs is visualized. Such locations will identify DNA fragments or bands within the control and patient DNAs containing the same sequence as the CpG

IDT Exhibit 1020.8189

diagnostic polynucleotide or oligonucleotide. Hybridization of the probe to a fragment within the patient DNA that is of higher molecular weight than that of the fragment within the control DNA to which the probe hybridized, indicates that a restriction endonuclease cleavage site flanking the target sequence of the CpG diagnostic polynucleotide or oligonucleotide was not cleaved due to methylation. Such result indicates that the tissue is from a cancer for which the CpG diagnostic polynucleotide or oligonucleotide serves as a diagnostic tool.

A second method for diagnosing cancer in a patient involves cleavage of patient DNA with a methylation-sensitive restriction endonuclease, such nuclease being the same first restriction enzyme used to identify the diagnostic spot in the RLGS profile that corresponds to the fragment. Such nuclease will cleave the patient DNA at the diagnostic recognition sequence only if the DNA is unmethylated. Using nucleotide information derived from sequencing of the library clone corresponding to the diagnostic spot on the RLGS gel, primers for PCR are selected that span the diagnostic recognition sequence. Using the primers, PCR is performed on the DNA. PCR amplification of the sequences will be successful only if the diagnostic nucleotide sequence in the patient DNA had been methylated and was not cleaved by the enzyme. Successful PCR amplification, therefore, is indicative of cancer in the patient.

Methods Employing a Chemically-Modified DNA Test Sample

Another group of methods for diagnosing cancer in a patient using CpG diagnostic polynucleotides and oligonucleotides are based on treatment of patient DNA with sodium bisulfite which converts all cytosines, but not methylated cytosines, to uracil. The bisulfite converted patient DNA can then be analyzed in a number of different ways. One method of analysis is direct sequencing of the DNA to determine whether the sequence contains cytosine or uracil. Such DNA sequencing requires primers adjacent to the sequenced region to be made. Such primers would be based on DNA sequence information obtained from the diagnostic RLGS spots.

Another method of analyzing bisulfite converted patient DNA is a method called "methylation sensitive PCR" (MSR). In MSR, primers are designed to comprise a sequence which is substantially complementary to the the CpG islands which are known to be preferentially methylated in DNA of cells found in one or more type of tumor tissues. Two sets of PCR primers are made to emcompass this region. One set of primers is designed to be complementary to the sequence that was changed by bisulfite (i.e., cytosines that were originally unmethylated and changed to uracil). As discussed above, these are the modified

IDT Exhibit 1020.8190

CpG diagnostic oligonucleotides. A second set of primers is designed to be complementary to the same sequence that was not changed by bisulfite (i.e., cytosines that were methylated and not changed to uracil). As discussed above these are the unmodified CpG diagnostic oligonucleotides, i.e the oligonucleotides which containe at least two CpG dinucleotides or dinucleotides which are complementary thereto. Two sets of PCR reactions are then run, one reaction with each set of primers, using DNA from the subject as the template. In the case where cytosines within the target sequence of the subject DNA are not methylated, the target sequence will be modified by the chemical reaction and the primers complementary to the modified sequence, i.e., the modified CpG diagnostic oligonucleotides, will produce a PCR reaction product while the primers complementary to the methylated sequence, i.e., the unmodified CpG diagnostic oligonucleotides, will not produce a PCR product. In the case where cytosines within the target sequence of the subject DNA are methylated, the target sequence will not be altered by the reaction with the sodium bisulfite, and the primers complementary to the unaltered sequence, i.e., the unmodified CpG diagnostic oligonucleotides, will produce a PCR reaction product while the modified CpG diagnostic oligonucleotides, which are complementary to the modified target sequence (i.e., unmethylated sequence) will not produce a PCR product

A modification of MSR is bisulfite treatment of patient DNA and PCR amplification of said DNA using primers designed to amplify either methylated or unmethylated sequences. The PCR product is then digested with a restriction enzyme that will cleave or not depending on whether said product contains uracil (rather, thymidine, the complement of uracil; found in PCR product if original patient DNA contained unmethylated cytosine) or cytosine (found in PCR product if original patient DNA contained methylated cytosine).

Another technique referred to as MS-SnuPE, uses bisulfite/PCR followed by primer extension, where incorporation of C (vs. T) denotes methylation.

Methods of Identifying Genes

In another aspect of the invention, the CpG diagnostic polynucleotides and oligonucleotides can be used as probes to to identify genes whose expression is increased or decreased in cancerous tissues. To do this, CpG diagnostic polynuceotides are reacted with individual clones of the DNA library. The clones which hybridize with the CpG diagnostic polynucleotide can then be analyzed to determine if they contain an open reading frames that could encode proteins. To determine if the CpG diagnostic polynucleotide hybridizes with the promoter region of a known gene, the open reading frame sequence is analyzed by

IDT Exhibit 1020.8191

searching existing DNA databases. For example, GenBank databases can be searched using the BLAST algorithm. If no known genes that correspond to a library clone is found, the sequence can be searched for open reading frames that could encode a protein. Such searching can be performed using commercially available sequence analysis programs commonly known to those skilled in the art. GCG is an example of one such program.

Sequences from clones of the DNA library that contain either known genes or open reading frames can be used as probes to determine whether genes encoded by the sequences are expressed in tumor tissues as compared to control, healthy tissues. To do this, RNA, preferably messenger RNA (mRNA) is isolated from healthy tissue and from tumor tissue from which it is desired to test expression. Such RNA is examined for the presence of expressed transcripts encoded by the sequences obtained from the library. Examination for the presence of expressed transcripts can be performed using a number of methods. One method is Northern blotting where the isolated RNA is separated by size using gel electrophoresis and then blotted to a hybridization membrane. A fragment, polynucleotide or oligonucleotide from the sequence obtained from a library clone is labeled and then used to probe the hybridization membrane containing the size-separated RNA. Detection of hybridization of the probe to the membrane indicates presence of a transcript encoded by the sequence and indicates expression of the gene encoded by that sequence.

Another method to examine isolated RNA for the presence of expressed transcripts is to use RT-PCR analysis. In such analysis, primers are designed and made that span a region of the gene whose expression is to be tested. The isolated RNA is reverse transcribed into DNA using reverse transcriptase. Such DNA is then amplified with the designed primers using PCR. PCR products are visualized after electrophoresis. The presence of PCR products on the gel indicates that the gene encompassed by the designed primers was expressing RNA transcripts. Such analysis can identify and determine genes whose expression is changed in cancer cells as compared to normal, non-cancerous cells.

The following examples are for purposes of illustration only and are not intended to limit the scope of the invention as defined in the claims which are appended hereto.

Examples

IDT Exhibit 1020.8192

Example 1. Identification of diagnostic markers using NotI and RLGS

A. Isolation and enzymatic processing of genomic DNA

Tissue from solid tumors was obtained as surgical tissue samples. Where possible, surrounding non-tumor tissue was taken and used as a control. Where it was not possible to obtain patient-matched normal tissue, normal tissue from multiple patients was used. Tissue samples from patients with acute myelogenous leukemis (AML) consisted of either bone marrow aspirates or peripheral blood. Normal samples were obtained from the same patients who were in remission after chemotherapy.

The surgically removed tissues were quickly frozen in liquid nitrogen and stored at -80°C prior to isolation of DNA. When DNA was ready to be isolated, 2 ml of lysis buffer (10 mM Tris, pH 8.0; 150 mM EDTA, 1% sarkosyl) was added to 100-300 mg of tissue in a 50 ml Falcon tube and frozen in liquid nitrogen. The frozen mixture was then removed from the tube, wrapped in aluminum foil, and quickly broken into pieces with a hammer. The broken pieces of cells were transferred to a chilled mortar and ground to a powder with a chilled pestle. For peripheral blood samples, cells were separated on a sterile Histopaque-1077 (SIGMA) gradient and stored at -80°C before DNA isolation.

Cells were transferred to a 50 ml tube and 15-25 ml of lysis buffer containing 0.1 mg proteinase K per ml of lysis buffer was added and mixed using a glass rod. The mixture was incubated at 55°C for 20 min with gentle mixing every 5 min. The mixture was then placed on ice for 10 min. Subsequently, an equal volume of PCI (phenol:chloroform:isoamylalcohol in a ratio of 50:49:1) was added and the tubes containing the mixture were gently rotated for 30-60 min. The tubes were then centrifuged for 30 min at 2500 rpm and the separated, aqueous phase was transferred to a new 50 ml tube using a wide-bore pipette. The PCI extraction was repeated one time. The collected aqueous phase containing the DNA was transferred to dialysis tubing and dialyzed against 4 L of 10 mM Tris, pH 8 for 2 hr. The dialysis tubing was then transferred into fresh 10 mM Tris and dialyzed overnight at room temperature. One additional dialysis was performed in fresh 10 mM Tris for an additional 2 hr. The DNA was then transferred from the dialysis tubing to 50 ml tubes and RNase A was added to a final concentration of 1 μg/ml. The mixture was incubated at 37°C for 2 hr. Subsequently, 2.5 volumes of 100% ethanol were added to the DNA and the mixture was gently rotated. The insoluble DNA was transferred to a microfuge tube, centrifuged briefly, and the remaining alcohol removed. The pellet was briefly dried in air. The DNA in the pellet was resuspended to a final concentration of 1 μg/μl. Such isolated DNA had an

IDT Exhibit 1020.8193

average size of 200-300 kb.

The isolated genomic DNA was blocked at ends where the DNA had been sheared. Blocking was done by addition of dideoxynucleotides and sulfur-substituted nucleotides. In a 1.5 ml tube, 7 µl of genomic DNA solution was added along with 2.5 µl of blocking buffer (1µl 10X buffer 1, 0.1µl 1 M DTT, 0.4 µl each of 10 µM dGTPαS, 10 µM ddATP, 10 µM ddTTP, and 0.2 µl 10 µM dCTPαS; buffer 1 consists of 500 mM Tris, pH 7.4, 100 mM $MgCl_2$, 1 M NaCl, 10mM DTT) and 0.5 µl DNA polymerase I. The mixture was mixed thoroughly and incubated at 37°C for 20 min. The mixture was then incubated at 65°C for 30 min to inactivate the polymerase. The reaction was then cooled on ice for 2 min. The DNA was digested with NotI by adding to the sample, 8µl of 2.5X buffer 2 (20X buffer 2 is 3 M NaCl, 0.2% Triton X-100, 0.2%BSA) and 2µl (10 U/µl) of NotI. The sample was incubated at 37°C for 2 hr. The DNA was then radioactively labeled. This was done by adding to the sample 0.3 µl 1 M DTT, 1 µl [α-^{32}P]-GTP, 1 µl [α-^{32}P]-dCTP and 0.1 1 µl[α-^{32}P]-GTP Sequenase ver 2.0 (13 U/µl). The mixture was incubated at 37°C for 30 min. The DNA was then digested with EcoRV by adding to the sample 7.6 µl second enzyme digestion buffer (1 µl 1 mM ddGTP, 1 ul 1 mM ddCTP, 4.4 µl ddH_2O, 1.2 µl 100 mM $MgCl_2$) and 2 µl EcoRV (10 U/µl). The mixture was incubated at 37°C for 1 hr. Then, 7 µl of 6X first-dimension loading dye (0.25% Bromophenol Blue, 0.25% Xylene Cyanol, 15% Ficoll type 400) was added.

B. First dimension gel set-up and electrophoresis

To make the 60 cm long agarose tube-shaped gel, a gel holder was made. To do this, a sharp razor was used to cut one end of PFA-grade teflon tubing (PFA 11 thin wall, natural; American Plastic, Columbus, Ohio) at an angle to make a bevel. The beveled end of the tubing was fed into glass tubes (4 mm inner diameter, 5 mm outer diameter, 60 cm long). Using a hemostat, the beveled end was pulled up through the tapered end of the glass rod until it protruded 2 to 4 cm. The tubing was cut horizontally at the same end, leaving a 2 mm protrusion (this is the top of the gel holder). The opposite end was cut horizontally, leaving a 5 to 6 cm protrusion from the glass tube. The gel holder was inverted and the top protruding end was pressed firmly against a hot metal surface (metal spatula heated by a Bunsen burner) to fold the edges of the teflon outward onto the rim of the glass support. A rubber stopper with cored center was pulled over the top end of the gel holder until it was just past the taper of the glass rod. A two-way stopcock was attached to a 10 ml syringe and then to the gel

IDT Exhibit 1020.8194

holder via 2 to 3 cm of flexible tubing. The stopcock valve was adjusted to the open position.

Then, to a clean 200 ml glass bottle was added, 60 ml 2X Boyer's buffer (20X is 1 M Tris, 360 mM NaCl, 400 mM sodium acetate, 40 mM EDTA) and 0.48 g Seakem GTG agarose (0.8%). The mixture was heated in a microwave oven until the agarose was dissolved. The mixture was then equilibrated to 55°C in a water bath. With the stopcock valve in the open position, the protruding teflon tube was lowered into the molten agarose solution. The gel solution was suctioned into the gel holder until the gel solution reached 1-2 cm from the top of the gel holder. The stopcock valve was then closed. Keeping the gel upright, the gel was suspended from a ring stand. The gel was allowed to solidify for 20 min.

The stopcock valve was then opened and the syringe and connecting tubes were removed from each gel. After adding 2X Boyer's buffer to the bottom of the first dimension gel apparatus (C.B.S. Scientific), the gels were lowered into the first dimension gel apparatus, seating the rubber stopper firmly into the appropriate holes in the top portion of the apparatus. The top chamber was filled with 2X Boyer's buffer.

Between 1.0-1.5 μg of DNA was loaded onto each gel. The sample was electrophoresed at 110 V for 2 hr, and then 230 V for 24 hr.

C. In-gel digest

After the DNA was electrophoresed in the first dimension in the agarose tube gel, the DNA was further digested with an additional restriction endonuclease so it could be electrophoresed in the second dimension. In order to perform this additional endonuclease digestion, the buffer and gel holders were removed from the first dimension apparatus. The gel was extruded into a pan containing 1X buffer K (10X buffer K is 200 mM Tris, pH 7.4, 100 mM MgCl2, 1 M NaCl) by forcing the gel out through the bottom of the gel holder. This was accomplished using a 1 ml syringe fitted with a pipet tip and filled with buffer K. The tip was firmly inserted into the top of the gel holder and the plunger depressed until the gel began to come out through the bottom of the gel holder. The 1 ml syringe was replaced with a 5 ml syringe, and the plunger was depressed until the entire gel was expelled. With a razor, a bevel was cut in the low molecular weight end of the gel and a horizontal cut was made at the high molecular weight end so that the gel was approximately 43 cm in length. The gel length was now the same as the width of the second dimension gel.

The gel was placed into a separate 50 ml tube containing 40 ml of 1X buffer K. The tube was incubated for 10 min at room temperature. The buffer was poured off and the gel incubated in 1X buffer K for an additional 10 min. The buffer K and gel was poured into a

IDT Exhibit 1020.8195

pan containing fresh buffer K. Using a 10 ml syringe attached to restriction digest tubing (PFA grade teflon, 9, thin wall, natural; 2.7 mm inner diameter and approximately 3.3 mm outer diameter; American Plastic, Columbus, Ohio), via a 1 to 2 cm segment of flexible tubing, the gel was suctioned into the digest tubing, low molecular weight (beveled) end first. The gel was suctioned into the digest tubing by placing the end of the tubing in line with the beveled end of the gel and pulling the syringe plunger. The tubing was positioned vertically, with the syringe at the bottom and remaining buffer from the tubing was suctioned into the syringe.

In a clean tube, a 1.6 ml mix of 1X HinfI restriction enzyme buffer (50 mM NaCl, 10 mM Tris pH 7.9, 1 mM DTT), 0.1 % BSA, and 750 U of HinfI restriction enzyme was made. The open end of the digest tubing was placed into the tube containing restriction digestion solution. Holding the syringe end up, suction was applied until a small amount of digestion solution appeared in the syringe. The digest tubing was removed and both ends were oriented upward in a U-shape. The syringe was removed and the two ends of the tubing were attached to form a closed circle. This was placed in a moist chamber and incubated at 37°C for 2 hr.

D. Second dimension electrophoresis

The digested DNA was now run in the second dimension using a 5% non-denaturing acrylamide gel with a 0.8% agarose spacer. To do this, the second dimension gel apparatus (C.B.S. Scientific) was first assembled. All glass plates were cleaned thoroughly and the non-beveled face of each plate was coated with Gelslick or Sigmacote (only once every 10 uses). The back half of the apparatus was laid horizontally on a table top with the upper buffer chamber hanging over the table edge. The two small clear plastic blocks were inserted at the bottom corners of each apparatus. A glass plate was placed in the apparatus, beveled edge facing upward and near the upper buffer chamber, followed by two spacers, one along each side. Glass plates and spacers were added in this manner until the fifth plate had been added. After the third plate, flexible Tygon tubing was slid down the side channel of the apparatus, with a bevel cut in the leading end of the tubing. The other end was cut, leaving approximately 10 cm protruding from the apparatus. The Plexiglas "filler" sheet was placed over the fifth glass plate. The front half of the apparatus was positioned by aligning the screw holes of the front and back half. These were secured with the teflon screws. The oblong oval "windows" at the lower, front face were sealed with Plastic tape (Scotch brand). The apparatus was stood upright in the lower buffer chamber.

Using a three-way stopcock, the gel apparatus tubing was attached in series with a 2 L

IDT Exhibit 1020.8196

reservoir and a 60 ml syringe was attached to the remaining stopcock outlet. The tubing was attached to the 2 L reservoir through a bottom drain (a 2 L graduated cylinder was used). The reservoir was secured above the gel apparatus to allow for gravity flow. The stopcock valve was adjusted to allow liquid to flow between the 2 L reservoir and the 60 ml syringe. Once the TEMED was added, the acrylamide solution (1X TBE, pH 8.3, 96.9 g acrylamide, 3.3 g bis-acrylamide, 1.3 g ammonium persulfate and 700 µl TEMED in a total volume of 2 L) was poured into the 2 L reservoir. The syringe plunger was pulled down to the 50 ml mark. The plunger was depressed to push the air out of the upper tubing. Once all air was removed, the valve was adjusted so that all three ports were open. Acrylamide flowed into the apparatus, filling all four gels simultaneously from the bottom upward. The flow was stopped when the level reached 3 mm from the top edge of the glass plates. The solution was allowed to settle for 2 to 3 minutes. After the valve leading to the gel apparatus had been closed, the syringe and reservoir were detached.

The ends of the in-gel digest digest tubing were separated and the first dimension gel was extruded into a pan containing 1X TBE, pH 8.3. The gel was transferred to a 50 ml tube containing 40 ml 1X TBE, pH 8.3. This was incubated for 10 min at room temperature, replaced with fresh TBE, and incubated for an additional 10 min. The first dimension gel was placed in a horizontal position across the beveled edge of each glass plate. Once all gels were in place, the space between the agarose gel and the top of each polyacrylamide gel was filled with molten 0.8% agarose (equilibrated to 55°C). This connecting agarose was allowed to solidify for 10 to 15 min and then 250 µl second dimension loading dye (1X TE, pH 8.3, 0.25% Bromophenol Blue, 0.25% Xylene Cyanol) was added along the length of each gel. Then 1X TBE, pH 8.3 was added to the upper and lower buffer chambers and electrophoresis was carried out at 100 V for 2 hr and then at 150 V for approximately 24 hr.

Buffers were then removed and the apparatus was disassembled. Each gel was lifted from the plates by overlaying with Whatmann paper cut to size for autoradiographic or phosphorimager cassettes. The perimeter of the paper was traced with the edge of a plastic ruler, removing any excess gel. The Whatmann paper and gel were lifted and placed, gel side up, on a second piece of Whatmann paper. This was overlaid with saran wrap and a third piece of Whatmann paper was added to the top and saran wrap was folded over the top of the Whatmann paper. This was placed in a gel drier, in the same orientation, for 1 hr at 80°C while applying a vacuum. The lower and upper Whatman paper was then removed, saran wrap folded under the remaining paper and exposed to X-ray film (BioMax MS).

IDT Exhibit 1020.8197

E. RLGS spots resulting from methylation-sensitive restriction enzymes identify CpG islands

Using this methodology, an RLGS profile of DNA from human cells produces a pattern displaying approximately 2,000 spots. Fig. 1B, for example, shows such an RLGS profile from normal peripheral blood lymphocyte DNA. First-dimension separation of labeled NotI/EcoRV fragments extends from right to left horizontally. Following in-gel digestion with HinfI, the fragments were separated vertically downward into a polyacrylamide gel and autoradiographed. To allow uniform comparisons of RLGS profiles, spots were defined based on their location in the gel by assigning each spot a three-variable designation (Y coordinate, X coordinate, fragment number). This can be more easily seen in the enlarged portion of section 2D of the RLGS profile (Fig. 1C) showing the numbers assigned to each spot.

From a set of 1,567 NotI spots comprising the central portion of the RLGS profile of normal DNA, 392 spots were eliminated from all analyses on the basis of having more than diploid intensity, less than diploid intensity, or a degree of positional overlap with neighboring fragments. In addition, a small fraction of loci in individual tumor profiles was not able to be analyzed due to poor local gel quality. In normal DNA profiles, the less-than-diploid copy-number intensities can result from polymorphism, partial methylation or spots derived from sex chromosomes. Thus, the analyzed spots were of diploid copy number in most samples. Tumor tissue and healthy tissue DNA profiles were compared by visual inspection of overlaid autoradiographs. In those cases in which matched normal tissue was not available, tumor profiles were compared with profiles matched for tissue type of four to five unrelated individuals. Each CpG island was defined as unmethylated or methylated (a visually apparent decrease in intensity on the RLGS profile, which, through corroboration with Southern-blot data for 26 CpG island loci and more that 100 loss events, corresponded to a 30% or greater level of methylation).

To determine if the NotI restriction sites which produced the RLGS spots, had characteristics of authentic CpG islands, DNA from 210 of the NotI/EcoRV RLGS spots was partially sequenced. This was possible because each spot on the human NotI/EcoRV RLGS profile had previously been assigned to a clone from a NotI/EcoRV genomic plasmid library (see description earlier in the specification). Of the 210 spots, 184 were randomly chosen. Another 26 spots were chosen because they were frequently lost from RLGS profiles from human tumors, suggesting that cytosine nucleotides within the NotI sequence of that spot were methylated in the tumor. From the sequences derived from these clones, the GC content (%GC) was plotted against the CpG ratio for each clone (Fig. 1D; CpG ratio = [(number of

IDT Exhibit 1020.8198

CpGs)/(number of guanines)(number of cytosines)(number of nucleotides analyzed)]). CpG islands have a GC content of greater than 50% and a CpG value of at least 0.6. Fig. 1D shows that, of 210 clones sequenced, 197 (94%) had sequence characteristics consistent with CpG-island DNA.

F. Tumor tissue samples analyzed

DNA used to perform the RLGS analyses was obtained from 98 primary human tumors and, where possible, matched normal samples. These samples were from 8 broad tumor types; breast, colon, gliomas, head and neck, acute myeloid leukemias, primitive neuroectodermal tumors (PNETs) and testicular.

Fourteen breast cancers included 2 adenocarcinomas, 2 lobular carcinomas and 10 ductal carcinomas. The samples were from obtained the Cooperative Human Tissue Network (CHTN). All tumors were from females, 38-89 years of age (average of 54 years). Breast tissue adjacent to the tumor was available for 6 of 14 cases, and 8 tumor profiles were compared with 4 breast samples from the matched sets.

Colon tumors were obtained from Roswell Park Cancer Institute and classified according the American Joint Committee on Cancer staging manual. The 8 primary tumors included 1 stage I tumor, 2 stage II tumors, 2 stage III tumors and 3 stage IV tumors. Patient ages ranged from 49 to 77 years (average of 63 years). Normal adjacent colon mucosa samples were obtained for all tumors.

Fourteen gliomas, including 12 World Health Organization (WHO) grade II astrocytomas and 2 WHO grade III anaplastic astrocytomas, from Saitama Medical School, the University of Tokyo, Teikyo University School of Medicine, Komagome Metropolitan Hospital and the University of Washington, Seattle. Patients included 10 females and 4 males with an age range of 7-57 years (average of 34 years). Brain tissue adjacent to the tumor was also obtained for 1 WHO grade II and 1 WHO grade III tumor. Twelve cases were compared with 3 unmatched normal brain samples and with the 2 brain samples from the matched sets.

Fourteen head and neck squamous cell carcinomas were obtained through the CHTN. Tumors were from 11 males and 3 females. Patients were 42-77 years of age (average of 57 years). Tissue adjacent to the tumor was available for 12 of 14 cases, and 2 tumors were compared with 4 samples from the matched sets.

Nineteen acute myelogenous leukemia samples (3 bone marrow aspirates and 14 peripheral blood) from the Cancer and Leukemia Group B Tissue Bank. Samples were

IDT Exhibit 1020.8199

classified according to the French-American-British system. Samples were obtained from patients at the time of initial diagnosis with AML and again at complete remission (24-154 days, average 45 days) after induction chemotherapy. Samples were from 14 males and 3 females. Patients were 22-61 years of age (average 40 years). All cases were compared with matched samples (either peripheral blood lymphocytes or bone marrow, but always matched with the origin of the cancer sample) obtained at remission.

Twenty-two PNETs, including 17 medulloblastomas and 5 supratentorial PNETs, through the CHTN, Pediatric Division. Tumors were from 15 males and 7 females. Patients were 2-26 years of age, with peak ages between 3 and 6 years. All tumors were WHO grade IV. Matched peripheral blood lymphocytes were available for 6 of 22 cases, and 18 samples were compared with unmatched normal cerebellum DNA.

Nine testicular tumors included 6 seminomas and three nonseminomas. Samples were obtained from the Norwegian Radium Hospital and from the Helsinki University Central Hospital. Patients were 21-77 years (average of 41 years). Adjacent testicular tissue was available for 7 of 9 cases, and 2 samples were compared with 4 samples of testicular DNA used in the matched sets.

G. Loss of spots from RLGS profiles is due to methylation

In comparing RLGS profiles of DNAs from different tumors with control, healthy tissue DNAs, loss of a fragment or spot from an RLGS profile (Fig. 1A) was frequently detected. Loss of such a spot could be due to either methylation of DNA sequences at the NotI site giving rise to that spot, or to deletion of DNA surrounding that NotI site from the genome of the tumor. The relative contribution of each mechanism was assessed by using clones from the NotI/EcoRV genomic library, specific for lost spots, as probes in Southern blotting studies. In Fig. 2A, a section of an RLGS profile, from normal, healthy tissue was compared with tumor tissue from two gliomas, J7 and J16. This RLGS section contains spot 3C1. In tumor J16, spot 3C1 is absent from the RLGS profile. If there was a deletion of DNA surrounding the NotI site, however, the expected result in the Southern blot would be either no hybridization of the probe to the J16 genomic DNA or hybridization to a band of a size different from those detected in the lane containing normal, healthy tissue DNA digested with NotI plus EcoRV, and tumor tissue DNA digested with EcoRV alone. This result is not seen. These results show, therefore, that DNA corresponding to a missing 3C1 spot in J16 glioma DNA is present in the genome, as shown by the Southern hybridization result.

Likewise, DNA corresponding to specific RLGS spots missing in certain leukemias

IDT Exhibit 1020.8200

(Fig. 2B) and neuroectodermal tumors of childhood (Fig. 2C) are found to be present when these DNA are analyzed by Southern blotting. Overall, in 26 tumors where specific spots in RLGS profiles were missing DNA corresponding to the spot, was found to be present in the genome by Southern blotting. These results show that loss of spots on RLGS profiles is due to methylation of the corresponding NotI site and not deletion from the genome of DNA representing that spot. Therefore, methylation is the predominant mechanism underlying loss of spots from RLGS profiles.

H. Heterogeneity in CpG-island methylation across tumors.

To compare the overall pattern of methylated CpG islands among different tumors of the same tumor type, 1,184 spots in each of 98 tumors (and their non-tumorigenic controls) were analyzed by RLGS. The analysis was performed by determining the number of RLGS spots lost, or of decreased intensity, as compared to the controls. Each lost spot or spot of decreased intensity is equivalent to one methylated CpG island. For each tumor type, the number of methylated CpG islands in each individual tumor, as compared to controls, was plotted (Fig. 3). These data showed that breast, head and neck, and testicular tumors had relatively low levels of methylation, with many such tumors showing no methylation. Colon tumors, gliomas, acute myeloid leukemias and primitive neuroectodermal tumors (PNETs) had a much higher frequency of methylation. Nonparametric comparison (Kruskal-Wallis procedure) of the methylation frequencies of the various tumor types showed significant differences between them (χ_6^2=56.9, P<0.0001).

Within a tumor type, the range of methylated CpG islands in individual tumors was variable. The data (Fig. 3) are not consistent with chance variation between tumors because, in the absence of heterogeneity, the variance of the methylation frequency would not be expected to be greater than the mean[1]. A formal test of this overdispersion was performed for each tumor type and the results are shown in Fig. 3 as a superimposition of the expected Poisson distribution on the dot plots. These data showed that aberrant methylation of CpG islands can be quantitatively different in individual tumors within a tumor type and more pronounced overall in particular tumor types.

[1]Heterogeneity of methylation frequencies across samples was assessed within each tumor type by a standard test for evidence that the variance in methylation frequency exceeds the mean. This test is motivated by the Poisson approximation, which applies even if the frequencies of methylation vary across CpG islands. Moreover, a simple result from the binomial distribution shows that the test is conservative, because under homogeneity the population variance cannot exceed the mean.

IDT Exhibit 1020.8201

I. Subsets of CpG islands were preferentially methylated in tumors

Through analysis of the RLGS spots lost in different tumors, it was determined that certain spots on the RLGS gels were lost in multiple tumors. This means that specific CpG dinucleotides were methylated in more than one tumor. This is shown in Fig. 4 where the number of tumors within a specific tumor type that had a particular CpG island methylated are shown.

To test the hypothesis that methylation of these common CpG islands was not random, a standard goodness-of-fit test was used.[2] This can be seen in the plots of Fig. 4 where the black line of each plot shows the expected distributions if methylation of specific CpG islands in multiple tumors was random. It can be seen from Fig. 4 that for breast tumors, colon tumors, gliomas, acute myeloid leukemias and childhood PNETs, the actual distributions were significantly different ($P<0.0001$) from the theoretical distributions indicative of randomness. Similarly, the results for head and neck tumors were significant ($P<0.025$). The results for testicular tumors ($P=0.365$) were not significant. However, tumors of this type have a low overall methylation frequency and larger sample sizes are needed. Overall, the data indicate that the patterns of CpG island methylation in tumors is not random.

J. Frequencies of aberrant CpG-island methylation of shared and tumor-type-specific targets

Because the data have shown that they are methylated in a nonrandom fashion, CpG islands that are methylated at a high frequency in one or more tumor types can be used for diagnosis of tumors. From analysis of 98 tumors using NotI/EcoRV RLGS analysis, a number of spots that are absent or of decreased intensity, as compared to control healthy tissue DNA, have been found. Table I lists these spots. Each fragment (CpG island) is identified in three ways in the table. First, the location of each CpG island is designated as the distance (in cm) migrated during electrophoresis, from the gel origin, in both the first dimension and the second dimension. Second, each CpG island is given a three-variable designation (Y coordinate, X coordinate, fragment number). The X coordinate indicates horizontal direction on the two-dimensional RLGS profile and is a letter from B-G. The Y coordinate indicates vertical direction and is a number from 1-5. Together, an X and Y

[2]Under the null hypothesis of equal methylation frequencies for each CpG island, a goodness-of-fit test (χ^2) was applied to the observed versus expected frequencies of islands exhibiting methylation in multiple tumors within each tumor type.

IDT Exhibit 1020.8202

designation divide the RLGS profile into 28 sections. Within each section, the spots/fragments are given a number. Such a profile is available at: http://pandora.med.ohio-state.edu/masterRLGS/. Third, the partial DNA sequence of individual spots has been determined by sequencing of library clones corresponding to each spot. These sequences are shown in the attached Sequence Listing and have been assigned SEQ ID NOS. from 1 to 82.

The diagnostic NotI/EcoRV spots are of two types (Fig. 1). The first type of spot is absent or of decreased intensity in a single tumor type. For example, the NotI site that is part of the CpG island designated 2.B.53, is methylated only in head and neck tumors. Similarly, the NotI site of CpG island 2.F.2 is methylated only in breast tumors.

The second type of spot is absent or of decreased intensity in more than one type of tumor. For example, the NotI/EcoRV spot designated 2.C.24 is missing in gliomas and AMLs. Similarly, the NotI/EcoRV spot designated 3.B.55 is methylated in breast, colon and PNETs.

Table I. Diagnostic CpG islands in tumors.

CpG Island[1]	1st-D (cm)[2]	2nd-D (cm)[2]	Type[3]	Methylated In[4]:
2.B.53	36.85	9.25	t	HN
2.C.24	30.3	5.32	s	Abt/Leu
2.C.29	27.8	5.45	s	Leu/Hn
2.C.35	29.45	6.9	s	Abt/Bre/Cln/Leu/Pbt
2.C.54	32.38	9.42	s	Leu/Hn
2.C.57	30.9	8.5	ND	Tst
2.C.58	31.2	9.2	s	Abt/Leu
2.C.59	30.4	9.35	ND	Hn
2.D.10	27.55	5.3	s	Leu/Pbt
2.D.14	24.25	4.47	t	Leu
2.D.20	26.3	5.3	t	Cln
2.D.25	27.15	6.4	ND	Bre
2.D.27	25.65	5.82	ND	Hn
2.D.34	23.62	6.6	s	Leu/Pbt
2.D.40	23.95	7.25	ND	Pbt
2.D.48	26.1	8.1	ND	Leu
2.D.55	24.2	8.3	s	Cln/Leu
2.D.74	23.95	9.35	s	Abt/Bre/Cln/Leu
2.E.20	20.6	5.95	ND	Pbt
2.E.24	19.35	5.7	s	Abt/Leu
2.E.25	18.27	5.65	t	Bre
2.E.30	20.35	6.4	s	Abt/Bre/Leu
2.E.37	21.42	7.1	ND	Bre
2.E.4	21.1	4.48	s	Leu/Pbt

IDT Exhibit 1020.8203

2.E.40	NA	NA	ND	Tst
2.E.61	19.4	8.08	s	Abt/Pbt
2.E.64	20.5	8.35	s	Abt/Cln
2.F.2	17.27	4.72	t	Bre
2.F.41	NA	NA	t	Tst
2.F.50	15.23	7	s	Abt/Leu
2.F.59	17.49	8	ND	Bre
2.F.70	15.88	13.3	s	Pbt/Tst
2.G.10	10.29	4.49	s	Leu/Tst
2.G.108	7.68	7.44	ND	Bre
3.B.30	35.4	12.55	ND	Tst
3.B.36	34.2	11.8	s	Abt/Cln/Leu/Pbt
3.B.55	NA	NA	s	Bre/Cln/Pbt
3.C.01	31.6	9.7	s	Abt/Cln/Leu
3.C.16	27.9	11.8	t	Pbt
3.C.17	29.2	10.57	t	Cln
3.C.30	31.61	10.37	t	Bre
3.C.35	31.6	11.5	t	Pbt
3.C.64	29.1	14.05	ND	Bre
3.D.21	24.2	10.75	t	Leu
3.D.24	23.2	11.03	s	Abt/Leu
3.D.35	26.1	11.65	s	Abt/Cln/Leu/Pbt
3.D.40	23.4	12.26	s	Abt/Cln/Leu
3.D.44	24.45	12.82	t	Leu
3.D.60	27.2	12.4	s	Abt/Cln/Leu
3.E.04	20.4	14.2	s	Hn/Pbt
3.E.50	20.55	10.7	s	Hn/Tst
3.E.55	18.78	10.55	s	Cln/Leu
3.E.57	18.09	10.9	s	Cln/Hn
3.E.59	18.4	9.72	s	Abt/Tst
3.F.16	16.6	9.75	ND	Leu
3.F.2	16.73	9.35	s	Leu/Tst
3.F.50	16.25	11.6	s	Cln/Leu/Tst
3.F.72	16.9	13.7	t	Leu
3.F.82	13.8	13.12	s	Abt/Cln/Leu
3.G.46	9.88	11.5	ND	Bre
3.G.78	10	12.93	ND	Leu/Pbt
4.B.44	33.7	18.53	s	Cln/Hn
4.B.56	33.2	19.45	s	Bre/Leu
4.C.05	30	14.9	ND	Bre
4.C.25	28.62	17	ND	Bre
4.C.42	NA	NA	ND	Tst
4.C.9	30.3	15.3	ND	Bre
4.D.07	22.9	14.5	s	Leu/Tst
4.D.08	23.5	15	s	Abt/Tst
4.D.12	25	14.85	s	Abt/Leu/Tst
4.D.13	24.95	15.3	s	Abt/Bre

IDT Exhibit 1020.8204

4.D.47	27.6	18.25	s	Abt/Leu/Pbt
4.E.53	19.39	18.43	t	Leu
4.F.15	13.25	15.45	t	Cln
4.F.17	14.1	15.6	s	Abt/Bre/Cln
4.F.22	17.56	16.2	s	Cln/Hn/Leu
4.F.6	14.85	14.59	ND	Bre
4.F.69	12.58	18.86	t	Abt
5.D.9	25.17	23.4	t	Hn
5.E.2	20.58	19.5	t	Bre
5.E.25	18.7	21.3	t	Cln
5.E.4	18.45	19.75	s	Abt/Bre/Leu
[1]Y coordinate, X coordinate, fragment number				
[2]NA, spots too close to analyze.				
[3]T, tumor-type specific target of methyaltion; s, shared target of methylation; ND, not determined.				
[4]Types of tumor in which CpG island is methylated: Abt, gliomas; Bre, breast; Cln, colon; Hn, head and neck; Leu, acute myeloid leukemia; Pbt, pediatric brain tumors; Tst, testicular germ cell tumors.				

Example 2. Identification of diagnostic markers for lung cancer using AscI and RLGS

Tissue from lung tumors was obtained as surgical tissue samples. Where possible, surrounding non-tumor tissue from the same patient was obtained and used as a control. DNA was isolated from the tissue as described in Example 1. In preparation for RLGS analysis, the ends of the DNA were blocked as described in Example 1. The DNA was then digested with AscI followed by digestion with EcoRV. The AscI restriction enzyme recognizes the sequence 5'GGCGCGCC3' and does not cleave said sequence if cytosines within the sequence are methylated. First dimension gel electrophoresis, in-gel digestion with HinfI, second dimension gel electrophoresis and autoradiography were performed as described in Example 1.

RLGS profiles from lung tumor DNA were compared with RLGS profiles obtained from healthy, non-tumor tissue DNA. Spots which were lost or present at reduced intensity in tumor tissue RLGS profiles as compared to profiles obtained from healthy tissue were noted. Eight spots were lost or altered in the RLGS profiles from multiple lung tumor samples. A compilation of such spots is shown in Table II (lung tumors).

DNA sequence information was obtained from the lung cancer-specific spots. This was done by sequencing individual clones of an AscI/EcoRV library that was made from DNA from healthy tissue. Individual clones of this library that corresponded to spots on the AscI/EcoRV RLGS profile were identified by overloading an RLGS gel with DNA from

IDT Exhibit 1020.8205

various groups of library clones, as was described earlier in the specification of this application for the NotI/EcoRV library. After individual clones were matched with spots in the AscI/EcoRV profile, the DNA from the spots that were missing in profiles from the lung tumor DNAs were sequenced. Such sequence information is shown in the attached DNA Sequence Listing.

Table II. Diagnostic CpG islands grouped by tumor type.

Library	Tumor type	Tumor type specific (+), shared (-), or not determined (ND)[1]	CpG island designation`
NotI/EcoRV	Breast	+	2.E.25, 2.F.2, 3.C.30, 5.E.2
		-	3.B.55, 4.B.56, 4.D.13, 4.F.17, 2.D.74, 2.C.35, 2.E.30, 5.E.4
		ND	2.D.25, 2.E.37, 2.F.59, 2.G.108, 3.C.64, 3.G.46, 4.C.05, 4.C.25, 4.C.9, 4.F.6
NotI/EcoRV	Colon	+	2.D.20, 3.C.17, 4.F.15, 5.E.25
		-	3.E.57, 4.B.44, 4.F.22, 2.D.55, 3.E.55, 3.F.50, 3.B.55, 4.F.17, 2.D.74, 2.C.35, 2.E.64, 3.C.01, 3.D.40, 3.D.60, 3.F.82, 3.B.36, 3.D.35
		ND	---
NotI/EcoRV	Glioma	+	4.F.69
		-	4.D.13, 4.F.17, 2.D.74, 2.C.35, 2.E.30, 5.E.4, 2.E.64, 3.C.01, 3.D.40, 3.D.60, 3.F.82, 3.B.36, 3.D.35, 2.C.24, 2.C.58, 2.E.24, 2.F.50, 3.D.24, 4.D.47, 4.D.12, 2.E.61, 3.E.59, 4.D.08
		ND	---
NotI/EcoRV	Head & neck	+	2.B.53, 5.D.9
		-	2.C.29, 2.C.54, 3.E.04, 3.E.50, 3.E.57, 4.B.44, 4.F.22

IDT Exhibit 1020.8206

		ND	2.C.59, 2.D.27
NotI/EcoRV	Acute myelogenous Leukemia	+	2.D.14, 3.D.21, 3.D.44, 3.F.72, 4.E.53, 2.C.29, 2.C.54
		-	2.D.10, 2.D.34, 2.E.4, 2.G.10, 3.F.2, 4.D.07, 4.F.22, 2.D.55, 3.E.55, 3.F.50, 2.E.64, 3.C.01, 3.D.40, 3.D.60, 3.F.82, 3.B.36, 3.D.35, 3.C.01, 3.D.40, 3.D.60, 3.F.82, 3.B.36, 3.D.35, 2.C.24, 2.C.58, 2.E.24, 2.F.50, 3.D.24, 4.D.47, 4.D.12
		ND	2.D.48, 3.F.16, 3.G.78, 4.B.56
NotI/EcoRV	Pediatric neuroectoderm al tumor of childhood	+	3.C.16, 3.C.35, 3.E.04
		-	2.D.10, 2.D.34, 2.E.4, 3.B.55, 2.C.35, 3.B.36, 3.D.35, 4.D.47, 2.E.61
		ND	2.D.40, 2.E.20, 3.G.78
NotI/EcoRV	Testicular	+	2.F.41
		-	2.G.10, 3.F.2, 4.D.07, 3.E.50, 3.F.50, 4.D.12, 3.E.59, 4.D.08
		ND	2.C.57, 2.E.40, 3.B.30, 4.C.42
AscI/EcoRV	Lung	+	
		-	
		ND	A.2.F.45, A.2.F.50, A.2.F.67, A.3.F.38, A.4.D.30, A.4.D.36, A.4.E.32, A.5.E.28[2]

[1]ND, not determined. Indicates that the designated CpG island was methylated in the indicated tumor type but its methylation in other tumor types was not determined.
[2]The "A" preceding the X, Y, number designation for the CpG islands indicates that these islands are from the AscI/EcoRV RLGS profile.

Example 3. Design of primers for cancer diagnosis

Primers are designed for diagnosis of cancer using methylation-specific PCR (MSR). The primers are designed to amplify regions of the human genome whose sequences are contained within the library clones disclosed in this application. Two sets of primers are needed for each library clone whose DNA sequence is to be used for diagnosis of cancer.

IDT Exhibit 1020.8207

Each primer set is designed to amplify the same region of the genome, said region beginning at the end of a library clone containing the methylation-sensitive restriction enzyme recognition site (i.e., the NotI site for the library described in Example 1; the AscI site for the library described in Example 2) and ending at a region contained within the clone up to 200 nucleotides from the methylation-sensitive restriction enzyme recognition site.

The first set of primers is designed to amplify template genome DNA whose cytosine residues are not methylated and, after bisulfite treatment, the cytosines of said genome DNA are converted to uracil. The second set of primers is designed to amplify template genome DNA which is methylated on cytosines that comprise CpG dinucleotides. Such methylated cytosines are unaffected by bisulfite treatment. Therefore, by using two sets of primers, one set that will amplify only unmethylated DNA and another set that will amplify only methylated DNA, methylation state of the template DNA can be determined. Such methylation state can be diagnostic for cancer.

The primers used for MSR are designed to be from 15 to 34 nucleotides in length and contain within their sequence either CpG dinucleotides or dinucleotides complementary to CpG dinucleotides that have been treated with bisulfite. It is preferred that the 3' ends of primers used to amplify unmethylated DNA are CpA dinucleotides. It is preferred that the 3' ends of primers used to amplify methylated DNA are CpG dinucleotides.

For each library clone to be used diagnostically, the first set of primers are designed to amplify genome DNA that is not methylated. After treatment of such genome DNA with bisulfite, all such unmethylated cytosines are converted to uracil. PCR primers that will use such DNA as a template and amplify it, will have adenine residues which are complimentary to these uracils.

For the first set of primers, the 5' end of one of the primers begins at the end of the library clone containing the methylation-sensitive restriction enzyme recognition site. The sequence of this primer is identical in sequence to the strand of the template which has its 5' end as part of the methylation-sensitive restriction enzyme site, except that guanine residues are replaced with adenine residues. The adenines allow the primer to hybridize with the template strand in which cytosines have been converted to uracils by bisulfite. This primer extends to a length of between 15 and 32 total nucleotides. Preferably, the 3' end of said primer ends with a CpA dinucleotide, the adenine of said dinucleotide hybridizing to a uracil which, before bisulfite treatment, had been a cytosine that comprised a CpG dinucleotide.

The diagram below shows implementation of these rules to select a primer that can be used to amplify clone 2.B.53 of the NotI/EcoRV library (Table I and attached sequence

IDT Exhibit 1020.8208

listing). In the diagram, I shows the end of the 2.B.53 clone containing the methylation-sensitive NotI site (NotI recognition sequence is shown in bold letters). CpG dinucleotides are shaded. To amplify a region of this clone rightward of the NotI site, the first primer is identical to the top strand of the duplex shown in I. However, since bisulfite treatment of the DNA in I converts cytosines to uracils, guanines within the PCR primer must be replaced with adenines. II shows the sequence of the bottom strand of I after bisulfite treatment converts cytosines to uracils. A primer complementary to the bisulfite-treated bottom strand has the sequence shown in III.

I

```
5'GCGGCCGCGGTTAGCTTCTCCTGTCCGAACCAGGG-----
3'CGCCGGCGCCAATCGAAGAGGACAGCGTTCCGTCCC-----
```

II

```
3'UGUUGGUGUUAATUGAAGAGGAUAGGUTTGUGTUUU-----
```

III

```
5'ACAACCACAATTAACTTCTCCTATCCAAACA 3'
```

III shows the entire sequence of one of the two primers used to amplify unmethylated genome DNA corresponding to library clone 2.B.53. This primer encompasses 5 CpG dinuceotides, as shown by the shading in I above. Encompassment of 2 or more such CpG dinucleotides is preferred so that this primer will not hybridize to a bisulfite-treated template which contains methylated cytosines. The 3'end of the primer shown in III ends in a CpA dinucleotide. This is also preferred in order to provide maximal discrimination of the primer between methylated and unmethylated template DNA in MSR. The primer shown in III has a length of 31 nucleotides.

The second primer is designed to work with the first primer in PCR amplification such that a fragment of less than about 200 base pairs is amplified. Therefore, this primer is made to a sequence rightward of the sequence shown in I. The sequence of this primer is complementary in sequence to the strand of the template which has its 5' end as part of the methylation-sensitive restriction enzyme site, except that guanine residues are replaced with adenine residues. This primer is preferably between 15 and 32 nucleotides in length. This primer is also designed to preferably encompass 2 or more CpG dinucleotides. Preferably, the 3' end of said primer ends with a CpA dinucleotide.

The diagram below shows implementation of these rules to select a primer that can be used to amplify unmethylated genome DNA corresponding to clone 2.B.53 of the

IDT Exhibit 1020.8209

NotI/EcoRV library. IV shows a region of the 2.B.53 clone about 70 nucleotides rightward of the sequence in I of the earlier diagram. The CpG dinucleotides within the sequence are shaded. To amplify a region leftward of this region, this second primer must be complementary to the top strand of the duplex shown in IV. However, bisulfite treatment of the DNA in IV converts cytosines to uracils. A primer complementary to this bisulfite-treated top strand has the sequence shown in VI.

IV

5′-----GGAGTCGCGGGTCGCGGGAGGCTGCGCCGCGCCACCGGA-----3′
3′-----CCTCAGCGCCAGCGCCCTCCGACGCGGCGCGGTGGCT-----5′

V

5′-----GGAGTUGUGGTUGUGGGAGGUTGUGUUGUGUAUUGA-----3′

VI

3′ ACACCAACACCCTCCAACACAACACATAACT 5′

VI shows the entire sequence of the second primer used to amplify unmethylated genome DNA corresponding to library clone 2.B.53. This primer encompasses 8 CpG dinucleotides, as shown by the shading in IV. Encompassment of 2 or more such CpG dinucleotides is preferred. The 3'end of the primer shown in VI ends in a CpA dinucleotide. This is also preferred. The primer shown in VI has a length of 31 nucleotides. Together, the first and second primers amplify a PCR fragment of 128 base pairs in length.

The above primers amplify genome DNA that does not contain 5-methylcytosine. The above primers will not amplify genome DNA containing 5-methylcytosines because 5-methylcytosines are not converted to uracils by bisulfite treatement. The two primers already described (III and VI), therefore, will not be complementary to bisulfite-treated genome DNA which is methylated.

Therefore, a second set of primers is designed to amplify genome DNA that is methylated. Methylation in human cells occurs at cytosines that are part of CpG residues. Such methylated cytosines are not converted to uracil by bisulfite treatment. Cytosines that are not part of CpG residues are not methylated and, therefore, are converted to uracil by bisulfite. The primers of the second set are designed to amplify the same region of a library clone as did the first set of primers. But, because the genome DNA contains both cytosines that are methylated and cytosines that are not methylated, the sequences of primers used to amplify such DNA are different than the sequences of the first primer set. Like the first set of

IDT Exhibit 1020.8210

primers, however, the primers of the second set are preferably between 15 and 32 nucleotides in length. Preferably the 3' ends of such primers contain CpG dinucleotides.

The diagram below shows implementation of these rules to select the first of two primers that can be used to amplify methylated genomic DNA corresponding to clone 2.B.53 of the NotI/EcoRV library. In the diagram below, VII shows the end of the 2.B.53 clone containing the NotI site (NotI recognition sequence is bolded). CpG dinucleotides are shaded. Cytosines within said CpG dinucleotides are methylated and are underlined in VII to indicate methylation to 5-methylcytosine. Treatment of the DNA in VII with bisulfite produces a bottom strand with the sequence shown in VIII. In VIII, only unmethylated cytosines are converted to uracil by bisulfite.

VII

5'GCGGCCGCGGTTAGCTTCTCCTGTCCGAACGCAGGG-----
3'CGCCGGCGCCAATCGAAGAGGACAGGCTTCCGTCCC-----

VIII

3'UGCUGGCGCUAATUGAAGAGGAUAGGCTTGCGTUUU-----

IX

5'ACGACCGCGATTAACTTCTCCTATCCGAACG 3'

A primer complementary to the bisulfite-treated bottom strand shown in VIII is shown in IX. Said primer will prime PCR amplification of sequences rightward of those shown in VII. The primer shown in IX encompasses 5 CpG dinucleotides. Encompassment of 2 or more such CpG dinucleotides is preferred. The 3' end of the primer shown in IX ends in CpG. This is also preferred. The primer shown in IX has a length of 31 nucleotides.

A second primer is designed to work with the primer shown in IX to amplify methylated genome template DNA. Design of such a primer is shown below. In the diagram, X shows the same region of clone 2.B.53 (approximately 70 nucleotides rightward of the sequences shown in VII) that is shown in IV. Treatment of the DNA in X with bisulfite produces a top strand with the sequence shown in XI. In XI, only unmethylated cytosines are converted to uracil by bisulfite.

X

5'-----GGAGTCGCGGTCGCGGGAGGCTGCGCCGCGCACCGA-----3'
3'-----CCTCAGCGCCAGCGCCCTCCGACGCGGCGCGTGGCT-----5'

XI

5'-----GGAGTCGCGGTCGCGGGAGGUTGCGUCGCGUAUCGA-----3'

IDT Exhibit 1020.8211

XII

3′ GCGCCAGCGCCCTCCAACGCAGCGCATAGCT 5′

A primer complementary to the bisulfite-treated top strand (XI) has the sequence shown in XII. Said primer will prime PCR amplification of sequences leftward of those shown in X. The primer shown in XII encompasses 8 CpG dinucleotides. Encompassment of 2 or more such CpG dinucleotides is preferred. The 3' end of the primer shown in XII ends in a CpG dinucleotide. This is also preferred. The primer shown in XII has a length of 31 nucleotides. Together, the first (IX) and second primers (XII) of the second set amplify a PCR fragment of 128 base pairs in length.

Example 4. Use of oligonucleotides to diagnose cancer

The library clones, and DNA sequences within, can be used to detect DNA methylation in a genome at the specific sequences identified by the sequences within the clone. Such detection can be diagnostic for cancer. Various methods can be used for such diagnosis.

A. Diagnosis of cancer using methylation-sensitive restriction enzymes followed by Southern blot

Cleavage or lack of cleavage by a methylation-sensitive restriction enzyme at a specific restriction enzyme recognition site can be detected by a probe for the specific recognition site, using Southern blotting. Genomic DNAs were isolated (as described in Example 1) from tumor tissue from a patient with acute myelogenous leukemia (AML). Cells from the same patient after chemotherapy and remission of the disease served as a source of control, healthy tissue DNA. The AML and control DNAs were designated as 26T and 26N, respectively. The DNAs were digested with NotI and EcoRV for 4 hours and then electrophoresed through a 0.8% agarose gel. DNA within the gel was depurinated by soaking the gel in 0.2 N HCl for 10 min. The gel was equlibrated in transfer solution (0.5 N NaOH, 1 M NaCl) for 10 min. and then blotted to Zeta Bind-GT nylon membranes (Bio-Rad). Blots were crosslinked with UV light, baked in a vacuum oven and then prehybridized for 1 hour at 65°C in a solution of 7% SDS, 500 mM sodium phosphate buffer (pH 7.2) and 1 mM EDTA. The blot was hybridized overnight at 65°C in prehybridization solution with 10 ng of α-^{32}P-labeled probe at a specific activity of 10^8-10^9 dpm/μg. The DNA probe used was the 2.C.40 clone from the NotI/EcoRV 2.C.40library. The purified NotI/EcoRV fragment (50 ng) was

IDT Exhibit 1020.8212

labeled with [α-^{32}P]dCTP by random priming using the Prime-It II random-prime labeling kit (Stratagene). The blot was washed with two quick rinses at 65°C in wash solution 1 (100 mM sodium phosphate buffer, pH 7.2, 0.1% SDS), followed by one 30 min. wash at 65°C in wash solution 1. The blot was next washed for 30 min. at 65°C in wash solution 2 (40 mM sodium phosphate buffer, pH 7.2, 0.1% SDS). Bands were visualized by autoradiography using Kodak X-OMAT AR film.

Fig. 2B shows the data. The first 2 lanes of the autoradiograph are relevant. The first lane, labeled 26N is the normal, healthy tissue DNA cleaved with both NotI and EcoRV. The 26N lane shows a band near the bottom of the autoradiograph labeled "NotI/EcoRV." This is fragment resulting when the NotI site present in the 2.C.40 clone is unmethylated. The adjacent lane, labeled "26T," is the tumor tissue DNA cleaved with both NotI and EcoRV. It is seen that this band, labeled "EcoRV," does not migrate as fast as did the 26N band. The reason is that the NotI site present in the 2.C.40 clone is methylated and the NotI enzyme was unable to cleave at this site.

B. Diagnosis of cancer using methylation-specific PCR (MSR)

MSR is a technique whereby DNA is amplified by PCR dependent upon the methylation state of the DNA. In this example, the specific areas of the genome whose methylation status is to be determined are the regions at the ends of the CpG islands that are demarcated by the methylation-sensitive restriction enzyme recognition sequence. In the case of the NotI/EcoRV RLGS profiles, this is the NotI site. In the case of the AscI/EcoRV RLGS profiles, this is the AscI site, at the end of each clone.

For the purposes of this example, the methylation status of genomic sequences corresponding to the NotI site of clone 2.B.53 of the NotI/EcoRV library is examined. Genomic DNA is first isolated from normal tissue and from tumor tissue, as described in Example 1. This DNA is then treated with bisulfite. This is done by taking 1 μg of genomic DNA in a volume of 50 μl and denaturing said DNA in a final concentration of 0.2 M NaOH. Thirty microliters of 10 mM hydroquinone and 520 μl of 3 M sodium bisulfite, at pH 5.0, are added, mixed and incubated under mineral oil at 50°C for 16 hours. The modified DNA is then purified using the Wizard DNA purification resin (Promega) and eluted into 50 μl of water. Modification is completed by NaOH (final concentration, 0.3 M) treatment for 5 min. at room temperature, followed by ethanol precipitation. DNA is resuspended in water.

Each genomic DNA is then used in two PCR reactions. One PCR reaction will amplify DNA that is not methylated and has, therefore, been modified by bisulfite. The

IDT Exhibit 1020.8213

second PCR reaction will amplify DNA that is methylated. Separate primers are used for each reaction. To determine the methylation status of the NotI site in the genomic DNA which corresponds to the 2.B.53 clone, the two sets of primers described in Example 3 are used. Each PCR reaction contains 1X PCR buffer (16.6 mM ammonium sulfate, 67 mM Tris, pH 8.8, 6.7 mM $MgCl_2$, 10 mM 2-mercaptoethanol), dNTPs (each at 1.25 mM), primers (300 ng each per reaction), and 50 ng bisulfite-modified DNA in a final volume of 50 µl. Separate control reactions are run which contain DNA that has not been modified by bisulfite. Reactions are hot-started at 95°C for 5 min. before the addition of 1.25 units of *Taq* polymerase. Amplification is carried out for 35 cycles (30 sec at 95°C, 30 sec at the annealing temperature, and 30 sec at 72°C), followed by a final 4 min. extension at 72°C. Each PCR reaction is directly loaded onto nondenaturing 6-8% polyacrylamide gels and electrophoresed. Gels are stained with ethidium bromide and visualized under UV illumination.

If input genomic DNA is not methylated at cytosines within CpG dinuceotides at the NotI site corresponding to the end of the 2.B.53 CpG island clone, the PCR reaction using the primers specific for nonmethylated DNA (primers III and VI in Example 3) will produce an amplification product of 128 base pairs in length. Using the same input genomic DNA, the PCR reaction using the primers specific for methylated DNA (primers IX and XII in Example 3) will not produce an amplification product.

If input genomic DNA is methylated at cytosines within CpG dinucleotides at the NotI site corresponding to the end of the 2.B.53 CpG island clone, the PCR reaction using the nonmethylation-specific primers will not produce an amplification product. Using the same input genomic DNA, the PCR reaction using the methylation-specific primers will produce an amplification product of 128 base pairs in length.

Example 5. Detection of gene expression

The library clones (Tables I and II) and DNA sequences (attached sequence listing) are useful for determining whether genes encoded within said clones are being transcribed in tumor tissue or cultured cells. To determine transcription, RNA was isolated from five different human glioma cell lines (U87MG, U178, T98G, U251 and LN235) using Trizol (Gibco BRL). Such RNA isolation reagent is known to those skilled in the art. RNAs were quantified using a spectrophotometer and then treated with amplification grade Dnase I (Gibco). The RNA (2 µg) was reverse transcribed by incubation with oligo-dT and random

IDT Exhibit 1020.8214

primers in a 20 µl reaction, heated to 70°C for 10 min. and placed on ice. A mix containing 1X reaction buffer (Gibco), DTT (10 mM), dNTPs (0.5 mM each), and RNAsin (80 U, Promega) was added to each sample. The samples were divided into two tubes, each containing 19 µl, and incubated at 37°C for 2 min. M-MLV reverse transcriptase (RT, 200 U) was added to one of the two tubes and each was incubated at 37°C for 1 h. DEPC-treated water (30 µl) was added to each sample and heated in boiling water for 5 min.

PCR amplification of the reverse transcribed RNA was then performed. In this study, transcripts encoded by sequences within the 2.C.24 library clone (Table I) were looked for. A computer search using the BLAST program had identified an open reading frame within the sequence of this library clone. PCR primers were made to this region. Primer 1 was 5' TGGTGCTGAAGTCGGTGAA 3'. Primer 2 was 5' GGGCCATCTTCACCATCTG 3'.

These primers (10 pmol of each) were used in 10 µl PCR reactions which contained 1.5 µl of the reverse transcription reaction, 1X reaction buffer, Taq polymerase (0.5 U, Boehringer), and dNTPs (250 µM each). For each gene, separate amplification reactions were carried out using RT-positive and RT-negative reactions as template. Amplification was not detected from the RT-negative reactions. The PCR reactions were carried out by heating the samples to 94°C for 5 min and then amplifying for 35 cycles, each cycle consisting of 94°C for 30 sec., a 30 sec. annealing step at 56°C, and 72°C for 45 sec. The reactions were then incubated at 72°C for 7 min and cooled to 4. The sample was then electrophoresed through an agarose gel containing ethidium bromide and PCR products were visualized using an Eagle Eye gel documentation system (Stratagene). The correct identity of the PCR products was confirmed by nucleotide sequencing of both strands.

The data showed that no transcripts encoded by this region of the 2.C.24 clone were found in any of the 5 glioma cell lines. Such expressed transcripts are present in RNA obtained from human fetal brain and adult brain.

In addition to examination of cell lines, tumor tissue obtained from patient samples can be similarly tested for the presence of transcripts by one skilled in the art. Other techniques to detect transcripts can also be used. Such techniques include, for example, Northern blot hybridization, RNase protection and primer extension assays.

IDT Exhibit 1020.8215

CLAIMS

What is claimed is:

1. A method of identifying CpG islands which are preferentially methylated in malignant cells contained within a tumor or neoplasm, comprising:

a) digesting genomic DNA obtained from the malignant cells with an infrequently- cutting, methylation-sensitive, restriction enzyme to provide a set of malignant cell restriction fragments;

b) digesting genomic DNA obtained from non-malignant, control cells with an infrequently-cutting, methylation-sensitive, restriction enzyme to provide a set of control cell restriction fragments;

c) attaching a detectable label to the ends of the malignant cell restriction fragments and the control restriction fragments;

d) digesting the labeled malignant cell and control cell restriction fragments with a second restriction enzyme;

e) separating the labeled malignant cell restriction fragments and the labeled control cell restriction fragments, wherein the malignant cell restriction fragments and the control cell restriction fragments are separated by electrophoresis on two different gels;

f) digesting the restriction fragments in each of said gels with a third restriction enzyme;

g) electrophoresing the restriction fragments in each of said gels in a direction perpendicular to the first direction to provide a first pattern of detectable malignant cell restriction fragments and a second pattern of detectable control cell restriction fragments; and

h) comparing the first pattern to the second pattern to identify diagnostic control cell restriction fragments in said second pattern which are absent or exhibit a decreased intensity in the first pattern, wherein said diagnostic control cell restriction fragments comprise a CpG island that is unmethylated in the DNA of the control cells and methylated in the DNA of the malignant cells.

2. The method of claim 1 further comprising the step of determining the sequence of at least a portion of a diagnostic control cell restriction fragment, wherein said portion is located at or near an end of the fragment .

3. The method of claim 1 further comprising the step of obtaining a clone from a DNA library which comprises a diagnostic control cell restriction fragment.

4. A method of preparing a polynucleotide or oligonucleotide for characterizing tissue

IDT Exhibit 1020.8216

obtained from a subject suspected of having cancer, comprising:

synthesizing a polynucleotide or oligonucleotide which comprises a sequence which is identical to or substantially complementary to a target sequence on one of the strands of a diagnostic control fragment identified according to the method of claim 2, wherein said target sequence comprises at least two CpG dinucleotides, wherein said oligonucleotide is from 15 to 34 nucleotides in length, and wherein said polynucleotide is from 35 to 2000 nucleotides in length.

5. The method of claim 4 wherein said target sequence is located at or near the control restriction fragment end which was cleaved by the methylation-sensitive, restriction enzyme.

6. The method of claim 4 wherein the target sequence is located from about 100 nucleotides to about 500 nucleotides downstream of the control restriction fragment end that was cleaved by the methylation-sensitive, restriction enzyme.

7. The method of claim 4 wherein the control restriction fragment comprises a sequence selected from the group consisting of SEQ. ID. NO.:1, SEQ. ID. NO.:2, SEQ. ID. NO.:3, SEQ. ID. NO.:4, SEQ. ID. NO.:5, SEQ. ID. NO.:6, SEQ. ID. NO.:7, SEQ. ID. NO.:8, SEQ. ID. NO.:9, SEQ. ID. NO.:10, SEQ. ID. NO:11, SEQ. ID. NO.: 12, SEQ. ID. NO.:13, SEQ. ID. NO.:14, SEQ. ID. NO.:15, SEQ. ID. NO.: 16, SEQ. ID. NO.:17, SEQ. ID. NO.:18, SEQ. ID. NO.: 19, SEQ. ID. NO.:20, SEQ. ID. NO.:21, SEQ. ID. NO.:22, SEQ. ID. NO.:23, SEQ. ID. NO.:24, SEQ. ID. NO.:25, SEQ. ID. NO.:26, SEQ. ID. NO.:27, SEQ. ID. NO.:28, SEQ. ID. NO.:29, SEQ. ID. NO.:30, SEQ. ID. NO.:31, SEQ. ID. NO.:32, SEQ. ID. NO.:33, SEQ. ID. NO.:34, SEQ. ID. NO.:35, SEQ. ID. NO.:36, SEQ. ID. NO.:37, SEQ. ID. NO.:38, SEQ. ID. NO.:39, SEQ. ID. NO.:40, SEQ. ID. NO.:41, SEQ. ID. NO.:42, SEQ. ID. NO.:43, SEQ. ID. NO.:44, SEQ. ID. NO.:45, SEQ. ID. NO.:46, SEQ. ID. NO.:47, SEQ. ID. NO.:48, SEQ. ID. NO.:49, SEQ. ID. NO.:50, SEQ. ID. NO.:51, SEQ. ID. NO.:52, SEQ. ID. NO.:53, SEQ. ID. NO.:54, SEQ. ID. NO.:55, SEQ. ID. NO.:56, SEQ. ID. NO.:57, SEQ. ID. NO.:58, SEQ. ID. NO.:59, SEQ. ID. NO.:60, SEQ. ID. NO.:61, SEQ. ID. NO.:62, , SEQ. ID. NO.:63, SEQ. ID. NO.:64, SEQ. ID. NO.:65, SEQ. ID. NO.:66, SEQ. ID. NO.:67, SEQ. ID. NO.:68, SEQ. ID. NO.:69, SEQ. ID. NO.:70, SEQ. ID. NO.:71, SEQ. ID. NO.:72, SEQ. ID. NO.:73, SEQ. ID. NO.:74, SEQ. ID. NO.:75, SEQ. ID. NO.:76, SEQ. ID. NO.:77, SEQ. ID. NO.:78, SEQ. ID. NO.:79, SEQ. ID. NO.:80, SEQ. ID. NO.:81, SEQ. ID. NO.:82, SEQ ID NO: 83, SEQ ID NO. 84, SEQ ID. NO. 85, SEQ. ID. NO. 86, SEQ. ID. NO. 87, SEQ. ID. NO. 88, SEQ. ID. NO. 89, SEQ. ID. NO. 90, SEQ. ID. NO. 91, SEQ ID. NO. 92, and SEQ. ID. NO. 93.

8. An isolated polynucleotide or oligonucleotide for characterizing cells that are obtained from a subject suspected of having a cancer which is associated with methylation of

IDT Exhibit 1020.8217

one or a plurality of CpG islands in the genomic DNA of malignant cells, wherein said polynucleotide or oligonucleotide comprises a sequence which is identical to or complementary to a target sequence on one of the strands of a diagnostic control fragment identified according to the method of claim 2, wherein said target sequence comprises at least two CpG dinucleotides, wherein said oligonucleotide is from 15 to 34 nucleotides in length; and wherein said polynucleotide is from 35 to 3000 nucleotides in length.

9. The isolated polynucleotide or oligonucleotide of claim 8 wherein said target sequence is located at or near the control restriction fragment end which was cleaved by the methylation sensitive restriction enzyme.

10. The isolated polynucleotide or oligonucleotide of claim 8 wherein the target sequence is located from about 100 nucleotides to about 500 nucleotides downstream of the control restriction fragment end that was cleaved by the methylation-sensitive, restriction enzyme.

11. An isolated polynucleotide or oligonucleotide for characterizing cells which are obtained from a subject suspected of having a cancer which is associated with methylation of one or a plurality of CpG islands in the genomic DNA of malignant cells, wherein said polynucleotide or oligonucleotide comprises a sequence which is identical to or complementary to a modified target sequence on one of the strands of a diagnostic control fragment identified according to the method of claim 2, wherein said modified target sequence is derived from a target sequence that has been modified by treatment with sodium bisulfite, wherein said modified target sequence lacks cytosines and comprises at least two UpG dinucleotides, wherein said oligonucleotide is from 15 to 34 nucleotides in length; and wherein said polynucleotide is from 35 to 3000 nucleotides in length.

12. The isolated polynucleotide or oligonucleotide of claim 11 wherein the modified target sequence is derived from a target sequence located at or near the control restriction fragment end that was cleaved by the methylation sensitive restriction enzyme.

13. The isolated polynucleotide or oligonucleotide of claim11 wherein the modifed target sequence is derived from a target sequence that is located from about 100 nucleotides to about 500 nucleotides downstream of the control restriction fragment end that was cleaved by the methylation-sensitive restriction enzyme.

14. An isolated polynucleotide for characterizing cells which are obtained from a subject suspected of having a cancer selected from the group consisting of glioma, acute myeloid leukemia, primitive neuroectodermal tumors of childhood, breast cancer, colon cancer, head and neck cancer, testiclular cancer and lung cancer; wherein said polynucleotide is from 35 to

IDT Exhibit 1020.8218

3000 nucleotides in length and comprises at least two CpG dinucleotides, and wherein said polynucleotide comprise a sequence which is identical to or complementary to a target sequence located within a sequence selected from the group consisting of SEQ. ID. NO.:1, SEQ. ID. NO.:2, SEQ. ID. NO.:3, SEQ. ID. NO.:4, SEQ. ID. NO.:5, SEQ. ID. NO.:6, SEQ. ID. NO.:7, SEQ. ID. NO.:8, SEQ. ID. NO.:9, SEQ. ID. NO.:10, SEQ. ID. NO:11, SEQ. ID. NO.: 12, SEQ. ID. NO.:13, SEQ. ID. NO.:14, SEQ. ID. NO.:15, SEQ. ID. NO.: 16, SEQ. ID. NO.:17, SEQ. ID. NO.:18, SEQ. ID. NO.: 19, SEQ. ID. NO.:20, SEQ. ID. NO.:21, SEQ. ID. NO.:22, SEQ. ID. NO.:23, SEQ. ID. NO.:24, SEQ. ID. NO.:25, SEQ. ID. NO.:26, SEQ. ID. NO.:27, SEQ. ID. NO.:28, SEQ. ID. NO.:29, SEQ. ID. NO.:30, SEQ. ID. NO.:31, SEQ. ID. NO.:32, SEQ. ID. NO.:33, SEQ. ID. NO.:34, SEQ. ID. NO.:35, SEQ. ID. NO.:36, SEQ. ID. NO.:37, SEQ. ID. NO.:38, SEQ. ID. NO.:39, SEQ. ID. NO.:40, SEQ. ID. NO.:41, SEQ. ID. NO.:42, SEQ. ID. NO.:43, SEQ. ID. NO.:44, SEQ. ID. NO.:45, SEQ. ID. NO.:46, SEQ. ID. NO.:47, SEQ. ID. NO.:48, SEQ. ID. NO.:49, SEQ. ID. NO.:50, SEQ. ID. NO.:51, SEQ. ID. NO.:52, SEQ. ID. NO.:53, SEQ. ID. NO.:54, SEQ. ID. NO.:55, SEQ. ID. NO.:56, SEQ. ID. NO.:57, SEQ. ID. NO.:58, SEQ. ID. NO.:59, SEQ. ID. NO.:60, SEQ. ID. NO.:61, SEQ. ID. NO.:62, , SEQ. ID. NO.:63, SEQ. ID. NO.:64, SEQ. ID. NO.:65, SEQ. ID. NO.:66, SEQ. ID. NO.:67, SEQ. ID. NO.:68, SEQ. ID. NO.:69, SEQ. ID. NO.:70, SEQ. ID. NO.:71, SEQ. ID. NO.:72, SEQ. ID. NO.:73, SEQ. ID. NO.:74, SEQ. ID. NO.:75, SEQ. ID. NO.:76, SEQ. ID. NO.:77, SEQ. ID. NO.:78, SEQ. ID. NO.:79, SEQ. ID. NO.:80, SEQ. ID. NO.:81, SEQ. ID. NO.:82, SEQ ID NO: 83, SEQ ID NO. 84, SEQ ID. NO. 85, SEQ. ID. NO. 86, SEQ. ID. NO. 87, SEQ. ID. NO. 88, SEQ. ID. NO. 89, SEQ. ID. NO. 90, SEQ. ID. NO. 91, SEQ ID. NO. 92, and SEQ. ID. NO. 93.

15. The isolated polynucleotide of claim 14 wherein said subject is suspected of having glioma and said polynucleotide comprises a sequence which is identical to or complementary to a target sequence located within SEQ. ID.NO. 78, SEQ ID NO:71, SEQ ID NO:75, SEQ ID NO:18, SEQ ID NO:4, SEQ ID NO:22, SEQ ID NO:82, SEQ ID NO:27, SEQ ID NO:38, SEQ ID NO:47, SEQ ID NO:49, SEQ ID NO:59, SEQ ID NO:36, SEQ ID NO:46, SEQ ID NO:2, SEQ ID NO:7, SEQ ID NO:20, SEQ ID NO:30, SEQ ID NO:45, SEQ ID NO:72, SEQ ID NO:70, SEQ ID NO:26, SEQ ID NO:54, or SEQ ID NO:69.

16. The isolated polynucleotide of claim 14 wherein said subject is suspected of having acute myeloid leukemia and said polynucleotide comprises a sequence which is identical to or complementary to a target sequence located within SEQ ID NO:10, SEQ ID NO:44, SEQ ID NO:48, SEQ ID NO:58, SEQ ID NO:73, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:9, SEQ ID NO:14, SEQ ID NO:24, SEQ ID NO:33, SEQ ID NO:56, SEQ ID NO:68, SEQ ID

IDT Exhibit 1020.8219

NO:76, SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:57, SEQ ID NO:26, SEQ ID NO:38, SEQ ID NO:47, SEQ ID NO:49, SEQ ID NO:59, SEQ ID NO:36, SEQ ID NO:46, SEQ ID NO:38, SEQ ID NO:47, SEQ ID NO:49, SEQ ID NO:59, SEQ ID NO:36, SEQ ID NO:46, SEQ ID NO:2, SEQ ID NO:7, SEQ ID NO:20, SEQ ID NO:30, SEQ ID NO:45, SEQ ID NO:72, SEQ ID NO:16, SEQ ID NO:55, SEQ ID NO:61, SEQ ID NO:63, or SEQ ID NO:70.

17. The isolated polynucleotide of claim 14 wherein said subject is suspected of having a primitive neuroectodermal tumor of childhood and said polynucleotide comprises a sequence which is identical to or complementary to a target sequence located within SEQ ID NO:39, SEQ ID NO:42, SEQ ID NO:50, SEQ ID NO:9, SEQ ID NO:14, SEQ ID NO:24, SEQ ID NO:37, SEQ ID NO:4, SEQ ID NO:36, SEQ ID NO:46, SEQ ID NO:72, SEQ ID NO:26, SEQ ID NO:15, SEQ ID NO:19, or SEQ ID NO:61.

18. The isolated polynucleotide of claim 14 wherein said subject is suspected of having breast cancer and said polynucleotide comprises a sequence identical or complementary to a target sequence which is located within SEQ ID NO:21, SEQ ID NO:28, SEQ ID NO:41, SEQ ID NO:80, SEQ ID NO:37, SEQ ID NO:63, SEQ ID NO:71, SEQ ID NO:75, SEQ ID NO:18, SEQ ID NO:4, SEQ ID NO:22, SEQ ID NO:82, SEQ ID NO:12, SEQ ID NO:23, SEQ ID NO:31, SEQ ID NO:34, SEQ ID NO:43, SEQ ID NO:60, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:67, SEQ ID NO:77.

19. The isolated polynucleotide of claim 14 wherein said subject is suspected of having colon cancer and said polynucleotide comprises a sequence which is identical to or complementary to a target sequence located within SEQ ID NO:11, SEQ ID NO:40, SEQ ID NO:74, SEQ ID NO:81, SEQ ID NO:53, SEQ ID NO:62, SEQ ID NO:76, SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:57, SEQ ID NO:37, SEQ ID NO:75, SEQ ID NO:18, SEQ ID NO:4, SEQ ID NO:27, SEQ ID NO:38, SEQ ID NO:47, SEQ ID NO:49, SEQ ID NO:59, SEQ ID NO:36, or SEQ ID NO:46.

20. The isolated polynucleotide of claim 14 wherein said subject is suspected of head and neck cancer and said polynucleotide comprises a sequence which is identical to or complementary to a target sequence located within in SEQ ID NO:1, SEQ ID NO. 79, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:53, SEQ ID NO:62, SEQ ID NO:76, SEQ ID NO:8, or SEQ ID NO:13.

21. The isolated polynucleotide of claim 14 wherein said subject is suspected of testiclular cancer and said polynucleotide comprises a sequence which is identical to or complementary to a target sequence located within in SEQ ID NO. 29, SEQ ID NO:33, SEQ

IDT Exhibit 1020.8220

ID NO:56, SEQ ID NO:68, SEQ ID NO:51, SEQ ID NO:57, SEQ ID NO:70, SEQ ID NO:54, or SEQ ID NO:69

22. The isolated polynucleotide of claim 14 wherein said subject is suspected of having a lung cancer and said polynucleotide comprises a sequence which is identical to or complementary to a target sequence located within SEQ ID NO: 83, SEQ ID NO. 84, SEQ ID. NO. 85, SEQ. ID. NO. 86, SEQ. ID. NO. 87, SEQ. ID. NO. 88, SEQ. ID. NO. 89, SEQ. ID. NO. 90, SEQ. ID. NO. 91, SEQ ID. NO. 92, and SEQ. ID. NO. 93.

23. An isolated CpG diagnostic oligonucleotide for characterizing cells which are obtained from a subject suspected of having a cancer selected from the group consisting of glioma, acute myeloid leukemia, primitive neuroectodermal tumors of childhood, breast cancer, colon cancer, head and neck cancer, testiclular cancer and lung cancer; wherein said olignoucleotide is from 15 to 34 nucleotides in length and comprises at least two CpG dinucleotides, and wherein said oligonucleotide comprise a sequence which is identical to a target sequence located within a region extending from nucleotide 1 through nucleotide 99 of a sequence selected from the group consisting of SEQ. ID. NO.:1, SEQ. ID. NO.:2, SEQ. ID. NO.:3, SEQ. ID. NO.:4, SEQ. ID. NO.:5, SEQ. ID. NO.:6, SEQ. ID. NO.:7, SEQ. ID. NO.:8, SEQ. ID. NO.:9, SEQ. ID. NO.:10, SEQ. ID. NO:11, SEQ. ID. NO.: 12, SEQ. ID. NO.:13, SEQ. ID. NO.:14, SEQ. ID. NO.:15, SEQ. ID. NO.: 16, SEQ. ID. NO.:17, SEQ. ID. NO.:18, SEQ. ID. NO.: 19, SEQ. ID. NO.:20, SEQ. ID. NO.:21, SEQ. ID. NO.:22, SEQ. ID. NO.:23, SEQ. ID. NO.:24, SEQ. ID. NO.:25, SEQ. ID. NO.:26, SEQ. ID. NO.:27, SEQ. ID. NO.:28, SEQ. ID. NO.:29, SEQ. ID. NO.:30, SEQ. ID. NO.:31, SEQ. ID. NO.:32, SEQ. ID. NO.:33, SEQ. ID. NO.:34, SEQ. ID. NO.:35, SEQ. ID. NO.:36, SEQ. ID. NO.:37, SEQ. ID. NO.:38, SEQ. ID. NO.:39, SEQ. ID. NO.:40, SEQ. ID. NO.:41, SEQ. ID. NO.:42, SEQ. ID. NO.:43, SEQ. ID. NO.:44, SEQ. ID. NO.:45, SEQ. ID. NO.:46, SEQ. ID. NO.:47, SEQ. ID. NO.:48, SEQ. ID. NO.:49, SEQ. ID. NO.:50, SEQ. ID. NO.:51, SEQ. ID. NO.:52, SEQ. ID. NO.:53, SEQ. ID. NO.:54, SEQ. ID. NO.:55, SEQ. ID. NO.:56, SEQ. ID. NO.:57, SEQ. ID. NO.:58, SEQ. ID. NO.:59, SEQ. ID. NO.:60, SEQ. ID. NO.:61, SEQ. ID. NO.:62, , SEQ. ID. NO.:63, SEQ. ID. NO.:64, SEQ. ID. NO.:65, SEQ. ID. NO.:66, SEQ. ID. NO.:67, SEQ. ID. NO.:68, SEQ. ID. NO.:69, SEQ. ID. NO.:70, SEQ. ID. NO.:71, SEQ. ID. NO.:72, SEQ. ID. NO.:73, SEQ. ID. NO.:74, SEQ. ID. NO.:75, SEQ. ID. NO.:76, SEQ. ID. NO.:77, SEQ. ID. NO.:78, SEQ. ID. NO.:79, SEQ. ID. NO.:80, SEQ. ID. NO.:81, SEQ. ID. NO.:82; SEQ ID NO: 83, SEQ ID NO. 84, SEQ ID. NO. 85, SEQ. ID. NO. 86, SEQ. ID. NO. 87, SEQ. ID. NO. 88, SEQ. ID. NO. 89, SEQ. ID. NO. 90, SEQ. ID. NO. 91, SEQ ID. NO. 92, and SEQ. ID. NO. 93; or a sequence which is the reverse complement of a target sequence located in a region

IDT Exhibit 1020.8221

extending from about nucleotide 100 through nucleotide 500 in said SEQ ID NO.

24. The isolated oligonucleotide of claim 23 wherein said subject is suspected of having glioma and said oligonucleotide comprises a sequence which is identical to a target sequence located within a region extending from nucleotide 1 through nucleotide 99 in SEQ ID 78, SEQ ID NO:71, SEQ ID NO:75, SEQ ID NO:18, SEQ ID NO:4, SEQ ID NO:22, SEQ ID NO:82, SEQ ID NO:27, SEQ ID NO:38, SEQ ID NO:47, SEQ ID NO:49, SEQ ID NO:59, SEQ ID NO:36, SEQ ID NO:46, SEQ ID NO:2, SEQ ID NO:7, SEQ ID NO:20, SEQ ID NO:30, SEQ ID NO:45, SEQ ID NO:72, SEQ ID NO:70, SEQ ID NO:26, SEQ ID NO:54, or SEQ ID NO:69, or said oligonucleotide comprises a sequence which is the reverse complement of a target sequence located in a region extending from about nucleotide 100 through nucleotide 500 in said SEQ ID NO.

25. The isolated oligonucleotide of claim 23 wherein the subject is suspected of having acute myeloid leukemia; and said oligonucleotide comprises a sequence which is identical to a target sequence located within a region extending from nucleotide 1 through nucleotide 99 SEQ ID NO:10, SEQ ID NO:44, SEQ ID NO:48, SEQ ID NO:58, SEQ ID NO:73, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:9, SEQ ID NO:14, SEQ ID NO:24, SEQ ID NO:33, SEQ ID NO:56, SEQ ID NO:68, SEQ ID NO:76, SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:57, SEQ ID NO:26, SEQ ID NO:38, SEQ ID NO:47, SEQ ID NO:49, SEQ ID NO:59, SEQ ID NO:36, SEQ ID NO:46, SEQ ID NO:38, SEQ ID NO:47, SEQ ID NO:49, SEQ ID NO:59, SEQ ID NO:36, SEQ ID NO:46, SEQ ID NO:2, SEQ ID NO:7, SEQ ID NO:20, SEQ ID NO:30, SEQ ID NO:45, SEQ ID NO:72, SEQ ID NO:16, SEQ ID NO:55, SEQ ID NO:61, SEQ ID NO:63, or SEQ ID NO:70, or said oligonucleotide comprises a sequence which is the reverse complement of a target sequence located in a region extending from about nucleotide 100 through nucleotide 500 in said SEQ ID NO.

26. The isolated oligonucleotide of claim 23 wherein said subject is suspected of having primitive neuroectodermal tumors of childhood; and said oligonucleotide comprises a sequence which is identical to a target sequence located within a region extending from nucleotide 1 through nucleotide 99 of SEQ ID NO:39, SEQ ID NO:42, SEQ ID NO:50, SEQ ID NO:9, SEQ ID NO:14, SEQ ID NO:24, SEQ ID NO:37, SEQ ID NO:4, SEQ ID NO:36, SEQ ID NO:46, SEQ ID NO:72, SEQ ID NO:26, SEQ ID NO:15, SEQ ID NO:19, or SEQ ID NO:61, or said nucleotide comprises a sequence which is the reverse complement of a target sequence located in a region extending from about nucleotide 100 through nucleotide 500 in said SEQ ID NO.

27. The isolated oligonucleotide of claim 23 wherein said subject is suspected of having

IDT Exhibit 1020.8222

breast cancer; and said oligonucleotide comprises a sequence which is identical to a target sequence located within a region extending from nucleotide 1 through nucleotide 99 of SEQ ID NO:21, SEQ ID NO:28, SEQ ID NO:41, SEQ ID NO:80, SEQ ID NO:37, SEQ ID NO:63, SEQ ID NO:71, SEQ ID NO:75, SEQ ID NO:18, SEQ ID NO:4, SEQ ID NO:22, SEQ ID NO:82, SEQ ID NO:12, SEQ ID NO:23, SEQ ID NO:31, SEQ ID NO:34, SEQ ID NO:43, SEQ ID NO:60, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:67, SEQ ID NO:77, or said oligonucleotide comprises a sequence which is the reverse complement of a target sequence located in a region extending from about nucleotide 100 through nucleotide 500 in said SEQ ID NO.

28. The isolated oligonucleotide of claim 23 wherein said subject is suspected of having colon cancer; and said oligonucleotide comprises a sequence which is identical to a target sequence located within a region extending from nucleotide 1 through nucleotide 99 of SEQ ID NO:11, SEQ ID NO:40, SEQ ID NO:74, SEQ ID NO:81, SEQ ID NO:53, SEQ ID NO:62, SEQ ID NO:76, SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:57, SEQ ID NO:37, SEQ ID NO:75, SEQ ID NO:18, SEQ ID NO:4, SEQ ID NO:27, SEQ ID NO:38, SEQ ID NO:47, SEQ ID NO:49, SEQ ID NO:59, SEQ ID NO:36, SEQ ID NO:46 or said oligonucleotide comprises a sequence which is the reverse complement of a target sequence located in a region extending from about nucleotide 100 through nucleotide 500 in said SEQ ID NO.

29. The isolated oligonucleotide of claim 23 wherein said subject is suspected of having head and neck cancer; and said oligonucleotide comprises a sequence which is identical to a target sequence located within a region extending from nucleotide 1 through nucleotide 99 in SEQ ID NO:1, SEQ ID NO. 79, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:53, SEQ ID NO:62, SEQ ID NO:76, SEQ ID NO:8, or SEQ ID NO:13., or said nucleotide a sequence which is the reverse complement of a target sequence located in a region extending from about nucleotide 100 through nucleotide 500 in said SEQ ID NO.

30. The isolated oligonucleotide of claim 23 wherein said subject is suspected of having testiclular cancer; and said oligonucleotide comprises a sequence which is identical to a target sequence located within a region extending from nucleotide 1 through nucleotide 99 in SEQ ID NO. 29, SEQ ID NO:33, SEQ ID NO:56, SEQ ID NO:68, SEQ ID NO:51, SEQ ID NO:57, SEQ ID NO:70, SEQ ID NO:54, or SEQ ID NO:69, or said oligonucleotide comprises a sequence which is the reverse complement of a target sequence located in a region extending from about nucleotide 100 through nucleotide 500 in said SEQ ID NO.

31. The isolated oligonucleotide of claim 23 wherein said subject is suspected of having lung cancer; and said oligonucleotide comprises a sequence which is identical to a target

IDT Exhibit 1020.8223

sequence located within a region extending from nucleotide 1 through nucleotide 99 of SEQ ID NO SEQ ID NO: 83, SEQ ID NO. 84, SEQ ID. NO. 85, SEQ. ID. NO. 86, SEQ. ID. NO. 87, SEQ. ID. NO. 88, SEQ. ID. NO. 89, SEQ. ID. NO. 90, SEQ. ID. NO. 91, SEQ ID. NO. 92, and SEQ. ID. NO. 93, or said oligonucleotide comprises a sequence which is the reverse complement of a target sequence located in a region extending from about nucleotide 100 through nucleotide 500 in said SEQ ID NO.

32. A method for determining whether cells obtained from a subject suspected of having a cancer are malignant or non-malignant, comprising:

a) digesting DNA which has been isolated from the cells with a methylation-sensitive restriction enzyme to provide a set of restriction fragments;

b) hybridizing said restriction fragments with a CpG diagnostic polynucleotide comprising a sequence which is identical to or complementary to a target sequence on one of the strands of a diagnostic control fragment identified according to the method of claim 2, wherein said method employed said methylation-sensitive restriction enzyme, wherein said target sequence comprises at least two CpG dinucleotides, wherein said polynucleotide is from 35 to 3000 nucleotides in length, and wherein said reaction is conducted under stringent hybridization conditions; and

c) assaying the reaction products of step b to determine the size or the sequence of the restriction fragment to which the CpG diagnostic polynucleotide has hybridized.

33. The method of claim 32 wherein said subject is suspected of having a cancer selected from the group consisting of glioma, acute myeloid leukemia, primitive neuroectodermal tumors of childhood, breast cancer, colon cancer, head and neck cancer, and testiclular cancer;

wherein said DNA sample is digested with NotI; and

wherein said polynucleotide comprises a sequence which is identical to or complementary to a target sequence located in SEQ. ID. NO.:1, SEQ. ID. NO.:2, SEQ. ID. NO.:3, SEQ. ID. NO.:4, SEQ. ID. NO.:5, SEQ. ID. NO.:6, SEQ. ID. NO.:7, SEQ. ID. NO.:8, SEQ. ID. NO.:9, SEQ. ID. NO.:10, SEQ. ID. NO:11, SEQ. ID. NO.: 12, SEQ. ID. NO.:13, SEQ. ID. NO.:14, SEQ. ID. NO.:15, SEQ. ID. NO.: 16, SEQ. ID. NO.:17, SEQ. ID. NO.:18, SEQ. ID. NO.: 19, SEQ. ID. NO.:20, SEQ. ID. NO.:21, SEQ. ID. NO.:22, SEQ. ID. NO.:23, SEQ. ID. NO.:24, SEQ. ID. NO.:25, SEQ. ID. NO.:26, SEQ. ID. NO.:27, SEQ. ID. NO.:28, SEQ. ID. NO.:29, SEQ. ID. NO.:30, SEQ. ID. NO.:31, SEQ. ID. NO.:32, SEQ. ID. NO.:33, SEQ. ID. NO.:34, SEQ. ID. NO.:35, SEQ. ID. NO.:36, SEQ. ID. NO.:37, SEQ. ID. NO.:38, SEQ. ID. NO.:39, SEQ. ID. NO.:40, SEQ. ID. NO.:41, SEQ. ID. NO.:42, SEQ. ID. NO.:43, SEQ. ID. NO.:44,

IDT Exhibit 1020.8224

SEQ. ID. NO.:45, SEQ. ID. NO.:46, SEQ. ID. NO.:47, SEQ. ID. NO.:48, SEQ. ID. NO.:49, SEQ. ID. NO.:50, SEQ. ID. NO.:51, SEQ. ID. NO.:52, SEQ. ID. NO.:53, SEQ. ID. NO.:54, SEQ. ID. NO.:55, SEQ. ID. NO.:56, SEQ. ID. NO.:57, SEQ. ID. NO.:58, SEQ. ID. NO.:59, SEQ. ID. NO.:60, SEQ. ID. NO.:61, SEQ. ID. NO.:62, , SEQ. ID. NO.:63, SEQ. ID. NO.:64, SEQ. ID. NO.:65, SEQ. ID. NO.:66, SEQ. ID. NO.:67, SEQ. ID. NO.:68, SEQ. ID. NO.:69, SEQ. ID. NO.:70, SEQ. ID. NO.:71, SEQ. ID. NO.:72, SEQ. ID. NO.:73, SEQ. ID. NO.:74, SEQ. ID. NO.:75, SEQ. ID. NO.:76, SEQ. ID. NO.:77, SEQ. ID. NO.:78, SEQ. ID. NO.:79, SEQ. ID. NO.:80, SEQ. ID. NO.:81, and SEQ. ID. NO.:82.

34. The method of claim 32 wherein the subject is suspected of having lung cancer;

wherein the DNA is digested with AscI;

and wherein said polynucleotide comprises a sequence which is identical to or complementary to a target sequence located in , SEQ ID NO: 83, SEQ ID NO. 84, SEQ ID. NO. 85, SEQ. ID. NO. 86, SEQ. ID. NO. 87, SEQ. ID. NO. 88, SEQ. ID. NO. 89, SEQ. ID. NO. 90, SEQ. ID. NO. 91, SEQ ID. NO. 92, and SEQ. ID. NO. 93.

35. A method of determining whether cells contained within a tissue sample obtained from a subject suspected of having cancer are malignant, comprising:

a) treating DNA isolated from the tissue sample with a compound which converts non-methylated cytosines to a different nucleotide base;

b) reacting a portion of the treated DNA with a CpG diagnostic oligonucleotide which is complementary to a target sequence which comprises CpG islands that are preferentially methylated in malignant cells of subjects known to have said cancer;

c) reacting a portion of the treated DNA with a modified CpG diagnostic oligonucleotide which is complementary to a modified target sequence in which the cytosines in said target sequence are replaced with the different nucleotide base; and

d) assaying the reaction products of step b and step c to determine whether the treated DNA has hybridized with the CpG diagnostic oligonucleotide or the modified CpG diagnostic oligonucleotide; wherein hybridization of the treated DNA with the CpG diagnostic oligonucleotide as opposed to the modified CpG diagnostic oligonucleotide indicates that the DNA has been obtained from malignant cells.

36. The method of claim 35 wherein the chemical compound is sodium bisulfite and the non-methylated cytosines are converted to uracil.

37. The method of claim 35 wherein the assay is a polymerase chain reaction,

wherein a portion of the treated DNA is reacted with a first primer set which comprises two diagnostic CpG oligonucleotides; and

IDT Exhibit 1020.8225

wherein a portion of the treated DNA is reacted with a second primer set which comprises two modified diagnostic CpG oligonucleotides.

38. The method of claim 35 wherein the subject is suspected of having a cancer selected from the group consisting of glioma, acute myeloid leukemia, primitive neuroectodermal tumors of childhood, breast cancer, colon cancer, head and neck cancer, testiclular cancer and lung cancer; and

wherein the CpG diagnostic oligonucleotide comprises a sequence which is identical to a target sequence located between nucleotide 1 and 100 in a sequece selected from the group consisting said polynucleotide comprises a sequence which is identical to or complementary to a target sequence located in SEQ. ID. NO.:1, SEQ. ID. NO.:2, SEQ. ID. NO.:3, SEQ. ID. NO.:4, SEQ. ID. NO.:5, SEQ. ID. NO.:6, SEQ. ID. NO.:7, SEQ. ID. NO.:8, SEQ. ID. NO.:9, SEQ. ID. NO.:10, SEQ. ID. NO:11, SEQ. ID. NO.: 12, SEQ. ID. NO.:13, SEQ. ID. NO.:14, SEQ. ID. NO.:15, SEQ. ID. NO.: 16, SEQ. ID. NO.:17, SEQ. ID. NO.:18, SEQ. ID. NO.: 19, SEQ. ID. NO.:20, SEQ. ID. NO.:21, SEQ. ID. NO.:22, SEQ. ID. NO.:23, SEQ. ID. NO.:24, SEQ. ID. NO.:25, SEQ. ID. NO.:26, SEQ. ID. NO.:27, SEQ. ID. NO.:28, SEQ. ID. NO.:29, SEQ. ID. NO.:30, SEQ. ID. NO.:31, SEQ. ID. NO.:32, SEQ. ID. NO.:33, SEQ. ID. NO.:34, SEQ. ID. NO.:35, SEQ. ID. NO.:36, SEQ. ID. NO.:37, SEQ. ID. NO.:38, SEQ. ID. NO.:39, SEQ. ID. NO.:40, SEQ. ID. NO.:41, SEQ. ID. NO.:42, SEQ. ID. NO.:43, SEQ. ID. NO.:44, SEQ. ID. NO.:45, SEQ. ID. NO.:46, SEQ. ID. NO.:47, SEQ. ID. NO.:48, SEQ. ID. NO.:49, SEQ. ID. NO.:50, SEQ. ID. NO.:51, SEQ. ID. NO.:52, SEQ. ID. NO.:53, SEQ. ID. NO.:54, SEQ. ID. NO.:55, SEQ. ID. NO.:56, SEQ. ID. NO.:57, SEQ. ID. NO.:58, SEQ. ID. NO.:59, SEQ. ID. NO.:60, SEQ. ID. NO.:61, SEQ. ID. NO.:62, , SEQ. ID. NO.:63, SEQ. ID. NO.:64, SEQ. ID. NO.:65, SEQ. ID. NO.:66, SEQ. ID. NO.:67, SEQ. ID. NO.:68, SEQ. ID. NO.:69, SEQ. ID. NO.:70, SEQ. ID. NO.:71, SEQ. ID. NO.:72, SEQ. ID. NO.:73, SEQ. ID. NO.:74, SEQ. ID. NO.:75, SEQ. ID. NO.:76, SEQ. ID. NO.:77, SEQ. ID. NO.:78, SEQ. ID. NO.:79, SEQ. ID. NO.:80, SEQ. ID. NO.:81, SEQ. ID. NO.:82, SEQ ID NO: 83, SEQ ID NO. 84, SEQ ID. NO. 85, SEQ. ID. NO. 86, SEQ. ID. NO. 87, SEQ. ID. NO. 88, SEQ. ID. NO. 89, SEQ. ID. NO. 90, SEQ. ID. NO. 91, SEQ ID. NO. 92, and SEQ. ID. NO. 93; or a sequence which is the reverse complement of a target sequence located in a region extending from about nucleotide 100 through nucleotide 500 in said SEQ ID NO.

39. An isolated polynucleotide for characterizing cells which are obtained from a subject suspected of having a cancer selected from the group consisting of glioma, acute myeloid leukemia, primitive neuroectodermal tumors of childhood, breast cancer, colon cancer, head and neck cancer, testiclular cancer and lung cancer; wherein said polynucleotide is from 35 to

IDT Exhibit 1020.8226

1000 nucleotides in length and comprises at least two CpG dinucleotides, and wherein said polynucleotide comprise a sequence which is identical to or complementary to a target sequence which originates between nucleotide 1 and nucleotide 15 of SEQ. ID. NO.:1, SEQ. ID. NO.:2, SEQ. ID. NO.:3, SEQ. ID. NO.:5, SEQ. ID. NO.:6, SEQ. ID. NO.:7, SEQ. ID. NO.:8, SEQ. ID. NO.:9, SEQ. ID. NO.:10, SEQ. ID. NO:11, SEQ. ID. NO.: 12, SEQ. ID. NO.:13, SEQ. ID. NO.:14, SEQ. ID. NO.:17, SEQ. ID. NO.:18, SEQ. ID. NO.: 19, SEQ. ID. NO.:20, SEQ. ID. NO.:21, SEQ. ID. NO.:22, SEQ. ID. NO.:23, SEQ. ID. NO.:24, SEQ. ID. NO.:25, SEQ. ID. NO.:27, SEQ. ID. NO.:28, SEQ. ID. NO.:29, SEQ. ID. NO.:30, SEQ. ID. NO.:31, SEQ. ID. NO.:32, SEQ. ID. NO.:33, SEQ. ID. NO.:34, SEQ. ID. NO.:35, SEQ. ID. NO.:36, SEQ. ID. NO.:37, SEQ. ID. NO.:38, SEQ. ID. NO.:40, SEQ. ID. NO.:41, SEQ. ID. NO.:42, SEQ. ID. NO.:43, SEQ. ID. NO.:44, SEQ. ID. NO.:45, SEQ. ID. NO.:47, SEQ. ID. NO.:48, SEQ. ID. NO.:49, SEQ. ID. NO.:50, SEQ. ID. NO.:51, SEQ. ID. NO.:52, SEQ. ID. NO.:53, SEQ. ID. NO.:54, SEQ. ID. NO.:55, SEQ. ID. NO.:56, SEQ. ID. NO.:58, SEQ. ID. NO.:60, SEQ. ID. NO.:62, , SEQ. ID. NO.:63, SEQ. ID. NO.:64, SEQ. ID. NO.:65, SEQ. ID. NO.:66, SEQ. ID. NO.:67, SEQ. ID. NO.:69, SEQ. ID. NO.:70, SEQ. ID. NO.:72, SEQ. ID. NO.:73, , SEQ. ID. NO.:75, SEQ. ID. NO.:76, SEQ. ID. NO.:77, SEQ. ID. NO.:78, SEQ. ID. NO.:79, SEQ. ID. NO.:80, SEQ. ID. NO.:81, SEQ. ID. NO.:82, SEQ ID NO: 83, SEQ. ID. NO. 86, SEQ. ID. NO. 87, SEQ. ID. NO. 88, SEQ. ID. NO. 89, SEQ. ID. NO. 90, SEQ. ID. NO. 91, SEQ ID. NO. 92, and SEQ. ID. NO. 93.

IDT Exhibit 1020.8227

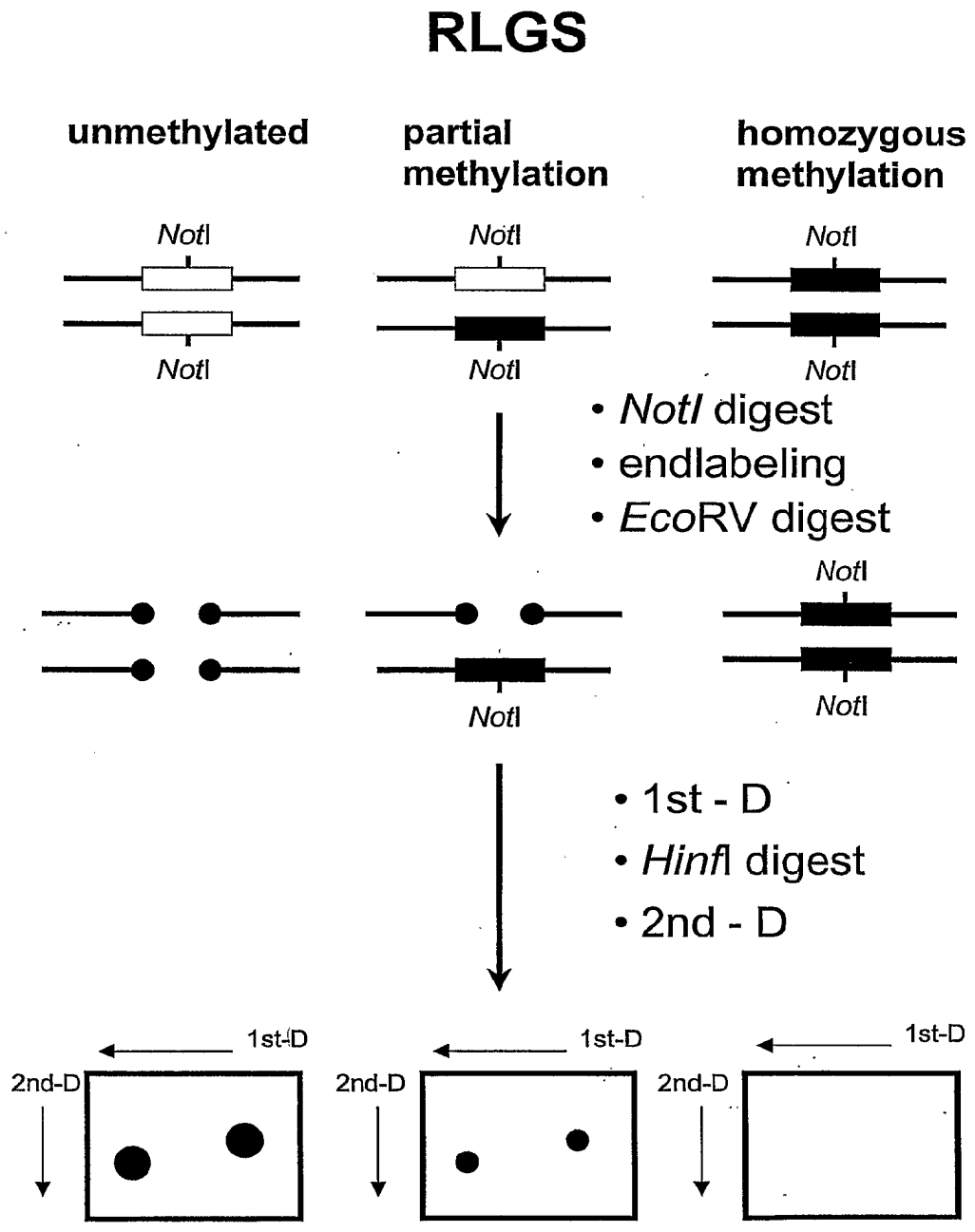


Fig. 1A

IDT Exhibit 1020.8228

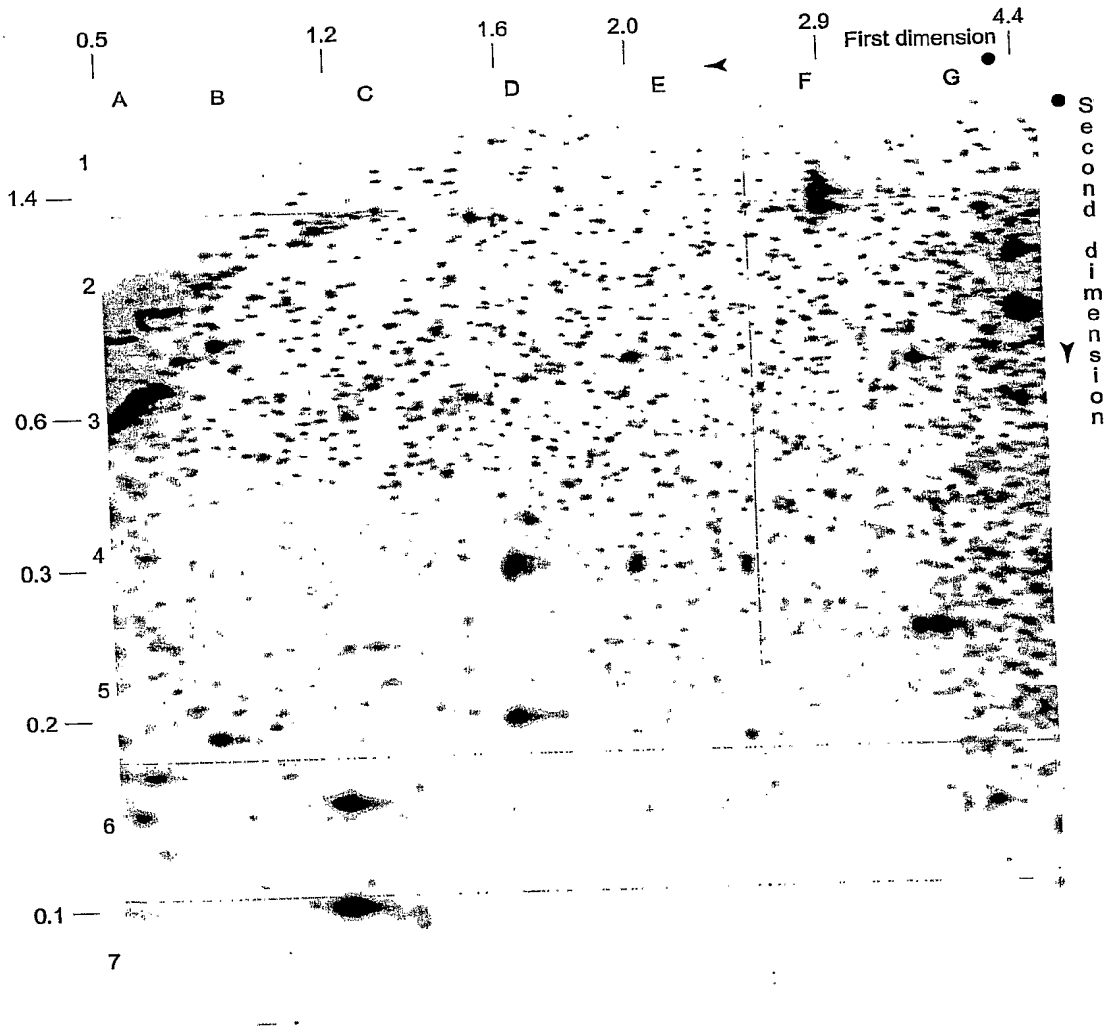


Fig. 1B

IDT Exhibit 1020.8229

Section D2

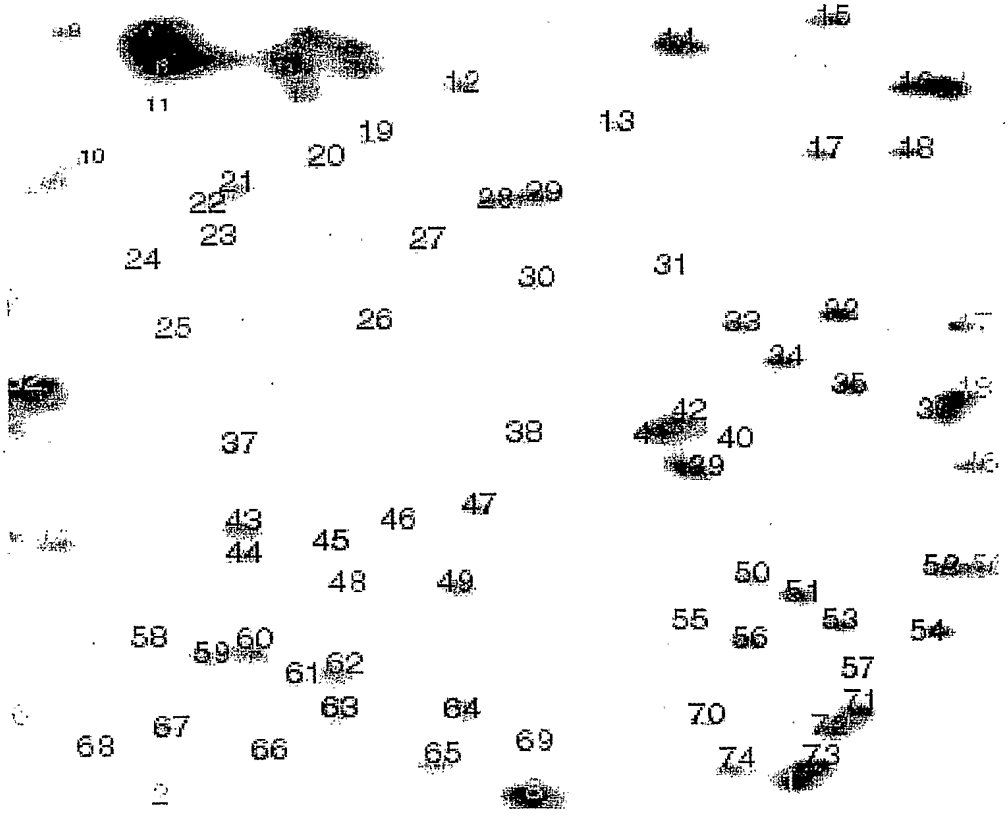


Fig. 1C

IDT Exhibit 1020.8230

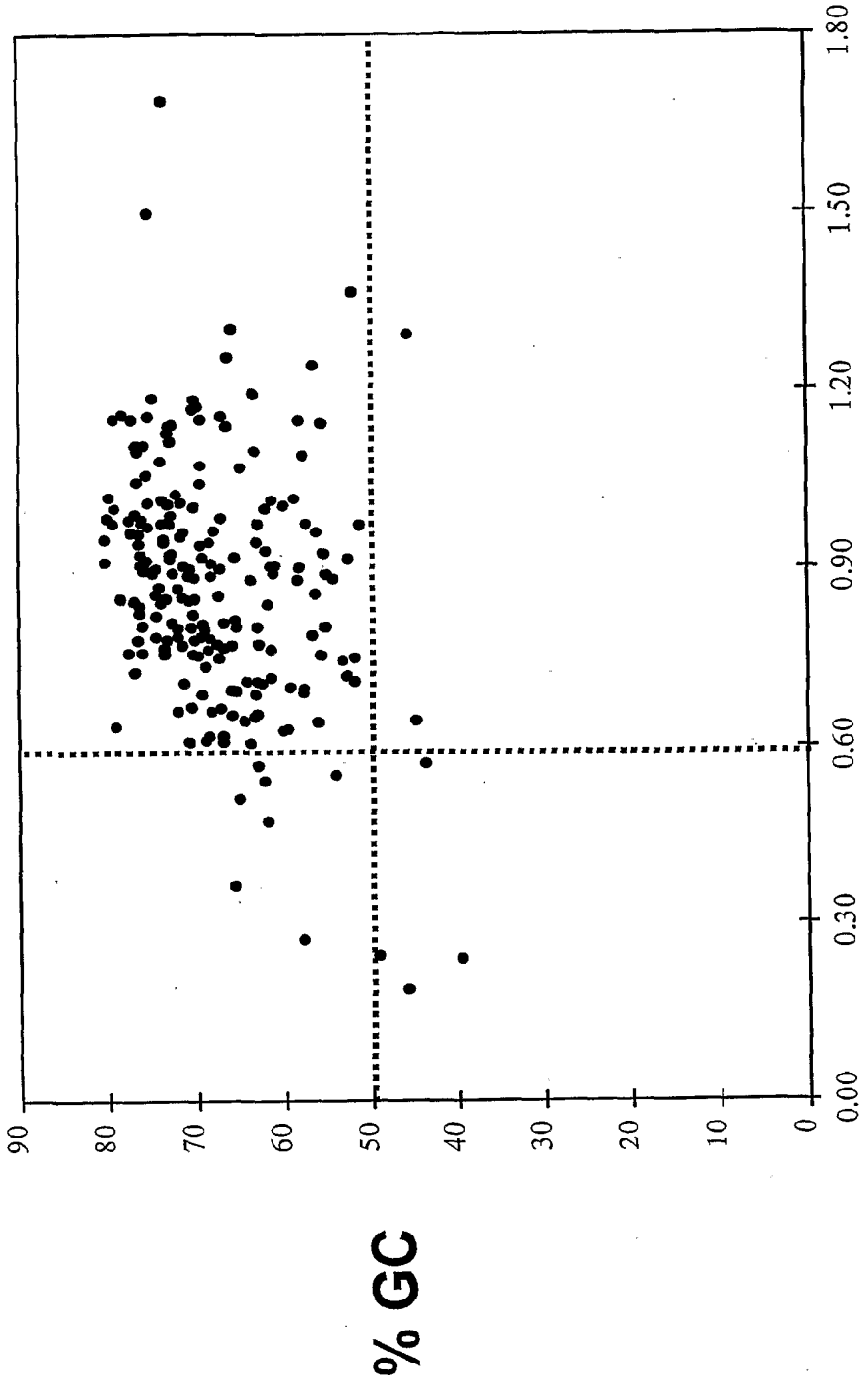


Fig. 1D

IDT Exhibit 1020.8231

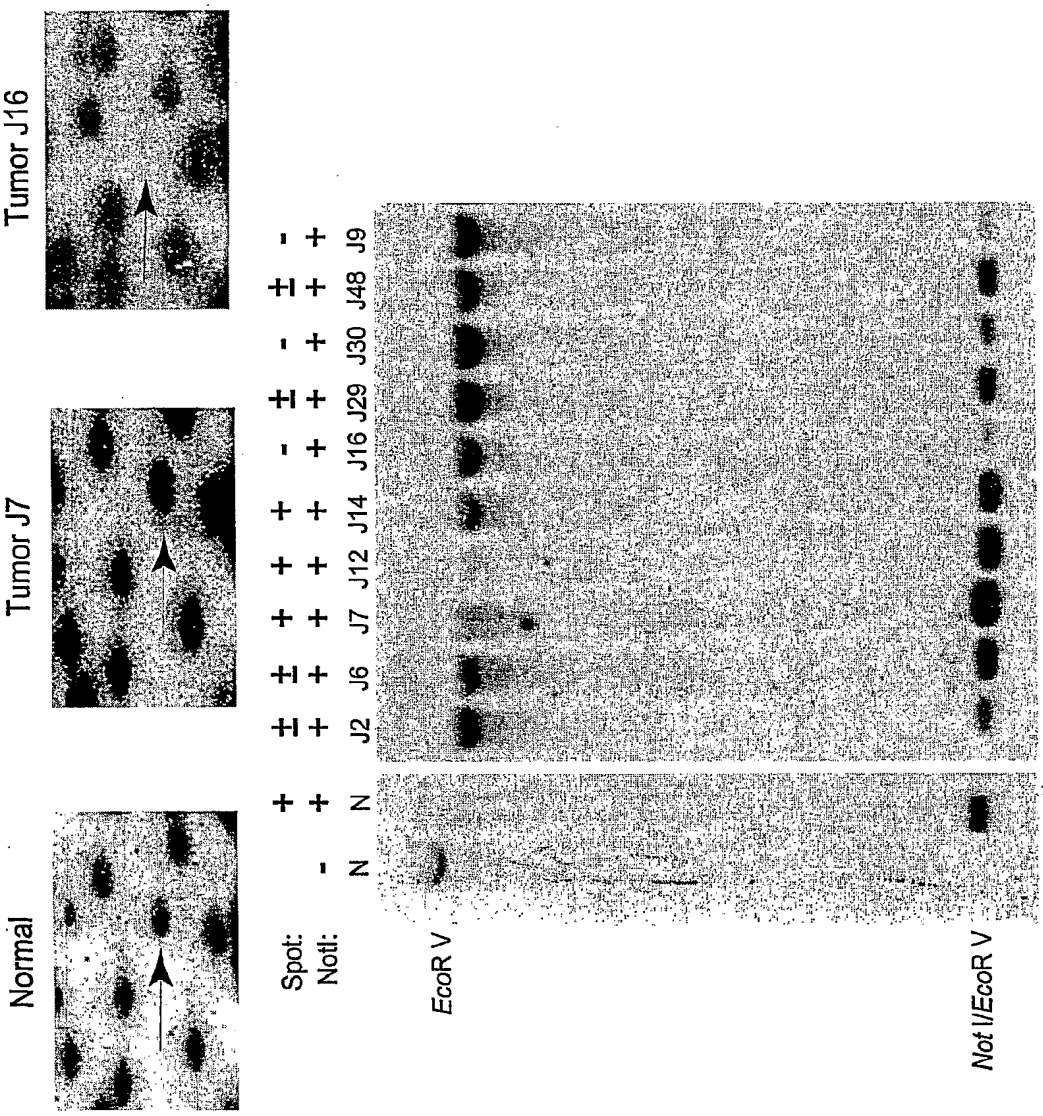


Fig. 2A

IDT Exhibit 1020.8232

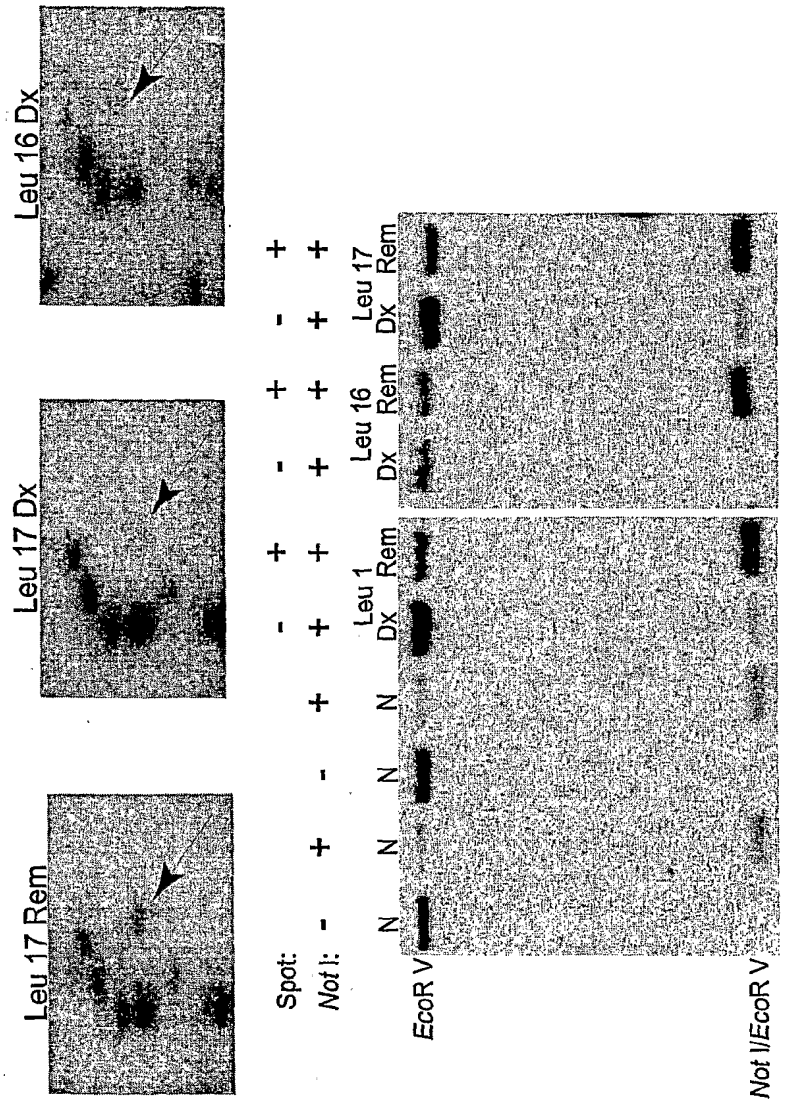


Fig. 2B

IDT Exhibit 1020.8233

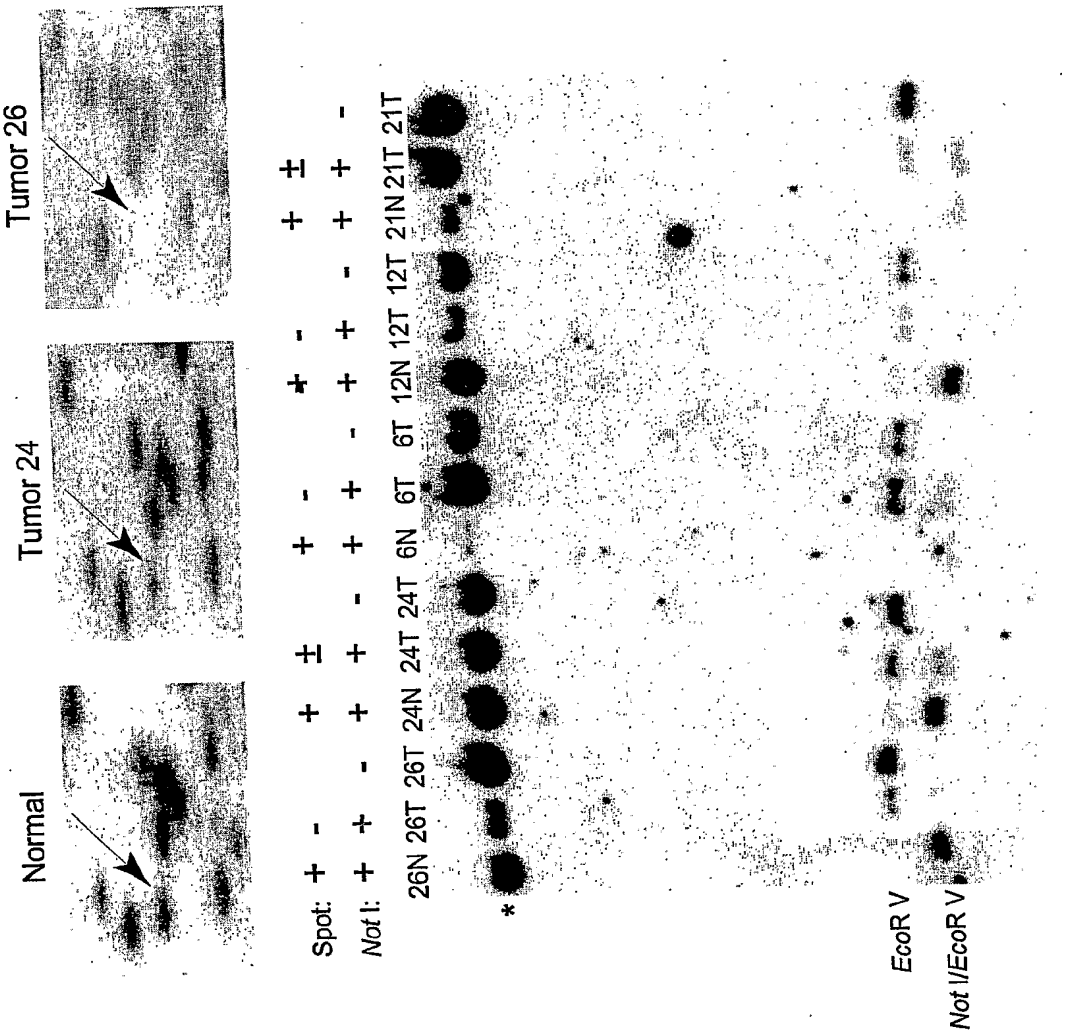


Fig. 2C

IDT Exhibit 1020.8234

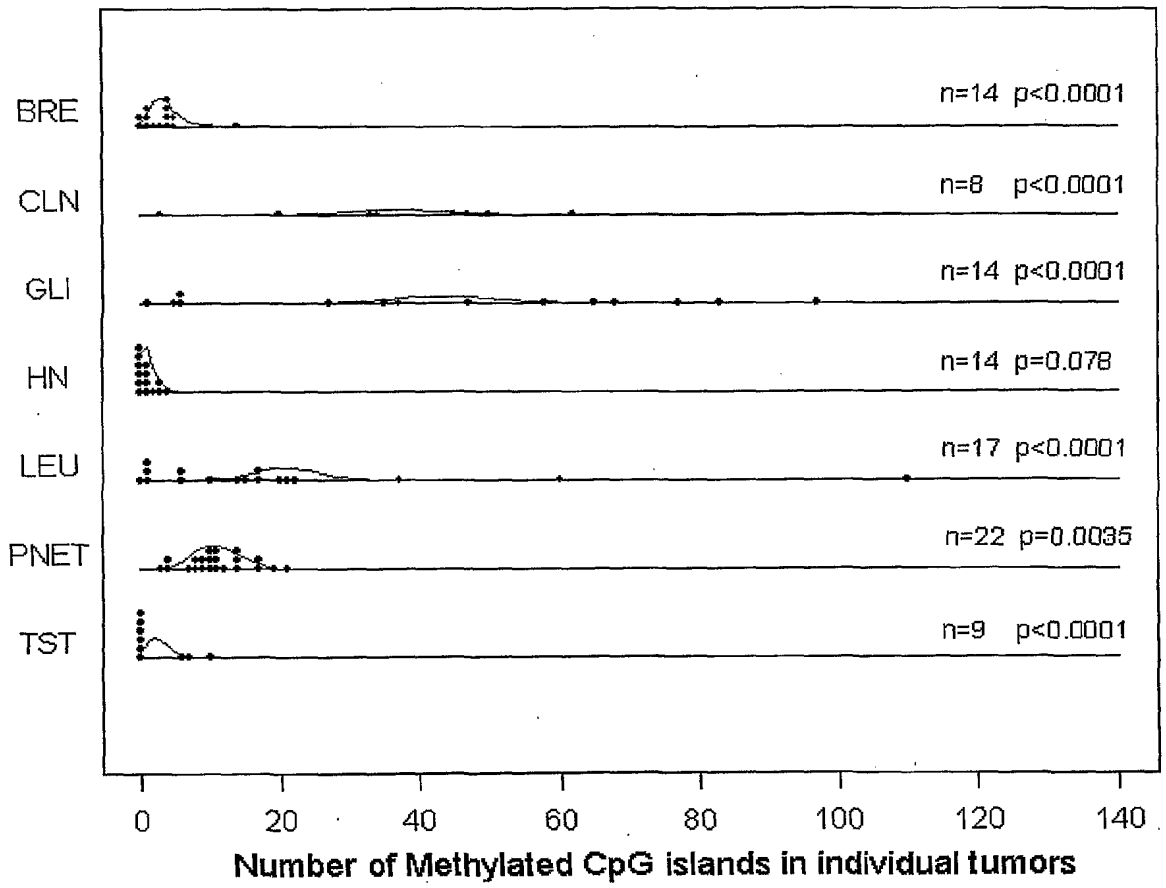


Fig. 3

IDT Exhibit 1020.8235

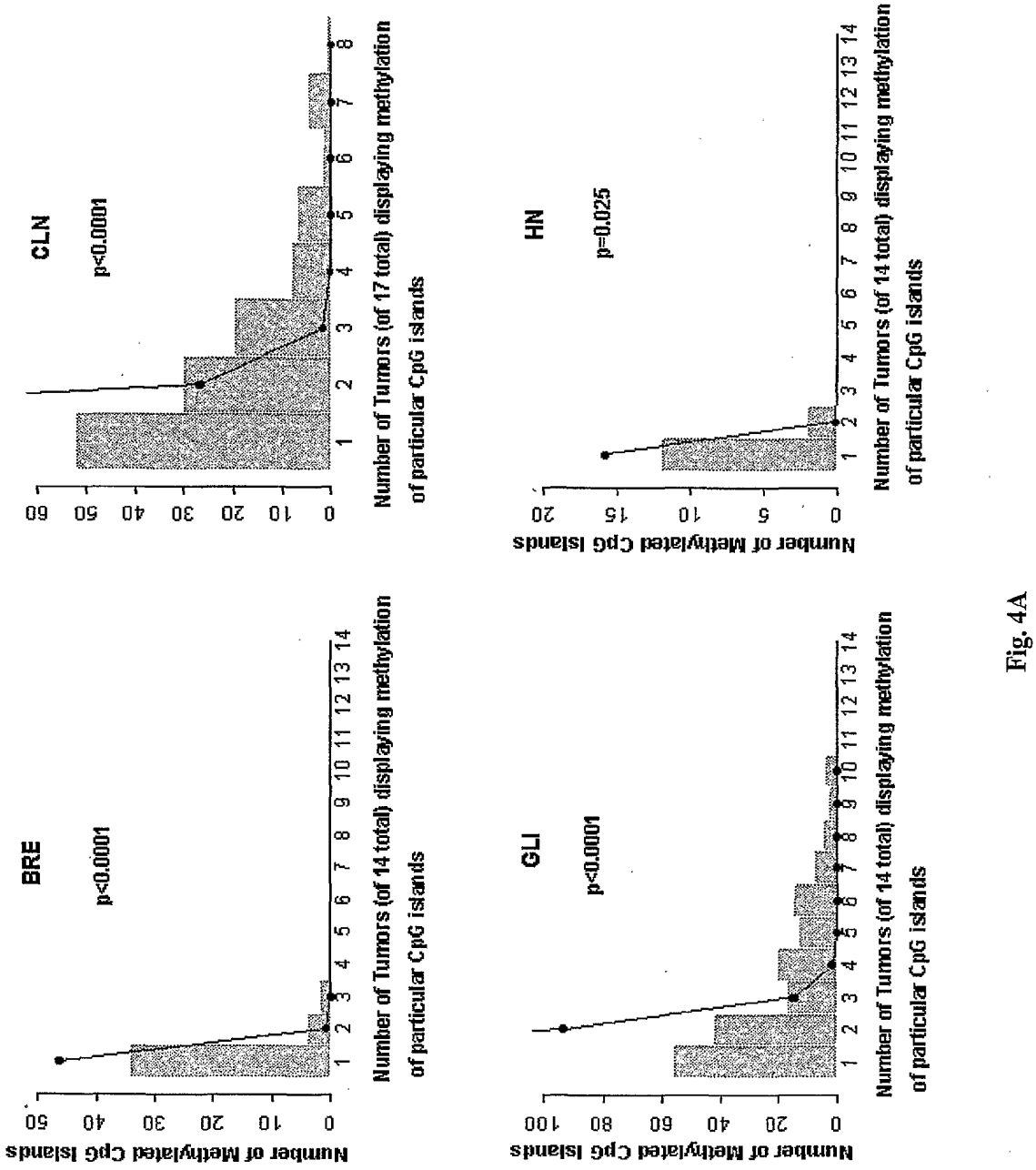

CLN
p<0.0001
Number of Tumors (of 17 total) displaying methylation of particular CpG islands
Number of Methylated CpG Islands
HN
p=0.025
Number of Tumors (of 14 total) displaying methylation of particular CpG islands
Number of Methylated CpG Islands
BRE
p<0.0001
Number of Tumors (of 14 total) displaying methylation of particular CpG islands
Number of Methylated CpG Islands
GLI
p<0.0001
Number of Tumors (of 14 total) displaying methylation of particular CpG islands
Number of Methylated CpG Islands

Fig. 4A

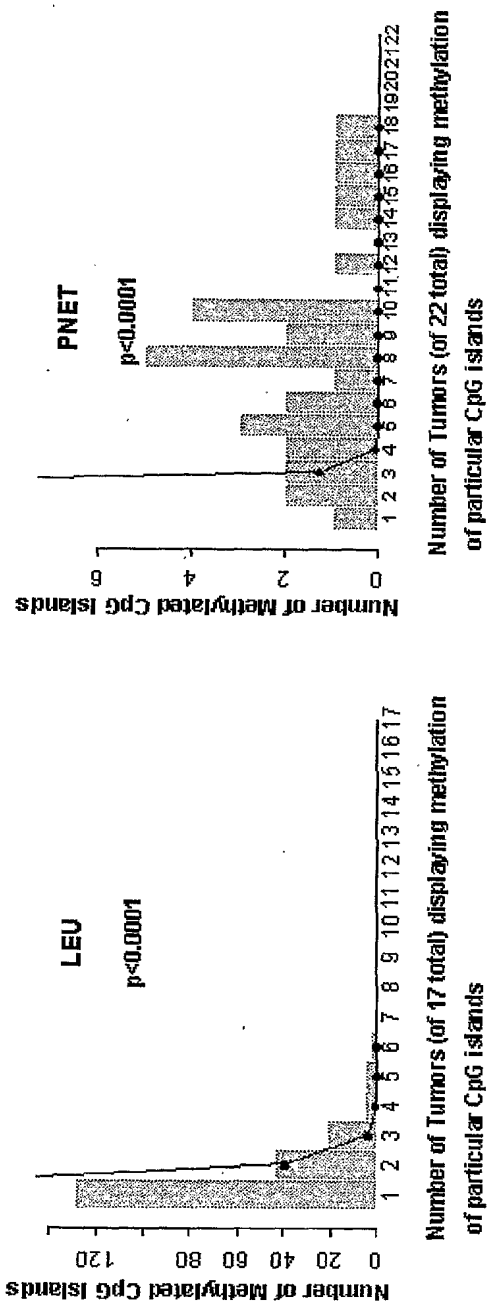


Fig. 4B

IDT Exhibit 1020.8237

SEQUENCE LISTING

<110> The Ohio State University Research Foundation
Plass, Christoph

<120> Detection of Methylated CpG Rich Sequences Diagnostic for Malignant Cells

<130> 22727/04110

<150> US 09/775,398
<151> 2001-01-31

<160> 90

<170> PatentIn version 3.0

<210> 1
<211> 677
<212> DNA
<213> Homo sapiens 2.B.53

<220>
<221> n
<222> (578)..(578)
<223> a or g or c or t

<400> 1
gcggccgcgg ttagcttctc ctgtccgaac gcagggtttc actggggcgc cgctacggtt 60

cctatggcaa cgcggctcct cgacgcagcc caggagtcgc ggtcgcggga ggctgcgccg 120

cgcaccgagc tcttccctgt ggccgccgca gccgccagcc tcttcctgct catgcttttc 180

ctcatcttca tctcggtctg agtgggctct ggacctctcc accagcctct gccccagaac 240

tgttaactgc gggggggaaa aaaggaattt gtcgtcgcaa cgcgcgttcc gatggagccg 300

cacgccacaa aggaagactc atgctgcacc ccgcggggca gatgcggcga cactggacat 360

cgctgcacag ctgggtctgc ccgtttccag agctgcttag cgccgacgcc cataaatgag 420

gaggactccc tgtgtattaa aaggggatc cgcagggttt aatttgataa ggattatagc 480

cttcataaag gcatttttaa caaaaagatg taggtggcat ggtaatcgag tattatttac 540

gcatctctcc gcacacgcac tcatacctga aaacgttntg gcaggcacaa aatgattttt 600

ttgtgtataa aagaatgtgt gtaactcgtg gatggtgggg ttcagcagga caagatagtg 660

acattagata aattaca 677

<210> 2
<211> 380
<212> DNA
<213> Homo sapiens 2.C.24

<220>
<221> n
<222> (246)..(246)
<223> a or g or c or t

<220>

IDT Exhibit 1020.8238

<221> n
<222> (297)..(297)
<223> a or g or c or t

<220>
<221> n
<222> (318)..(318)
<223> a or g or c or t

<220>
<221> n
<222> (325)..(325)
<223> a or g or c or t

<220>
<221> n
<222> (327)..(327)
<223> a or g or c or t

<220>
<221> n
<222> (345)..(345)
<223> a or g or c or t

<400> 2
gcggccgcct tgaaggcgct ggacgggatg gtgctgaagt cggtgaagga gccccggcag 60

gtgagctcgc ggcccgccag cccgctgccc acgcagtagt ggaagaggcc gaagtagcca 120

ggcttggggg tgctcacgct gtcgcccacc cagtagggct ggatgaagac caccacgttg 180

atgatggcga agcagatggt gaagatggcc cacagcacgc cgatggcccg cgagttccgc 240

atgtantgct cgtggtagag cttggaacct cctgcgaggg cagcatggtg cccggangcg 300

gggccggcgg cggctgtngc tggcngnggc cgtcggcccg ggacngacgc ctggctgccg 360

ggcgggaact ggggactcac 380

<210> 3
<211> 566
<212> DNA
<213> Homo sapiens 2.C.29

<400> 3
gcggccgccg cgctagtgac tacttcctcc tactccttct cctcctgctc cggcctcctg 60

gcgccctgct ccaggctctc cggcgccctg ctccaggctc tccggcgccc tccagccagg 120

caccggccga accgggtagt gccgcaaggt gtaattactg ctttgaaact ttaaaggcat 180

ttggaaagaa actacgggtt atgcttactt tttttgtttt tgattattat tttgtaggag 240

acacaaagtt taaaaataga aagcaaaaag tgtgacacat ttaaagagtt aaaggaaata 300

aacgtttcca atttaccttа taacatgatt ttcatacact ggatttgttt aaaacagact 360

gactacatgg ataacttttc taggaattgt tcttaactct gatagctggc tcaactgatg 420

taggcattaa aataacgtca tattaccatc tttcctccac gaattgatga tatttgacta 480

IDT Exhibit 1020.8239

tagctttgtc agggttatgt ccaactattg tataatatgt gtcagtttcc tattgctacc 540

gtaacaaatt accccaaatt tactgg 566

<210> 4
<211> 1297
<212> DNA
<213> Homo sapiens 2.C.35

<220>
<221> n
<222> (1046)..(1046)
<223> a or g or c or t

<400> 4
gttcacttct cgctgcgccg cgggttctgt agaagcgcaa gaatggggct gattattccg 60

gtgcccacat gccgcccccа cacgcccccа cccccgtccg gcgcaagact tcccttggcc 120

aaaagaggcg tttaattagt tctggggccg cggagagcca gcgtggccga caaagcccgg 180

ctccccaggt aacccgggtt ccctgcggac ccgggagggg gcgcgcgggg ccggagcacc 240

ggccttgggc tgcgcgctcc ctccggcgac actgccgtcc ccctggcctc cggcccggtc 300

ccccgcaggc caaaggctca tctgccgggc ttgggtggcc cgggccagcg ccgcctgcgg 360

tccccgagtg cggctggctc taaggccggc gcccctctccc cggctttcag tgctcagagc 420

caggccagcg ggaaagaagg cagcatggtc cgcaaaagac aggtggcagt ggcagtcttg 480

catgatactt gtccttcttc cctgttcccc attttgggga aacactggaa acacttttct 540

ctttatgcgc attcgcgtct cagcaccgag tgctccaagc cctgcgcgca gcgccgggct 600

tggaaggcgg cgaatggctg cctagccgcc gcccctacta gtgacactcg gccgccagcc 660

cccgcccagg atgtgcacat ctgctggcag cactggcccg ggtggcagtc accgggccac 720

ccactccaca ggtacaaccg cacccaatcc aacctggaac tcggagggct gtgcgcgccg 780

agctgggatc gcgccccaac gagccgggcc tttggctgcg ccagggccca ggccgagtca 840

tccccccgct cgcgtcgccg cgaggcggga caccgtgtaa tacctttgcc gtgggctggg 900

cgtcggccgc gggccggaga gcgggtgtcc cacctcgcct catcatttga tttccgccag 960

cgtctgagga cggcgcaccc aattcgttcc actcgctgcg ctctgtgaac cagcggcggg 1020

cagggcgggg gaggccgggc tggggnaggg cagggtggtc ccaatccccc gcccccgccc 1080

cgccggcctc gcggagcaca agtgttggga ttcccacggg caggcgtgct ctgcggctgg 1140

aggcccgagc gcccagggcc caggagacgt ggcggacaca gaggggtttg taggcacggt 1200

gacctccgtg ctcctgctct gaaagggcct gaaaggagcg gtttatggtg cattaccagt 1260

caagggctca ggtaccagcg cctgtgtcgg gaacccg 1297

<210> 5
<211> 651
<212> DNA

IDT Exhibit 1020.8240

<213> Homo sapiens 2.C.54

<400> 5
gcggccgccg cgggggacgc tcagatctcg cgagaagagg gcgagcgcgc tgccccctg 60

gtgggcgggg cgaagcccgg gagagggtgg gcgccaccgg aggggaggag gggaacaggg 120

aactgaagga agtgggaggg gccggcgggg cggggaagcg gaaaggggge gtggctgagg 180

gcgggaggat taagctgcct ttttgaaagt ggagcgccag gtcccgggtt ctgggtggag 240

gtggttgctg attggtggag ctcggagcgg cggttgggag ggtcctggtc acatggtggg 300

gagtgggagg ggggaagttc ggagagcggg agcgggatgg tagtgggctg ggccccactg 360

ggctgggaca gcaggaggat agtcttgagg aggagcgtgg ggtgctagat gtgtaactac 420

gtcccgaact ggttcctgtg ttttttctagg gcatgtggac tagggatggg tacttgagta 480

gaagcctgca acttgaagag tttgtgcagg agttagctgc agtgtcggaa attagtgtcc 540

tgtatgctca acaaggtatt cggactgggt gtgcacacca cagctctcag gactggaagg 600

tggaaattta atctacgaag ttcccttaaa ctgcataagc ttcgggacct c 651

<210> 6
<211> 710
<212> DNA
<213> Homo sapiens 2.C.57

<220>
<221> n
<222> (652)..(652)
<223> a or g or c or t

<220>
<221> n
<222> (690)..(690)
<223> a or g or c or t

<220>
<221> n
<222> (695)..(695)
<223> a or g or c or t

<220>
<221> n
<222> (710)..(710)
<223> a or g or c or t

<400> 6
gcggccgcac ggagttgaag acactaaccc agctaagcca catacagacc ctcacggccg 60

cctggtctac acaggccgcc acagctacac aggctcaggc ctcagcctgg tcacaatggt 120

cacacccaca ctctcgggtc ccacagtttt gcgggagcgg tgacacacac ccgctcccaa 180

ctgaccacgc ccacacacgc tggcttcagc cgcacacgca cacagtagcc acgccccctt 240

atgctccagc cttgccagca cccgccctcg ccacgctggt cacgcccaca cacacacaca 300

IDT Exhibit 1020.8241

cacacacaca cacacgcacg caggcctggg gcacgcccct cccccacacg caggcgtgcg 360

gcacgccttc ccatacacac acacacgcgc gcgggcctgg ggcacgccct ccacacacat 420

gcaggggtag ggcacgcccc cacacacaca cacgccggcc tggggcacgc tcgcgcgcac 480

atgcacacac atacacacgc acaggcctgg ggcaggcccc acccccacac acgcaggcct 540

ggggcacgcc cccccacaca tgcaggcctg gagcatgcgc acactcgcag gccttgggca 600

cacgcgcaca cactcatgca cagacacgca cgcacacatc gagccccgcc cncggaagca 660

catgagaggc acttgctttc actgactgan ggcanggctt tgggcccgcn 710

<210> 7
<211> 1204
<212> DNA
<213> Homo sapiens 2.C.58

<400> 7
gcggccgctc ctctttattc tactctcacc cgaggcccgc gcccgtcccg gggagcggct 60

ctgccaggaa aacggcccga ccagtgcccg gcgcctgggc tgcgtccgag cccaccttct 120

tccctcgtcg tcgtctccca gactaaatcc cggaaaggga aagcgggatg tttgcgccca 180

ccgcgctgta gctggtcctg acacttgcaa aatggtcagt ggctcctgct cggccaggct 240

gagtgtgtgc gtgtgtgtga gcaagggagc gagggtgtgc ggtgtgcagg gggtgcgctg 300

tgtgtgcgcg cgtctccggg aaggtctcgc ggcggctgga gccgggactg acagcccggg 360

cggagcgcag gcagctccac acgctaaacc tctcgcctct cccctcaccc ccaccccttc 420

cactccctc tccttccccc accctccccg gcccccttcca agctctctga ttggccaatg 480

ggacaaaagt ttctgtggag acggctgggc gctgacgtca cgggcagaat tgtcccattt 540

agggatcccs ggggcagtgc gcatgctgca ggctgcaggt tagaggcaga aggaggtagc 600

agcgggcccg gcggcagcca ggtggcagaa aggagcacgc agcatccagg tgggggggacg 660

actccagcag ggtttccatg gagattcctc tgggtctagc ctaaaaacag cagatcagct 720

gacaccatta gctcaggacc taattactgc ttattggagc aacaaatgag ggaaagggcc 780

agctgcaaag gaagagtttt tatcccccca ccccattccc ccatctcctt tctcccctc 840

tctccatccc tcttgagtcc cgggtgaatt ctcattaact tgcaagattc ctgcaacaac 900

agctcccctt ctccagaggc caccccgact gcttttattc ttttatttcc ttcttttgta 960

ttaaaaagaa atgctaaaat aaatcagttg ttgagtcctt gaatttttgt tcaatacgta 1020

ttagaccata gagctcagag aagacactgt ccaatgaagt cacaagtgaa tctaatacaa 1080

gggactcagg ggaaaaatat cactttcaat ttattgaggt gaatctttag atatttcaca 1140

ttaaaaaaat cttaatatct taaatacata aatatttgaa acacgcaatt ggacagaaga 1200

tatc 1204

<210> 8
<211> 687

IDT Exhibit 1020.8242

<212> DNA
<213> Homo sapiens 2.C.59

<220>
<221> n
<222> (650)..(650)
<223> a or g or c or t

<400> 8
gcggccgcac aagcgcacac gcacacgtcc agggcggagg aacactacta gtaacacccg 60
cctccttcta gcctccctat cccaaagtta tggtgccgat tttgtccgcg gcaggggctc 120
caggggcaca ctcataaatt cggtgcggag gaacacaact agcagcacca cacccccgcc 180
actgccagaa ccaaagtgac ggtgccgaca cccctccgca agcgcaaggc cgacttccat 240
aagtaattag ccagagcacc gtcccgttcc tgtcagcacc gagccccagc caggacaccg 300
gtattcccag caccatacaa gaactacttt ttcgatgaag caacccaaaa gctgcgagcg 360
gttcccggtg aggccgccca ctcacctggc cggcgcagac aagctccgtg cgtcaagaca 420
taacagcgta agtgtacgac gttgcgcagc gacgcggggg ccttcgggaa atgtagtcta 480
caactggaaa ccggccggat cgtgtctgcg caggcccagc agctaagatc gggtccggcg 540
ctccagaaca gaacgatccc tgaggctccc ttgctcgaac tgtgggactt accctactat 600
ggtccgagcc taccctattt cattatactc aagtaacgcc ccagaaattn cagagaatct 660
acacaaagag gttgagtctt gccgtgg 687

<210> 9
<211> 1520
<212> DNA
<213> Homo sapiens 2.D.10

<400> 9
gcggccgcga ggacagctcg gacgggggag agaaaggagg tttccagtaa aaataataac 60
gccagagaga aaaccgtaac tcgcgtgaca cagacagaaa tttccagtaa taatcatcag 120
gtgatagaga aggaaggctt ccaaaatgaa gaacaagtga aataaaggtt ttagtcatga 180
attacagcac gtgcgatgga tgagtggtga tttctcatca taaatggtaa ctcgggagat 240
agagaaacgt gtccagccct aaactacaac agggtttggt ttgaaagaga ggtgctgtca 300
taaagcggaa ctcaggggat ggggaagacg gcctccgtcc caaatgacaa ctcaatgaca 360
gagaacaaaa gatccaaact aaagtgatgg agaaaaaggg tttccaacca ccacacaaat 420
gaagagaaag actgatcaca taatgaagta ttcagtcatt aatacatgat aaacccggtg 480
atagagaaag aggcttagtc acaaattact cagataatgg agaaaaaagc cttattcatg 540
tatcactcag gtagatacat caaggcaggt ttcctgccat aaaggataac acagctaaaa 600
gagaaataaa ggttttagta ataagtgaca attcatataa cagagaaaga aggcttctgg 660
ccataaggat aactcatgta ataaagaaaa gttttagtca taaataatag agaaagaaag 720
gtttccgata gaaaatggta gagatagaaa ggttctaggt aacaaacggt aactgaagtg 780

IDT Exhibit 1020.8243

atagagcaag gtcacaaata ataactcagg taatagagaa agatttctag tcataaataa 840

tacatctgct acagaaataa gggttttgat tcataaagtt atgtcataag tgataagtgg 900

tagaaaagga aaggttttag ttataaatta tgattcaagg gatagaaaaa caaaggtttc 960

aagttataaa tatcatttca atggtcaaga aaggttttca gtcatgaatg aaaactgggt 1020

gaagttttcc agtcacaggt tataactcag gcaatggaca gagaaggaaa gatttttgtc 1080

atcaatcaac tcaggtggag aaggaaaggt ttttcaataa gaaataactc agttgagtga 1140

aagaaggctt gaggtcatga atgataatta ggtgatagag aaagaaatgt tccagtcata 1200

agggttaaat cagatgctag agaaagaaag gtttttagtc ataaataaaa ctcagctgct 1260

agaaagaata gggctaccag tcataattga taactcaggt gagagaaaga ttgctggtca 1320

taaattgtaa cccaggtgac agaaaagaag gtgtcactca cacatgataa ttcgggttat 1380

gaggaaggtt tccagccaca gtggtaactc aggtgctagg gaaagaaggt ttgggcaata 1440

atgacaactc aggtaataca gaaaaacgat tacagtcata aatgacagag aaggaaaggc 1500

ttttattcat aaaggatatc 1520

<210> 10
<211> 575
<212> DNA
<213> Homo sapiens 2.D.14

<400> 10
gcggccgcgg ctgtggctcc tcttggccgc gcagctgaca ggtaaggcgg cggcgcgcgg 60

gctacccaag ggtctgcgct cccggggcct gagcggggag gtgataagtg gctgtcctgg 120

ccctggtcct ggcagggtgc agcgtcgagc ccgcggtggc ggggcgcccg ggaggcagct 180

tggcaggcac ggtccctaag ggtggaaata aaataccccc atatcgcatt accccggggg 240

accggagagc ccctggactg aggccacctc ccctcaaaag cctggacgca ggagaagggg 300

aggcagtgaa aaggggagcg agtgagggaa ggaaagagag ggtcgctgga ggtcaccagg 360

ggaaggaaac aggtccctgc ccagggtccc cgcaggatgt gctcggagga aggttggcca 420

ggccatgggt cctgtggaca catttttatt acttccgggg aagtgtttgt agtacaatca 480

gacaaacatc gggcgttctc agttctcgga gggctagggc agggtgatcc ctctggctcc 540

cgttctccct gatgtcgctg gtgttgggtg tcatg 575

<210> 11
<211> 741
<212> DNA
<213> Homo sapiens 2.D.20

<220>
<221> n
<222> (691)..(691)
<223> a or g or c or t

IDT Exhibit 1020.8244

<220>
<221> n
<222> (732)..(732)
<223> a or g or c or t

<400> 11
gcggccgcgt cgtcgctgag tacaccagct gcctcatcta tctggagccc ggcctccatc 60

tcgccaggct cagcgcccgc gtccgtgtcg gtgccggagc cattggccgc gcctagcaac 120

acctcgtgta tgcagcgctc cgtagctgca ggcgccgcca ccgcagcagc ctcttatccc 180

atgtcctacg gccagggcgg cagctacggc caaggctacc ctacgccctc ctcttcctac 240

tttggcggcg tggactgcag ctcataccta gcgcccatgc actcacatca ccacccgcac 300

cagctgagcc ccatggcacc ctcctccatg gcgggccacc atcatcacca cccacatgcg 360

caccacccgt tgagccagtc ctcaggccac caccaccacc atcaccacca ccaccaccaa 420

ggctacggtg gctctgggct tgccttcaac tctgccgact gcttggatta caaggagcct 480

ggcgccgctg ctgcttcctc cgcctggaaa ctcaacttca actcccccga ctgtctggac 540

tataaggacc aagcctcatg gcggttccag gtcttgtgag cccaggaatg aaagaggaga 600

agaaacgcaa ctacctgcgc cctccgtggt cccgatcctg ttgctgctgc tgcaccgccc 660

gcctttgcct cgtcttctcc aaaaactgat tntcaccccc caaaagatgt ccattgcctg 720

cactgccgcc cncatttttg t 741

<210> 12
<211> 458
<212> DNA
<213> Homo sapiens 2.D.25

<220>
<221> n
<222> (333)..(333)
<223> a or g or c or t.

<220>
<221> n
<222> (372)..(372)
<223> a or g or c or t

<220>
<221> n
<222> (426)..(426)
<223> a or g or c or t

<400> 12
gcggccgcca gtagcagagc ccagcacatt gcgggtgccc agttcatctt cgtggggtta 60

aacctgcggg aagagaggga aagggccctt agtttccatg gagatcgggt gcccaggggc 120

ggagggctca aggctggaga gcagagggac ccccatcttt tgtgggatca gggtgccccc 180

agcatcttgg aggcccactg aggcctgggg gggcgcggtt taacttctag catcagggac 240

IDT Exhibit 1020.8245

ttaggcctgg gggaggcgct gggaagtggc aggtggggca ggagggttct gcacctgaag 300

gttgtgcacc tggattgggg gtgtagaagc ggngcaggag cgccgcggtg ggggcgtcca 360

ggcccgggcg gnggagcaag cctgggggag ggagctctgc acgcgttgct gggatgtggg 420

gggcgngggg aggcggcatg gggggagggg cgttgtgt 458

<210> 13
<211> 615
<212> DNA
<213> Homo sapiens 2.D.27

<400> 13
gcggccgccc ggcgtcccgc tctggggggc cgggaccgaa gcgctcacgg cccggggacg 60

cggggttggt ccaggctgcg gcctgtggcg cgtgcaggcc tgaaggaggc gagatgccga 120

tgccgccacc gctggtccgg tggaccaggc cccttggtcc agcctcccct cccgcagccg 180

cccgtctggg ggtgttcgca gccccgggct ccccccggccc gcccgccggg gagtgggagg 240

gcgatggcgc cccgcctccg gctcttacgg agagcgcgcc tccccctcaa ctccggcggc 300

ggtgagccgg ggtgcgatgc gcggccgagg cctcgcccgg accgccggtc cccatcgcgt 360

ccctgggcga gggaggggcg gttggccgga gatggcggag gggcgtaccc gccccgcctg 420

cccgccgtcc ccagccctca gcgcctgggg aagcccctgc tgtggcagtg ctcgggcgct 480

atccggagga agaggagcag ttcctctttc ttggctgcgg cagggctgct tgggccggaa 540

aactaacttg tgtcggcgcc cagccgcccc gcgccggctg ccggctagct caggccgacg 600

ccgaggggag cggcg 615

<210> 14
<211> 669
<212> DNA
<213> Homo sapiens 2.D.34

<400> 14
gcggccgcgc ggggcagcgc gaggaactgt tgatttgcct gcgccttggg cccctgcgtc 60

tctcccaggc ggcggctccc gctttcctca aaggccgtgt cgggtttgtt gtttggtgtg 120

ggtgccggga aagggcgctt ctccccagtg aggtggggaa cttgggtgat gggaccacgg 180

aggcgccggt tcgtgcccgg tggggacggg tgaggcaggg gagagtgaga ttttattctc 240

ccccaaggaa ggagtgtccc cttctcctta ttttgagggc tattcaagct tattgaaacc 300

agaaagcggt gtttcttgtc aatctctcag ccccttcttc caaccaagaa caattgtcga 360

tgagtttcca tcacaggcgc ttgtgagaga accggtaaac ccagtacagc aaaatccaag 420

cccttggttt ccacatgcat tttgctagca gtttttggca ttgaccctcg ccctcccgtg 480

tttccactcg acatcattta gcgtttgagg tttttttccc tcctcaaaat tgcaaatgag 540

aaaaaaagag gaaaccagga aaaggggtg gggggtagca tttaaattgg atgtgagttt 600

ctgctgagaa ttctagcgaa gtcccctgta cactgaagcg ccgagagatt tttccgtttg 660

IDT Exhibit 1020.8246

tgtatcttc 669

<210> 15
<211> 998
<212> DNA
<213> Homo sapiens 2.D.40

<220>
<221> n
<222> (929)..(929)
<223> a or g or c or t

<400> 15
gatatccatt ataatactat ttgacctcaa agtgaatttt attgttccac acaagcaaca 60

gattacacca atttcacaac tcccagaatc caaacctaca aagacccttc ccaccaagca 120

ctttaccaaa aacgggcttc atctccatct tcctttcttt cacagttgaa aaactgccct 180

tcctaattaa gccaaccaac ttcttacctc aataaaatcc ttgtttttca gtagcatgta 240

cagtatttcc agtgatgaac agtgaactgt ctttcgtctc acacagtaac ctccgtgaag 300

aagatccacc ttgttcttta ctgtatattc ctggcatgct aactgcatcc tcagacaatt 360

ttaagtgact gaaaactcag gcaaagaaag gcaagagggc aaatagaagg gcacaggaga 420

caacgctttt caaatttttc tcactgcgac ctacagaaac acactgtaga acacctccta 480

gtacactcac acgtgtgtgt acacctgaag tgtcaagaaa caatacccta agtgcaacac 540

cctctgatat tttctatttc aagtggccgt gatctactaa actgatttcc aactcaccaa 600

taggattcag tttgaaaaac actgcaataa atcaaacctt acagttgcat tccacaagct 660

actaatgaac tcttgaaaat ccagcataca gcagagacgc tgaccaacta caagatccaa 720

acccccccagg tgggcagtgt ccttctgttc agcagtggca gttccccacc accaccagcc 780

ctgagagtta attatctccc aaactcccag agtttcccaa gtagcctgag gtgtctgtca 840

tatgcccttt taacctcttt ataaattcag tcccgtccgt ctcttacggt ggcaaagttc 900

atttatcgtc ggctgtggaa agcaatacnt tctttttgtc cccttcagga acccagaatt 960

aatgaccagg ttggtgcccg gtgtgccttt atgatcta 998

<210> 16
<211> 797
<212> DNA
<213> Homo sapiens 2.D.48

<220>
<221> n
<222> (679)..(679)
<223> a or g or c or t

<220>
<221> n
<222> (687)..(687)
<223> a or g or c or t

IDT Exhibit 1020.8247

<400> 16
gcccctctga gttacgggga gccctgcaga cacccagccc ctggggatcc tctccccgac 60

ctgcccttcc cctccgacac ttgccagtac tccccggcct ggtattcctt tcgagacccc 120

ctcacctatt ccaggctgtc ctccactgag gcgaagctct atgaagtagc ccaatttcaa 180

tataattcac gttgtgtaaa agaactttga agacggacta catcgtgcaa ggacaccgtc 240

acccgaaaac cattggtgga acgttaaaac aaacaaaaaa caaaacggca aaacctttt 300

gaaggcaatt ttgacattta tgaatttaca gttattattc ggtttgtccc tgaaatgtca 360

cttctgaaaa tttgcatagt tttcattatc actaaaataa tctagtaaat attcccgaat 420

gaatgcattc aagaatattc actaaattat tttagtgata aggaaaaagt ggaaatagct 480

gacagtcatc aatttataaa taaaatgatg gttaaataaa atgatgaaca ttcatataaa 540

ggaatactct atattcagac gagatctgtg tgctcacagg caaacaggtc taagcttact 600

ttaaatgaaa aaggataaat tgcaaaaaga atagtttgtg taatatgatt ccacatttgt 660

aaaaatggag aaagaaatng taagcanatg tctgcaagca atcagatatg attagtgact 720

taatttcatg gatagttata taggaaatat atgtatattt tatatgcaca tagatatgga 780

ggaatatact ttcactg 797

<210> 17
<211> 1024
<212> DNA
<213> Homo sapiens 2.D.55

<220>
<221> n
<222> (499)..(499)
<223> a or g or c or t

<220>
<221> n
<222> (768)..(768)
<223> a or g or c or t

<220>
<221> n
<222> (846)..(846)
<223> a or g or c or t

<220>
<221> n
<222> (847)..(847)
<223> a or g or c or t

<400> 17
gcggccgcgg cgctgcacgg gcgtgacgtc atggcgccgc ggagccgcgt cctccccgcc 60

ccgcccccgg ccggggtcac ccacccgctg ccggggctga cagagaccct ggcccgcggt 120

ctgcagcctc ctcagtcgtg cgtgcgttca ttccgctcat agcttctgtc actcagcaag 180

IDT Exhibit 1020.8248

cgctcaacac agacgcatga gataccctgg ctggaaggcc ctgaaaggta gtcgtccatt 240

caacacgtgc ttagcgcgct gctgatctgt gccaggcact gggccagggc cccgacacgc 300

gtcagggtag aagcaagcag aagcctggcc ctgttggagc ttacattggt aaataaccaa 360

gataatttca ggtaaatatt aggtcctatt aaaaatatgc gtcttcgcca ggcgcggtgg 420

atcacgcctg taatctcagc actttgagag gtcgagcacg ggcggatctc ctgaggtcaa 480

gagttcgaga ccaacctgng taaatggtga aaccgcatct ctacaaacat acaaaaaaaa 540

aaattagcag tgagctgtga gcttgcacca ctgcactcca gtctgggcaa caggacgaga 600

tcttctaaca acaacaaaaa aaaagtatgg gccacctagt ccagccaaaa aaacaaagtg 660

ctttttttt gctttttttt ttttttttt ttttgagat ggagtctcgc tgtgtcgccc 720

aggctggagt gcaggggcgc gatctcagct cactggaagc tccacctncc gggtttacgc 780

cattctcctg gctcagcctc ccgagtagct gggactacag gcacatgcca ccatgcctgg 840

ctaatnnttt gatttttgg ttgggtgttt agtagagacg ggttcatcgt gtagccagat 900

ggctaactct gactgtgatc tgcacttgcc tccagtgtgg atacagggga ccacttgcag 960

caaagctcta ttcctgtagg aggggtgttg tgaatcagac ccaatttgga aatcaaattc 1020

tagt 1024

<210> 18
<211> 1854
<212> DNA
<213> Homo sapiens 2.D.74

<220>
<221> n
<222> (258)..(258)
<223> a or g or c or t

<400> 18
gcggccgctg cagaccctgc tccaggcgcc gtagccttgc aggaagagca gacaaagaca 60

ggagagaggc aaagcgccgc ttgcccagag atgcagtcgg ctcagtcaat agagggaaat 120

cgcctccaaa cccaggctgg gaatgaggga ggaggggcga ggcggctggg gactagaaaa 180

agcagcaggg aattaacgtg acagtcagag cccagccagt gcctcgccgg cgctgctctc 240

tcgcctcgcg gttgcggngt ccggaatgga gagaggaggc gggggctgag ccgttggctg 300

ccggagacca gctgaggtag gagtattaac tccctctgct gctctcgcct gccttcctcg 360

caccccctta cacagctcta cttgcagcag gctatggccc cattctttct cctatttttc 420

taactactga gatcagagct gaattaagct ggtgaaagga gcaaaacgtg caagggattg 480

attgccctcc ttgggggaaa agcggaggct taaaatcaat tcgacaaatg agtgtttact 540

gggtgctgag tactgtgctc cgctattgtg agggagggtt atgaataagg taccccctc 600

ccgccccagg gtccgttgtc agatctcaga atcagtttcc cctgcagttc tggaagccca 660

IDT Exhibit 1020.8249

aagtttcggg gttgagttgt ggtccctgat cccgatcctc aaccaatcta gctttctaaa 720

tcagaagaag gtggaattca attttccttt ctccttcctg ggatgacttt aacctgcagc 780

cgaaatggag tctataggcc ccttaaaaaa gcgcgcgcac gccagtgtgt gtgtgtgcga 840

gcgcgctcgc gtgcgcgcgt gtgttttaag agtaagtcaa attaatggtt ttagtgatgt 900

tcttatttca tgattttaat tatttaccat atctgcagta gacaccagtt tggggcagag 960

gaacccgcct ctccagactc tacaaatacc acctttttt ctaaagcttt tttccgctac 1020

cccagtcctc tgactcgagg cagaaatctt tcccctctct ttgccctctc agaattttat 1080

ttgccaatca cttgcggaac ttatatattt atagatttat ctcttcactc acatatgagt 1140

attccctgtg ctttttgttt gtttgttctc actgcaacat ccagcagtgt tttgtatcta 1200

atgggtactc aaggaaagct tatccagttg aaggtcattt tctccttctg tatgagctaa 1260

atctcagtgt ctctagaatt aaagagactc cagggatgga acttttgatt tagggtgtgg 1320

tgaagggacc cacacataca gttagactca cagccccttt actggaaagg taataaagta 1380

tttaattcat tttggtctct agacaatcaa ccttctccca ctgaccaccc acctctgttt 1440

cctgaattcc caaaagcaaa agaaaaccaa actgctaagc aactgcctag agcaagacat 1500

gtatgttcag ctgccaacac ctagagcaaa cccattccaa gtggagaatg accaaaaaat 1560

cttgattatt tcttgacctg tgtcaagtat gttgaaagcc tgccaaagtt tcctcatttc 1620

tattgaagca ctcttattct ggatgcattt tagaacagtt tgaacagtgt tacattgctc 1680

agaggtgaag aaaattgctt tgtagtttaa ggatatttaa gatttgtttg tttgtttgtt 1740

tgttttctgt cccaccttct acaaattgca cgatagatac ctcagatcag gaatgctgca 1800

tgaaaaagta tgtccataat gcaggagatt agactaaatg actcttaaga tatc 1854

<210> 19
<211> 674
<212> DNA
<213> Homo sapiens 2.E.20

<400> 19
gcggccgcct tcccttccca ttcactggct gcctcctttg tgaactaatg actgtaatta 60

ttacctccca gagctctttt gttatctcca accccaagcc ccggagaggg ggaatgggct 120

ctttagtgaa atgaaagtca ttacaaagca aattaccgtc tagggaggga cagccttcag 180

gaaagacaaa tcagatctcc atctgcatct gaagtagggt gtgtttaaat aaaaaatgta 240

aatatcacca ttagatccaa agtactccag agctgtggga tttaatggag tttaaacggt 300

agcacttgaa gccattgctt taccaaaaag aaaaaaaaat cagttaaatt caggtgtttt 360

aatccgtttc ttctttgggg gttttgtgtg atttaaacgc ttgcttttaa gaacctttat 420

gttttcaacc actcatccat agtagaaaag ttctgcaacc ctagactgct ggcttgaagg 480

aaaacctttg caggatttga tatggatttc aacaaagaac cagcctctgc gaggctggag 540

agagctgcgg agctgccatg cctgaagtgc agatggctga accacaagtc tttaggtttc 600

IDT Exhibit 1020.8250

cggagttgtt attgtggtga cctagagtgt cagagccagg agagcaagaa agaggagcca 660

aactgagccc tgag 674

<210> 20
<211> 676
<212> DNA
<213> Homo sapiens 2.E.24

<220>
<221> n
<222> (493)..(493)
<223> a or g or c or t

<220>
<221> n
<222> (505)..(505)
<223> a or g or c or t

<220>
<221> n
<222> (508)..(508)
<223> a or g or c or t

<220>
<221> n
<222> (533)..(533)
<223> a or g or c or t

<400> 20
gcggccgcag acgcgccagg cccgccaggg cgccgcacgc cgggcgcgcc acgatgtcca 60

cgaagcccac gatggacagc aggaaggcgg cgtcggtgtc gggcacgccc gcgtccttgg 120

cgtagttcac cagcaggatg gcggggacga agagcccgag cgccatcagg aacttggtga 180

cggcgtacac ggcgaaggcg cggtcggtgc acactgccaa gtccagcagg cgccggcggg 240

gccggaccct gggggatgcc tcgcgcagct gcagccccgc accgtcagcc tccgcctcgc 300

ccggagcgtc cccggcgcgg tcgccggcgc tgtccctgcg cggtcgcggg cccggcccgg 360

gcggcggcct catgacagcc ccgcaggcgc agcagtgcag caggagcccg ccgagcagca 420

ggaagccgcc gcgccagccg aagcgctcca gcagctgctg gccgagcggc gacagcgcgg 480

acaggaacac ggngctgccc gccgncgnca gcccgttggc cagaggccgc cgncgctcga 540

agtacagccc cagcatgatg agcgacggct ggaagttgag ggccaggccc aggcctgcgg 600

gcgaggcggt gctgtgccgg ggtccccgga gagcccctcc ttgggcccca caggagggag 660

gggccaggcc ccggaa 676

<210> 21
<211> 455
<212> DNA
<213> Homo sapiens 2.E.25

IDT Exhibit 1020.8251

<400> 21
gcggccgcgg ctgggggcgg ggaggggggc gcaggacccc aagtgggggt cccggagcca 60

gaggcaagtg tcctggggtg ctgggggcgc cgtgccggcc gggccgctgc cctggcctag 120

gctggtccgg gggctagcgc gccgggggct gcggccgatg ggcggggcga ggggccgcgg 180

gggtggcgag ccgggggggc acgggggtcg ggggtgcccg aggggcgcg gccgggcggg 240

ggtggccagg gatgggggtc actgggggca aaggggatcc agtgggggggg tcccgatgga 300

ggcgtgcagg gccaggggcg cccgaggcgt gcgggggtcg ggtgccccag actggtggcg 360

tcagacaggc gtgggtcgtt gggggcctgg gtcgcggctt gactgagggc ccggccgggg 420

ctgtggggcg tcaggagagc gtggggtgtt atggg 455

<210> 22
<211> 156
<212> DNA
<213> Homo sapiens 2.E.30

<400> 22
gcggccgcgc ttcgacgacg acgacgactc cttgcaggag gccgccgtag tggccgccgc 60

cagcctctcg gccgcagccg ccagcctctc tgtggctgct gcttcgggcg gcgcggggac 120

tggtgggggc ggcgctgggg gtggctgtgt ggccgg 156

<210> 23
<211> 978
<212> DNA
<213> Homo sapiens 2.E.37

<220>
<221> n
<222> (712)..(712)
<223> a or g or c or t

<220>
<221> n
<222> (819)..(819)
<223> a or g or c or t

<220>
<221> n
<222> (938)..(938)
<223> a or g or c or t

<220>
<221> n
<222> (956)..(956)
<223> a or g or c or t

<400> 23
gcggccgcta cagtgcgtca acaggcgctg taatccgagc gcataaacga ggggtccggg 60

ggtgggggcc cggggcggcc gtggcagtgg cccggggctg gcagcccgct ttgaaaatct 120

ggcgaagtcg gggagcctgc gtttgctttg gcagctgcga aggcgcacag gtgcacgggg 180

IDT Exhibit 1020.8252

gcgggggggct ggctggcggc gccaccaccg accgtcactg acagagcctc gccatgggcg 240

cccaaattcg ttcacttgcg aattgcgtaa gcggccctcc ggtacccaac ctctgggaat 300

tacgcgggct tgtgcctgtg gccaccttgc taggccccac cgctccagcc tgaactccca 360

ccgctccctg ccttgcgctt gatgttccag caacttcgaa ctgtttttat ctcctgtaaa 420

ccaagccgct tctctccttg acgctggcct tcctgcctgg cttgccctcc cgccttcttt 480

tgccttttaa gaccgggcag ctatcccacc ccgccagtat atgcccctct tctgggctcc 540

ttggcttcct gtttatacct acgtgactgt gcttactttt ttgcacatgg tttttcttat 600

ccttctgtaa gtttcttgaa ggtaggagcc atgtcttacc ctgccaagca cattgtctgg 660

cacgtagtag ctgttcagta gaggaagtgg tccctttccc taaagggctt tncgtctcac 720

tggagagaaa ggctagcctg gtaccaggga ctgccgagat caagtgatgg cagtacgtgc 780

gattcgatgg tgccgaaagt gacctagaga ggcagctgng agtgctctgg tgctcgcgga 840

tagagctttg gcgatattgt catttacaat gaggactgta ctctgagacg tggaccttct 900

aacagaccat tataaccttt gctctggagg agtgagcnag caacggactc tgacancatg 960

ttttgacaat gggtattg 978

<210> 24
<211> 321
<212> DNA
<213> Homo sapiens 2.E.4

<400> 24
gcggccgcac cggctcgggc tctgccaagg gacccggcct gccccaatgc cgccggcggg 60

cggtgcccgg tcgaccctgc acctgactgc gaggcgcggg aaatgaccgg gtctgtcagc 120

ctcccatcgc ggcttccgtc tacaggtact acctgtgctc tgtccagcct cagccactgg 180

acgatccttc ccgtagccgt aggaaggggc ggcgcttcct tggaggggat attagaggcc 240

cgaattcgcc cgggaagcgg cgggagggcg ggggtgccgg gaaggaggga ggggagaagg 300

agtgagggaa gtgggtgtat g 321

<210> 25
<211> 1023
<212> DNA
<213> Homo sapiens 2.E.40

<220>
<221> n
<222> (628)..(628)
<223> a or g or c or t

<220>
<221> n
<222> (651)..(651)
<223> a or g or c or t

IDT Exhibit 1020.8253

<220>
<221> n
<222> (830)..(830)
<223> a or g or c or t

<220>
<221> n
<222> (837)..(837)
<223> a or g or c or t

<400> 25
gcggccgcgg gctgggggcg agcgcacacc ccgcgccgct ggagttcact gccgggcgcc 60

ggcatgggcc tgggggaggg gtgcacaggg cccggagggt gcgtgggtgt ggggtgcgcc 120

cggaggagag cgaggctgcc agagtgcgtg tgccgactga gccagtgtga gtgtgcaggg 180

gctggcggag agactgggag cgagtgtgtg tgcatctaac cgggaggttg tgagtttgtg 240

tgcgcgcacg cccgcagaga agttgtgagc ctgtgtgtgc acctaacaca gaggttctaa 300

gtgtgtgcac ttgtatgtgt gtgtgcacac gcggacagag tgattgtaag gatatgtgtg 360

cacctcacag agaggttgtg agattgtaag ggtttgcgca cctaacggag atgttgtgag 420

tgctttttt cctgacaggc tgtgagtttg tgttgtgtgt attagaggtt tgtatggacc 480

tgactgaggg gttgtggaat gtgtgtgcgt gagcatgagc ctggagaggt tctatgcctg 540

ttcactcctg acagagtttg tgagtgtgta tgattgtgtg actacaccac ccaactggcg 600

gattgaatgt gttgtataca tctactgnga gggcgtgtgt gtgtgtaaat tgtatacaat 660

gaggctgtgt gcatcagtgc acctaaccac gaacctgtgt gtacagatgt gtgtgccttt 720

ctgtgtatca gacatgaggc catgtgtctg ngtgtgttta gttggttgtg caagtgctgg 780

agtctggggg ggagagaggc agttcggagc cttcccgctt tctccttctn cactctntgc 840

ttgtctcggc caccagcatg ttggaggact acaaggctgc ccttcaggcc ctttagaccc 900

gcttaaggca cttgtgatcc tatatgccag atgccctccc aaagtgccag gctaccacat 960

ggcttggctg attgattggc attgaccacc catttgttct ttgcttcctg ggcgggtcat 1020

aaa 1023

<210> 26
<211> 964
<212> DNA
<213> Homo sapiens 2.E.61

<400> 26
agccacatgt gtacccatct tcctcctctg tggaaggcgg aaggaaacag atgccctcca 60

aatatggaca gctgaaatga tgaagtgctg aagccctggc ccagaccctc agagagatgt 120

actcaaccac ctccccaccc ttggacaagc acaaaaccag agaaaacaaa ggccagcaac 180

tgtggctcag cccgcataaa tttcttctgg acactggcct gtctatttga atatctgtaa 240

tgtttggtgg agtcaggggt gagggtctca gcctttggct gctgcatctc cagacaccaa 300

IDT Exhibit 1020.8254

tcatggggtt cttttctttt ttttaatttt ttttttttt ttggaaccgg attccaaggg 360

gccaatttaa gttaacttcg gcttccaagg ttcaaggcaa ttcttctggc ttaaccttcc 420

aaagtggctg ggaataccag gattgcacma cmatgccsgg ytaatttkgw attttwagka 480

raracarggt ttytccatgt kggtwaggyt ggyctmaaac tytsgacctm aggwgatcca 540

cccgcytsgg cctccmaaag tgctggratt acagsswtga sccaccgkgc csggcccatc 600

atggtcttac taatgggtat ttcccccttà acatgtcatt tgagcccctg cctgctcatc 660

agtaaactgg gctaattaat aataccctcc tgtagggctg ttgtaagaat aaaatggact 720

atttgagaaa agggcttaac aacagggtat agtgacagag gactcggtaa ctgcttttt 780

gtgcttatta agagagaata ctacagcaac ctatgggaag atttggagtc acgaaaacct 840

gttctccgtc cttggagcca cagctggact acatttccca gccttccttg cagctgggca 900

tggtcacatg actgtgctcc agccaatgga atgtgaatgc aagtgatatc aagcttatcg 960

atac 964

<210> 27
<211> 748
<212> DNA
<213> Homo sapiens 2.E.64

<400> 27
gcggccgctc cgttgactgc agggccccgg cggtcttcct ccgctgttcc gaggccgttg 60

agggctgatg tgctccatcc tcccacttgt ggtttggcaa gccatccagc cgactacaaa 120

cccacgtttg tgagttacct gctggctgtg acgcttccgt caaatctgag taacagtttc 180

ctcatctcta agatgggtaa catagtatct acctcacagg atcgtgtggg cagtacatgc 240

atagaaagga tttaacacgc agtgtactca gctagtttta ttatttatcc gtaatgatca 300

tttgttcttt tcccctaact gtgcctcaca agcatgaaac agaatccacc aaacatttag 360

gtctgggtag tggttggatg gaaacccatc gcgggttaac gcttccaaca ccagtccctt 420

gacactctcc cgccgaggag gctgatttgt aaacttgctg agaagagaat acccagcaga 480

tctttcaggt ttcaaatcca cgttctttac aagttgtgtt aattgtttgt atatgctttc 540

gatatagagt ctctaggaag taatactagt acatgtttta aaattcaaat actgccaaac 600

agtgagatgt aagtctccct cctaacttct gtttcccaaa tcccatgtcg tttcttctga 660

tgcaatagac attgtatgtg tgtgtgtcta gatagataca tatgtgtatc tctcggcttt 720

ttttttttct tttaaagagt aaaccaag 748

<210> 28
<211> 250
<212> DNA
<213> Homo sapiens 2.F.2

<400> 28
gcggccgccg gggaagggcc ctggaagagc aggaccaggc agagcgggcg ctggggtctg 60

IDT Exhibit 1020.8255

cgctggagct tgcgctgagg ccggggtctg gccaggagcc gcagttgcag ccgctgctgc 120

cgcagggtct gaggatgagg ctggagccgc agcgggaacc ggagccgcag ccggtgctgg 180

cgttggcgct ggaactgagg ctggggccgc cgccgggact ggggttggcg tggccggagg 240

agcacttact 250

<210> 29
<211> 657
<212> DNA
<213> Homo sapiens 2.F.41

<400> 29
gcggccgcgt acggacagcc agtgcattag gcagggctcc cctacgcgcc cggagagcgc 60

ggaccgctgc ctcgggccgg cgccgcctcc tgccgcctgc cgccgctcgc ggagcccgag 120

ccccagcccg agccgccgcc taccccaggc cggggcgtcg agcagccggc ggcctgtcca 180

tgtggggcta gccctcgcgc ctggcctgca tcaggaccag caacatggag gcggccgttt 240

gcgaccccga cacgcgagga ccagggcggt gcggagcccc gcgaggacgc gacgcccatg 300

gacgcctgtc tgcggaaact gggcttgtat tggaaactgg tcgacaagga cgggtcgtgc 360

ctgtttctgg cccgggcgga gcaggtattg cactctcagt ttcgccatgt ggaagtcaga 420

atggcctgta ttcactcgct tcgagagaac agagagaaac ttgaagcgat tatagaacga 480

ccatttgaag gaattttaaa gcgcttcgga aattcacagg aatgggtatg acaaatggaa 540

aaaagagccc tttctcttat gtacaggaaa gattttattc ctaaactgga gccaaaggtt 600

ctttctcaca agtaactgaa aatattttcc tgaaaggggt tactggtgtt tttaaat 657

<210> 30
<211> 318
<212> DNA
<213> Homo sapiens 2.F.50

<220>
<221> n
<222> (39)..(39)
<223> a or g or c or t

<220>
<221> n
<222> (84)..(84)
<223> a or g or c or t

<220>
<221> n
<222> (189)..(189)
<223> a or g or c or t

<220>
<221> n
<222> (223)..(223)
<223> a or g or c or t

IDT Exhibit 1020.8256

<220>
<221> n
<222> (256)..(256)
<223> a or g or c or t

<220>
<221> n
<222> (260)..(260)
<223> a or g or c or t

<220>
<221> n
<222> (261)..(261)
<223> a or g or c or t

<220>
<221> n
<222> (279)..(279)
<223> a or g or c or t

<220>
<221> n
<222> (296)..(296)
<223> a or g or c or t

<220>
<221> n
<222> (297)..(297)
<223> a or g or c or t

<220>
<221> n
<222> (305)..(305)
<223> a or g or c or t

<220>
<221> n
<222> (317)..(317)
<223> a or g or c or t

<400> 30
gcggccgcgg agcgattgca tgcaggggcc gcgtaccgng aagtgcagaa gctgatgcac 60

cacgagtggc tgggcgcggg cgcnggccac cccgtgggcc tagcgcaccc ccagtggcta 120

cccacgggag gaggcggcgg cggcgattgg gccggcggcc cgcacctaga acacggcaag 180

gcaggcggng gcggcaccgg ccgagccgac gacggcggcg gcngcggagg tttccacgcg 240

cgcctggtgc accagngggn ntgcccacgc ggtcgcagna tgggcgcagg gcaatnncaa 300

aacancactt gggcccng 318

<210> 31
<211> 525
<212> DNA

IDT Exhibit 1020.8257

<213> Homo Sapiens 2.F.59

<400> 31
gcggccgcct cccgccagga agggtggcgg gcccggaagg ccagagatgc cccagtgctt 60

cccgcgccgc tacgcaccta gctgcccgcg ggtcccacat ggctgcggcc ggagggtccg 120

caccaggacc gccgccgcct ggggaagcgc ttccctgtgg gcagggcgcg gcgggcagtg 180

cggaagcccg aaagctaccg gagcccgggg caggggcggc gcgatgcaga ggcggcgttc 240

gggggccccc agctgcctgc ggctcggcta cccagccgcg atcagagggg gcgggggacg 300

caggaacccc ggcgtccggg cggtgtgcag ccgcagacct attccaagtt tccacgtagt 360

tgcgagagcc caaaaactgt cacgtgcacg tcgctgctga gtgggaggag gtgtttgtca 420

tcgcgttcaa aaggggcgtt tcggtgtctc ccgtcatgca agcaaatggt atggctctcg 480

gccgcctttg aataaacgag tgcttcgaac cctttaccag gaggg 525

<210> 32
<211> 1032
<212> DNA
<213> Homo sapiens 2.F.70

<220>
<221> n
<222> (687)..(687)
<223> a or g or c or t

<220>
<221> n
<222> (689)..(689)
<223> a or g or c or t

<220>
<221> n
<222> (885)..(885)
<223> a or g or c or t

<400> 32
gcggccgcgg ccgggggggct gagaagggcc tgggtgcctg tcgcccggga gccgaggttt 60

cccggcctcc cccgaccccg ggcgccaaga gcagtcggtc cccccggcct cccgccggca 120

aaggggcccc ggggcccagg cgcgcggccc ctgcgtggcg gcaggcggcc caggccagcg 180

ccggcggcta gagaaggcct ccagtccagg cctcatggaa gggcctgcct cgagcggccc 240

ctcaacgccc cgcagtgtgg cactggaagg gacctaaaaa cccacctggc tttctccttt 300

ccccttcccc acgcttccca gggcccaatg cccgcatctc agtttcgctt tccggcaggg 360

tcaggggtga gagggaggaa ttctcaggtg tcacctcctc acccgcctgg aggcggaggc 420

tagaaagacg tcggggcact ctggagggga ggaagaggtg tgcctagaat tctctctctt 480

aaacgctcgc gttatcacgg aggagacttt ataaacactt taaacacaac accaaccatt 540

ttatcagcaa aagcgagggg agggggcgt acagtaaatg ctgagagatg ttcgagaagc 600

IDT Exhibit 1020.8258

cccaagacgt tccctgcgga aggagaacgg aagaaagaaa ttacgggcgg aaaaagagta 660

aatattagct ccacacctaa ccacttncnc agccccaaac taggagagaa tctgctaaga 720

ttcgctttat atttatatag tctatgtgat gttaacaata ggggttgcaa atattgcatg 780

ggggcattct tagagtaaaa aattggtatc tacctgaaat tcaaaaattt aactgggcat 840

cctgtatttt tattggctaa tcctgcaatt ctaactaaaa aacancttgt gaagaaatca 900

tatagaagga agctaattgc tgatgaatac agtattggga actgttatgg aactggctgg 960

aaagaaatga ttctctacga tactttgagc catgtaggtg agagagatga tgagcactgg 1020

atgtctacta tt 1032

<210> 33
<211> 708
<212> DNA
<213> Homo sapiens 2.G.10

<400> 33
gcggccgcgc ccaggcgccc cttcccctgt ggggcaaccc aagccgggga cgcgtgaacc 60

acctccgtag ccgccccgcc agcaccccca gccgtgcgcc cctgcaccac gcagctgccc 120

tgcgcatgga gcccagaggg acagcaggcc cggcccccag caccaccggc ctgccgggag 180

gttcgggaaa ctggcgtcgc agcggagagg gcatcaggcc aacgcctccc ccgaggctca 240

gctgcgggct cccaggcgta ggcacccacg gcccttacgc tgaccgtagc ttggacgccg 300

ctgccgccgg ggtccaatgc cggtcatgcc catcccgcgg gggttgtgct ccttccatgg 360

tccacacacc acctgcctgc atgcggtctg tgggcccgtg ggcgcctccc acctggcccg 420

caccaagtac aacagcttcg aggtgtgcat caagacgcgc tggctgtagg gcttcatcca 480

cttcctgctc tacttcagct gcagcctgtc actggggcac gctggccgcc ttcttctgcc 540

tgcagtactt gggcgttagc gtcctcctgt gcttccaaca caagctgtgg gtgctgctgc 600

tgctgcttgg cccgctggcg cgttgaaatt tcgctgttga acgagctgct catctacagc 660

atccacgtca acatgcttgt tgtatggggg cctgggctgg atgcctaa 708

<210> 34
<211> 569
<212> DNA
<213> Homo sapiens 2.G.108

<400> 34
gcggccgcac acgtgtccag gcgtcacgtc cgcgcgcgcc cccggggctt gcgtcagcgg 60

ctgttccaga agcgggtggg ccagggctct gcgcaccgct ggggttcggg gcccgggacg 120

ccgccgggag gagggcaccg cgcggggtcc gacgcggagg cgtgctcgga acgccggggg 180

ctgcggagtg catcagcgcg gtccagccct ccgcctgccg ggcgccgagc gtctccgccg 240

cccggacctg ggctgggcgc cgtggcgttg cctcggagct cgctgcccgc ggggcgcgca 300

ccgccttgac ccgggcggcc ccgcggcagg caggcgcccg cagttccatg gttggttcgg 360

IDT Exhibit 1020.8259

agcgcgatga gccgcccgtc ctccaccggc cccagcgcta ataaaccctg cagcaagcag 420

ccgccgccgc agccccagca cactccgtcc ccggctgcgc ccccggccgc cgccaccatc 480

tcggctgcgg gccccggctc gtccgcggtg cccgccgcgg cggcggtgat ctcgggcccc 540

ggcggcggcg gcgggccggc ccggtgtcc 569

<210> 35
<211> 916
<212> DNA
<213> Homo sapiens 3.B.30

<400> 35
gcggccgcgc tgagctcact ccgggccctg cggaaagaat tcgtaccgtt cctgttgaac 60

ttcctgaggg agcagagcag ccgcgtcctc ccgcaggggc ccccgacccc cgccaagacc 120

ccgggcgcct cggcagcctt gccagggagg ccgggaggcc cgccgcgggg tagccgcggg 180

gcgcgcagcc agcttttccc tccgaccgag gccctgagca ccgctgccga ggcccctctg 240

gcccgccgcg ggggcaggag gcgggggccg gggccggccc gcgagcgtgg aggccgcggc 300

ctggaggagg gggtcagcgg ggagagcctg cccggagccg ggggccggag gcttaggggc 360

tctggcagcc ctagccgccc cagcctcacg ctgtctgatc cgccaaacct cagcaacctg 420

gaggagttcc ctcccgtagg ctcggttccc cccggcccta cagggtgaga ctcagctctc 480

atgcaggaga tgggtaccac gaaggctctg gggagtcagt cattcgagct cggcgctccg 540

cagtggagcg ccaggatggg tagaaggctg ggggtgatgg tgagggtttt tgtggggttt 600

cttcgcagcg gccatgctct gccccgtggg ccgtcatttt gtcgtttcgt tttctctata 660

atgtaataac taactaggca aaaagtgtta aaattaataa ctactaaata tccgatgtca 720

ttacaacatt tataatatat aacaatatta aaacatataa ttaataataa aaaaaacctt 780

attttaatct ttttcttttt gttaatttat atcaccttat ataccatttt tctcaatacc 840

attcgataca atcataaatt tatttattgt atattgtcaa aataaaatat tcctctatat 900

aaaaataact ctccta 916

<210> 36
<211> 998
<212> DNA
<213> Homo sapiens 3.B.36

<400> 36
gcggccgcag catggctttc ggccactact cggagcactg gaaggtgcag cggcgcgcag 60

cccacagcat gatgcgcaac ttcttcacgc gccagccgcg cagccgccaa gtcctcgagg 120

gccacgtgct gagcgaggcg cgcgagctgg tggcgctgct ggtgcgcggc agcgcggacg 180

gcgccttcct cgacccgagg ccgctgaccg tcgtggccgt ggccaacgtc atgagtgccg 240

tgtgtttcgg ctgccgctac agccacgacg accccgagtt ccgtgagctg ctcagccaca 300

acgaagagtt cgggcgcacg gtgggcgcgg gcagcctggt ggacgtgatg ccctggctgc 360

IDT Exhibit 1020.8260

agtacttccc caacccggtg cgcaccgttt tccgcgaatt cgagcagctc aaccgcaact 420

tcagcaactt catcctggac aagttcttga ggcactgcga aagccttcgg cccggggccg 480

ccccccgcga catgatggac gcctttatcc tctctgcgga aaagaaggcg gccggggact 540

cgcacggtgg tggcgcgcgg ctggatttgg agaacgtacc ggccactatc actgacatct 600

tcggcgccag ccaggacacc ctgtccaccg cgctgcagtg gctgctccct ctctttcacc 660

aggtaaagcg ctctgggagg cgtgggccag gtcttttctc ctctgaaaar ggcggagtag 720

agacagaata tgctgagttt gcaagcaggg ccccsggttt ggggtttcgc tccaggtccc 780

caccctcaa aaccaagaat cgcgtcggta arggractca cagtgagggc tgcgacacgc 840

gcacgcgccc cacccagcgg tgccccgaac ccctccggtc yyctatctkg yytctatcgt 900

cccctcmcyt gcttkcgagt gagaacacat ttgcaaagac ccctccaccc cccggaaaaa 960

caagagtttt taaatgcttg gagatgagcc ctgatatc 998

<210> 37
<211> 514
<212> DNA
<213> Homo sapiens 3.B.55

<400> 37
gcggccgcgg cgctgttggg ccagcagggc agcaccgagc ccgacttggt gccgcagtac 60

tgcgggggac tgcgggcgcc ccagcccgac gggtcggcgt agtagccgag cgggcggcca 120

gtgcagcctg cagcctgcag cggcagcgcc ttcacgcccg ccgccgcgta agagagcagc 180

gtggccgcgt tgcccgcgaa gtccgtggcc gtgtcatagg ccgaggccgc gaagtccagc 240

cggttgttgg ccggcgtcac aaaccagcgt tgcggcgagg gcgcgcccgg gtcctcggcc 300

tgctgcggcg acagcagccc gttggtgtgc ggcacgctgc ggtccgtacc cggcccgggg 360

cccgcgcccg cgcccgggtg gaagcgggcc ttggcgtagt tgctcacgaa ctggtcctgc 420

aggaaagagc cggccatggc gtagcgggcc ccgggcacga tctgcgagcg cggcgagtcg 480

ttgggcgagg gggtcaggcg gtccatgtca cagc 514

<210> 38
<211> 608
<212> DNA
<213> Homo sapiens 3.C.01

<400> 38
gcggccgcgg cgcagcggag gggctgcggg cccggaaccc aggccggtca gcgtgtaagc 60

gccccagccg gccgggctcc gtgggggggtc agctccctga cccctacagc gcggtagcgc 120

ctctccgaga gctccgggac cagcggcccg gccgccccca aagccagcct ccctctccct 180

tccccgcacc gggatcccag accagggagg gggcgcacgt ccgacggctg aggaatagca 240

gggcgcgagc cggcccggca ggtgcccatc gtcgccctct gggaccccgg tggcgcgctc 300

tgtcctccgc gccacgctca gccaccaccc cggctgtttg ggacccggca cccagccgag 360

IDT Exhibit 1020.8261

cgcgccgccc cctcggggac ccgctgggcg gggctgagcg aggcttggag tgcgggcgaa 420

gggacgtggg gcgaacccgg ggcgctgcgc cacctcggct gtctccagcg gagaccggcg 480

ccctcgcccc ccgtctccgt tcattgtgct gtattcatcc agcagatttt gaaacaattc 540

tcgtgtaaaa aggcatttta ctccgcgcgt cttccttaca gccatttagt tgggagtttg 600

cggtgggc 608

<210> 39
<211> 1025
<212> DNA
<213> Homo sapiens 3.C.16

<400> 39
gatatcctcg ctgggcgccg ggggctgcag ctcgctctgc tgctgctgct ggtagaagtt 60

ctcctcctcg tcgcagtaga aatmcgsctg caccgagtcg tagtcgaggt catagttcct 120

gttggtgaag ctaacgttga ggggcatcgt cgcgggaggc tgctggagcg gggcacacaa 180

agcgggaggc agtcttgagt taaaggggtc ttggtgcgra aacctggcgc agcgcgcagt 240

gcgcgccaca gtcccgaacc tctccccttg cagagctatc ccctaaagcg gctgggtggt 300

cttggtgggg gaataaaggg agcacccttt cacccccttt ggacagtccc ctgctatctc 360

ggagacgcac ttagtgaacc agcggcttgg tgcccgccga gcccccgctc ccccgggagc 420

ccggagcgca aagcccggga gtcggccccg cagcggcaga ggaatcgaaa tcggccctgg 480

cgcccttaag aagccgcggg aggtggcggt gaggaaaaca atttgccaaa atccaaggca 540

caaagttttg cgccacctga aggagaaggc gagaggcgcc tgggcgctag cggctgcgtg 600

aaccccgctc cgcgccgggg cccctccgct gcggctgttc ccactcgcgc cctagccgct 660

ctcctacccc cgccggcacc gcagcccctc ccaaccttcc ytytccaccg scccgtccc 720

caccccagt accgcccccg tccaacactc cttttgccag cttttcttct ttctctcgcc 780

ggctggagtg gcgagctcag ccgcgggctt taacacccct ccataaatac arggggggtg 840

tcaaataata ataggggcac ctcccttcgc actcaatacg gagatgcaac tgcgccagag 900

accccgctgc gatacctccc ccggagccac cccaccaagg gtagcagctg ttctggaacc 960

gcccagagcc ccgctcctcg cagttcctyc gcatctcggg cgcgaggaca cccgagggcg 1020

gccgc 1025

<210> 40
<211> 1010
<212> DNA
<213> Homo sapiens 3.C.17

<220>
<221> n
<222> (829)..(829)
<223> a or g or c or t

<220>

IDT Exhibit 1020.8262

<221> n
<222> (875)..(875)
<223> a or g or c or t

<220>
<221> n
<222> (906)..(906)
<223> a or g or c or t

<400> 40
gcggccgcgg accgacttcc ttcgccggcc accggaggga gggggcgccc ctaccccggg 60

aggggggctgg gcgagccggg agacggtcaa gttggggtcg ggggagcgcg ggcgctccgc 120

actctggggc acgcggggac gagcccggcc gcattgtctg cgcggcctcg gaacaagcac 180

ggccggcggt ggcaccggcg ggcgcgggga ggagttgccg tcccctttcg ccgccgccgc 240

ccaccgcgtt ctttgtgtgt ctctcgccgc cctccagccg cttcgccgct cgcctgacag 300

ctgatgggct caccgcgccg ggtcccgcgt cctctcggcc gcagccggcg gagcccggcc 360

cggcaggagg aggaggggag aagaggagcg ttgacagatg ctgtcttgga gcgggcaccg 420

ccgggggaaa agtctggact gcctcggcga gaagcggccg gtaggcaacc ggccccagcc 480

tcgcattcgc ctcaaagacc ccaattggct aggagccctt ccctccgcag cggctcgcgc 540

agctccgctc ttgcgccccg cgcccggctc agcggacgga ctagcgcgcc cggtcaagaa 600

tcctggggaa cccgctccgc cccctggctc cagcgccctc caatggatgt cggcgtacag 660

aggggctgtt ccgcccaatc aggtgtcggg aagcccagcc agtccccggg gagtgtagcc 720

aatagaaggc gacttcggca cacacccgcc ctgatccact aggacaaacc gctcgagccg 780

gggtggtgga ccgatcctga ggcagatcag ccagtccgcc aaactgtgng caagtagatc 840

tgagacggtc cgtgttaatg actatatcta agagntggat gggaacgggg cgcccaattt 900

tccctngtat acgcttttgg caagttgggt tgaaaactga caacctgagc tgttaatgag 960

gcttctttaa ctgtttatgc tatacgccta gtggctcaga caacgttttt 1010

<210> 41
<211> 413
<212> DNA
<213> Homo sapiens 3.C.30

<400> 41
gcggccgccg taaagcgcgg atgcgcggcg tggccacgcc ccttcagtgc ttgtgacgca 60

ggcgccctgg gctttttggg cgcgaaaaag aagcagtcct gggttgtacc cggcgcagct 120

gggagcggct gcttcctccg gggtcgtatc tccgcccggc atggggctgc tggacctttg 180

cgaggaagtg ttcggcaccg ccgaccttta ccgggtgctg ggcgtgcgac gcgaggcctc 240

cgacggcgag gtccgacgag gctaccacaa ggtgtccctg caggtacacc cggaccgggt 300

gggtgagggc gacaaggagg acgccacccg ccgcttccag gtatgcaggg acccgccccg 360

aagacgaccg gctgcgcggg cctcccccta gacttttggc taccgggccc cgc 413

IDT Exhibit 1020.8263

<210> 42
<211> 927
<212> DNA
<213> Homo sapiens 3.C.35

<220>
<221> n
<222> (595)..(595)
<223> a or g or c or t

<400> 42
gcggccgccg ctccttgcct gaccgcttgc tccccgcccg cccgcccgcc gggttgtcgg 60

cgcggggcca ctggcgggtc gtgatgagca ctcgctcgcg cccccgcacg cacacgcgaa 120

acccggcccg gcccgccgcg ccgccccgcc tctcgcactc ccggagctcg cccaccggcc 180

gcgctggctc acactctccc tcacagcacg ccggccgagg gaggaagggg gcggtccggg 240

ctcccgaggc gtggggaggg ctgtttattt tgggggggagg aggggcgcga ggcaggaacg 300

agctgactgg ccgggatcct ccgacccgcc actgtggcag caccgggaag gcggggagag 360

agaaagaggg agggagggag ggaccgggat gtagaactcc agcccgcgcg ggaggctacg 420

gcgaggggggg cggtggcggc ccgcgggggg ggcggtgcca ggccccctcg gcaatctccg 480

tagtctcctc gctggctgcc cgagggaggc cgggaagcga tcggggaagc tcgggaatct 540

ccggcacggg cctgggattg tcctggaggc acagcgcggc tggagtgcgg ggcancgcgg 600

gggggcgggg gtctgtctcc tttctgggcg gggccgtatc ctgaagcagg cggggcttga 660

gagacccgaa agccacggag tggctcctgc ttgcggtact agttggacag agtaaagtcc 720

tggagttacc tcgcctgagc accctggttt cccgagaggg aatgggcact ctgtgagagg 780

caagctattt gcctgctttc cctccgcaga agaaaaaagg ctcaattgga aggtggagga 840

tgaagccacc ctctatggtc accccaatct gagagcttta ctttatataa ctacattcta 900

aggagtagta aaatacccga ggtggaa 927

<210> 43
<211> 1365
<212> DNA
<213> Homo sapiens 3.C.64

<400> 43
gcggccgcaa ggaccggctg agarmtgkgg gscsctgtgc tgggggcgsg arggagrcgg 60

ccytraggac tgcscscccc ccacaccggg gcccgggcgg gacacacgcc caacgggacc 120

cctgagcccc caggctgggg accggcaggg gctccgggga ggctggtgag gccaggacgg 180

agccgccycc acgcgtagcc gtgaagcggg aggtacgcgg ccccctggag ctgccccgac 240

tgcagccgag ggcgcgacct gtggtgccaa ccgcctgacc ctgcttggcc gccgccgcct 300

gcgggtctcc agcaggtccc accccacgcg cccgcggggc ccgctccaga ggctcctcca 360

aggccgctgc agaggcgcgg ccaggctccc atttctgcgc atccctggcg ctcagacacg 420

IDT Exhibit 1020.8264

gcctgagccg ggtacccggc gactcccttc ggcctccacc gcctcctggg gagggaccgc 480

gcgctgctcc cacgcgggcc cgggggtctc cgcagccctg gcctgggtgc gtccgtcggg 540

ctgctcggct cggagcaccc cccgccccgc cgccccacca gcgcctytyc ggagcgctca 600

ccccgccccc gactcgtgtt gttgttgcgt gggtttttc tctaattctc cggagttact 660

cttttgttgc caattgtttc tatgcccgga ggccacgctg taaatgagat gttacatctg 720

caccgagcta agtaaacact ttataaatga ataaataagt gaataaataa cgaaatcgtc 780

atctcggggc ggcccggctg ccagggctcc ggccgccggc ctgcgggggt ctgtgtggtc 840

ccgggccctg ccctggggtc ggggaggcgc cgggaggggc cgtttcccag ccgtgtccct 900

accctgaccc catcttcctt cctctcccaa atcatcctcc agactctggg cgtttggtcc 960

ccagatgtcg tgtgggattc gtggcttcca cccaccgctt ctcaaacaaa aacgggttgt 1020

caccgcggct cttaaccctg ggcgagccac ggagcgtttc ttcccgggat cgggatcggg 1080

ccgcggctcg aaccggcatc tgcagaagga agacccggcc ctgtaggccg ccgccgcccc 1140

aggaccggac tggtggcctc tccacgtcgt gtccggaccc gacwcatcgc ctccaacgcs 1200

aacaaacgga agcagcggag cctccgcctc cmasscykgc cyctgyscgs yswgmcmggc 1260

gcattsragt gcwcsakkym sgcycaatym mgagagckct gracktckca aytatcwcgg 1320

actarsrrsr rcawwtkmww argsactcay tgagtaactg atatc 1365

<210> 44
<211> 608
<212> DNA
<213> Homo sapiens 3.D.21

<400> 44
gcggccgcac tgccctggcc gccacgctcc gcgcctgcgc cgcgcacctc aggggcccgc 60

cgagagggcg gggaggtgac gaggtgaggt gggggcaggg agcgggctgc gcgaacgcac 120

cgcacacgcg gcctgggagg gaccaccggc ccgcagcccc gggggaggcc cagcggcccg 180

cgccccctgc cggaggcctt gcgccgccgc agtctccctc tgggccggga agagcccctc 240

ccgagccccg agggcgatcc caccctctag gattactcca cgccaggcgg ccagcgaatt 300

tatcccgccc gcctccaccg ccccttcaag ccctggggaa ctgggagaaa cgtggcgcgc 360

agcggcacct ttcccacgct gctcctcaag ggaaaggacg cgagtggtct tgcccaggtt 420

aggcaaggca gatggcatct cagaccccga agtgtgccag ccgcctgttg gggacagaga 480

ggccgaggac ctcgtcacgg ttttactgag gccacaccag agaaccacct agggctagga 540

tgctgccctc agggcaagag ggtgaaacct gaagactgcg agtcgttgtt gagtttcacc 600

cgattcct 608

<210> 45
<211> 1947
<212> DNA

IDT Exhibit 1020.8265

<213> Homo sapiens 3.D.24

<220>
<221> n
<222> (748)..(748)
<223> a or g or c or t

<400> 45
gatatcatct attttaaaag acatatgtaa aacccaaccc ttaagaaagg attcctatca 60

ctgttcccca caggcatcct cctcagtctt acacctttcc acccccaaa acaaatcatt 120

cagcatattt atttcatact gtaatatagg aaatagctat tttttagact ttttatatta 180

ttagcactga tcatacaaac atggaataga aattccttat gttttatctg gatttaaggt 240

gatacataat ggaatatatt tctatcaagc cgtacacatt agagataatg aaatcacttg 300

tgttctagtt taaacattat gggaatttca gaactgcaac ataacaaata atcctcggat 360

gaaaactaaa tctctcctct ggtcaggcat ctatgtgcat cagwgatgag aagacaggga 420

ctgtggaagg gaaaacagcg agtcaggaag gactgtggcc acgtccattc cctggtccct 480

caagtaatta aatcctgacc tcctctaccc cagtctgtcc tggggaatgg ccaacactgg 540

cctttcacaa ctgtgtgtta ctagaaatgc aacagaaacc cagctgaatc ccctcctctg 600

cccttctcaa aggaaagatc tgtcccagga ccatttgttc caacattttc aattatgaga 660

actgggaaga taaagttatt tttacattta taaagaaaca catatttatt cacmctcatt 720

wcaagraagg tcaagaatct atmcaaanac caagaggaat ttttaaaatc ccataatwcc 780

accatcaaaa gagccacact tagcatgttg gtccacaggc ttctttagca ccctcttyyg 840

ttggtgtatg cacaaaatgc acaatcacat tctgtctaca ttttataatt tgcctgtttg 900

ttgattamca ctatatattg aacaattttt aagacctgca acatatgttg acaacattac 960

ttccaaacaa tgtatttaca aataaatgca cacacacact atctgtctta tatacaacgt 1020

gtcttacttt ctaattctcc actcttgaag atttaggttt ttccaacttt ttcttaatat 1080

attcaccagg agtcagcaac ttttttccat aaaaggccaa agagtaggcc gggcgcagtg 1140

gctcacgcct gtaatcccag cattttggga ggccaaggcg ggcagatcac gaggtcagga 1200

gatccagacc atgctggcta acacggtgaa accctgtctc tactaaaaac acaaaaaatt 1260

agctgggtgt ggtgagtgtg gcggcggaca cctgtagtcc cagctactcg ggaggctgag 1320

gcaggagaat ggcgtgaacc cgggaggcag aggttgcagt gagccaagat cgcaccactg 1380

cactccagcc tgggcgacag agcaagactc tgtcacaaaa gmaaagaaaa aaaaaaggcc 1440

aaagagtaga tattttaaac tctgcaggcc ataggtttct gttgcaacac tcaactctgc 1500

tgttgcaggg aaagaagcca tacacaattt gtaaatgaat gggcatgact gtgttcttcc 1560

cgacatggtt tgccagcccc tgatgtataa cactacagag gatgctgtta gaatgaaawt 1620

tctttacata tctctgatga tctccttagg actaattact agacatgaca tcatggtagc 1680

tgtgggtcaa agggcatgca tgctctggga tgtacattcc cagattgctc atcatgagcc 1740

IDT Exhibit 1020.8266

tttctcatgt caaaatgttt tgtgaccacc agaaaggctg gttctgcttt tawtacccat 1800

ggawtgagga atagaaatga catggcatgg cccttcccca cagcaccacg gcttctcttc 1860

ctcagcacgg cgacaggggc ttccccttg ccgccgccgc ccgccaagct ccgccgccgc 1920

cggccaagct ccgccgcgcc cgcggcc 1947

<210> 46
<211> 1637
<212> DNA
<213> Homo sapiens 3.D.35

<220>
<221> n
<222> (612)..(612)
<223> a or g or c or t

<220>
<221> n
<222> (715)..(715)
<223> a or g or c or t

<220>
<221> n
<222> (1014)..(1014)
<223> a or g or c or t

<220>
<221> n
<222> (1265)..(1265)
<223> a or g or c or t

<400> 46
gatatcttct gataaagaac caatctgcct gggagtttca aatctgaaaa agcaaatcat 60

agtttactgg agtaaactgc tgtttaaaaa taaaagagaa aggaaaaaaa aaagaatgtt 120

tcctagttcc agaactgaca actagagcct aaataaatac ctggacaagg gtaaatatga 180

cctcaaattt ataaccgccc tgaacgcaga acatcaaccg cgacagctgt ggcatcagcg 240

gcgacagtaa ttttctccct ggcattcaac cagagggcag ttggactgtg caccgactgc 300

actagtggtg ggtagccaaa gctagcctcc aaagtgaacc acggtctggg gcctggtccc 360

gtttgaccga aaatgctatc cagaacmccc wycgagactg caggcccttc ttcctgattg 420

agctagaggt gagtgaagac agggtctggg gtagggaggg gcgtccacgc cagcttgccc 480

attacctgcc ggccttggtg atgatcatct cagtgcctat ctcatgaaag cgcttccaga 540

gctcggctcc ctgcagatcc acccgcgggg cctgcggcga gggcagaggg gtcccgggcc 600

gggccaggga gncgcgccgg agacccctg ggggaagcct cccggtgacg ccagagggga 660

agctccytgc tggaagccgt cctcacagcc gcctggacag caaaggacag agaanaggra 720

actggtgagg gaaaacagag gggaagcmag ccgcggagac ggscccacct ggtggctgag 780

IDT Exhibit 1020.8267

aagargaaaa tgaccgggag aaaaggggaa gctttggtgc catcaggtcc tcctaaagaa 840

caagccagtc gatagacacc cacattctgc ctgtcgaagg ggcgcattca gagctccagt 900

gtggcctgct tggtccccaa gtcccaagcc cggrakaggc gygcggsmag cgtccacmcc 960

accccgctgk gcctccgcag gkcsarggcm cmasmaraaa aggcttcacg ccgnccgccg 1020

gggtctggga cgcttgcccg acggagtcag agragctccc sggtcmagag tccacagtgc 1080

aaactycgac gcaacctgcg ccttgaarcg caagcagcaa aagcgcccsg cactctgktc 1140

ccaagagcyt gggcctcctt aagccataag cgtytgcggc gcctcgcttt kggccttctt 1200

ttgggccggg ccggaggmat cttctagaar gctcttyaga acmccgcttt ygycaaactm 1260

ycggncgccc tgcgcttcca rcccarcaga agaaaagtgt gaaaagcaag cccgcggtcg 1320

ccgtcggcct tggcagagaa atcaagagga gaagggaagg gaaccgctca actacccttc 1380

gggaaaccaa gtttccaaat atgccgccct cttcctggtt tgcacaaacg gtttagggca 1440

ttcgttccgg tttcagggtg gggtatgccg tcgctcccct cctccccgcc ctgtgctttt 1500

aaaagttagg aaacaaaaaa gagcacccat tggctggaac cccaagggag gcagatgcag 1560

gaagcacaga gctgcaccgc taggcgcagc aaacagccgc ggccgaaggc gcgggtcgcc 1620

gagtgggcgg cggccgc 1637

<210> 47
<211> 900
<212> DNA
<213> Homo sapiens 3.D.40

<220>
<221> n
<222> (671)..(671)
<223> a or g or c or t

<220>
<221> n
<222> (702)..(702)
<223> a or g or c or t

<220>
<221> n
<222> (739)..(739)
<223> a or g or c or t

<220>
<221> n
<222> (746)..(746)
<223> a or g or c or t

<220>
<221> n
<222> (752)..(752)
<223> a or g or c or t

IDT Exhibit 1020.8268

<220>
<221> n
<222> (762)..(762)
<223> a or g or c or t

<220>
<221> n
<222> (782)..(782)
<223> a or g or c or t

<220>
<221> n
<222> (816)..(816)
<223> a or g or c or t

<220>
<221> n
<222> (837)..(837)
<223> a or g or c or t

<220>
<221> n
<222> (867)..(867)
<223> a or g or c or t

<220>
<221> n
<222> (874)..(874)
<223> a or g or c or t

<220>
<221> n
<222> (879)..(879)
<223> a or g or c or t

<400> 47
gcggccgcgg cccggaccag ccgctcccac ccgccccagc tactacggcg cggcgcgacc 60

gcgggctccg gccccagccc aggcacgtgc gcccaggccg cggggaggcg ccggcgcctc 120

ccggaacgcg ctcctggcct gcgagtgctg cccgctcagt ctccgggtgg gaagtgcgct 180

cgccccggac cgaggggaaa gcccaacatc cccgggatgg aacagagagg cggccacccg 240

tgagtgggcg tgacccattg gttcccttgc gcagcatctg tggagaatta ggctttcccc 300

tcctctcttg ccagccgttg ttcctaatct tgtctttttt aagggaggaa agcaggagaa 360

ctcatgacac tttgtatcac aggaaatcaa gttggtggag agagggtttg ctgacctctc 420

ccgtcccttc tcagggtccc taggagaatt tttgaagaag taatcggcag caaggagatg 480

ggggcaatag agagtctcag actcgcaggg acccatgttc gtccccagcg ccactacttt 540

caaaccgtta tccctcagag ctgtttcctc acctccacaa caactctccc gggttcgatg 600

acactatata tcccaccagt tcatcttggt acaggccaaa aggtaattca aaaagcgaaa 660

cgaatctcat nttctgacct gtgccctcgg taaagtcccc angtttccac cccaagtaca 720

IDT Exhibit 1020.8269

cttggaagcc aggcccctnc acacangctg ancaccacct tncacaaact gaaaacaaag 780

anaatccctt ggtttcaaag ttagaatagg gatacngcgt gagtggggtg aattgcnatt 840

gggtcaagga aaaaaaaaaa gtaaatnaat taantttttnt tgacctcctg cgctgcccac 900

<210> 48
<211> 1511
<212> DNA
<213> Homo sapiens 3.D.44

<400> 48
cgggcgcggc gagccccact ttctcccggc aggaaggggg gaggccgaga gcatttcctg 60

ttgtgcagct gagccctgcg gagacgtcat tgcattcatg ctccctcggg tgtcagcgga 120

cgggggccca aagttcaagc cgcgtccagg gcaggcagcg cgcggcggcg cggcggcgcg 180

gggcgggcgg ccagggctcc cctctcccgc tggcgctccc ggcgcctccg tccccggccg 240

gcccagcgct gctaccggag gccagccctg gggctccgcg gggaagagct gctcttcctc 300

ccggaggaaa ccgagctcgc aagcccagcg ctcccagccg cagactgcag agctccagta 360

aggtgaaagt aggcaagaag gccccctgag acgtttctaa aagcatattc tatatgtttt 420

cattatgaaa acacccactg cactcctttt atttattagg accttaagtt atcctatctc 480

aactaatact tttaacaatc agaatctctt aagaatcttt caatcttata cttatccact 540

ttaatagcca acaaaacctt tagccagagt gttttaaaat ggaaattacc tgttcatgtt 600

tcttaaagat ttttaaagtc tccttctaaa tttccagcct tccatttagt ttcaagccat 660

aaaccagatt ataacaatgt gtaattgtag agaagctgtg gcttacggtt aataacgatt 720

aaaaataagg ccataaggta ttttatgatc attttgaaat aaaaaattga aatagtttaa 780

tttcagcttg tgcagtttga gacagatcgt caactacaaa acaaattgta gattctgttc 840

tcatggtgaa caaacattac agatgtttta ctgtgtcaac atctctaaca tttgaactaa 900

gcaatgtttc acatcagaac atgaattaaa acaatgtaaa ctatggacct ggggtgacca 960

tgatgtgtcg atgtaggttc ttggattata acaaatgtac cactctagcg caagacttcg 1020

atagtggagg aggctgtgtg tatgtgggga caggaagtac atgggaaatc tctgtacctt 1080

ccgctggatt ttgctgagaa gctaaaacta ccctaaaaat ataaactcta tttttaaaca 1140

tatgtttagg gttttatgag tatcctgata cttaaaatgt gcattgcatt gtaacctatg 1200

aattgacaag aaattaatct taagaattgg cacagaaatc atctcgatgt tttcatgaag 1260

ttcatcctcg gttctactgc ttcttgataa acaagtttca tgtttagaag gttactgaaa 1320

tttttttata tggtaaaggc acatcaaaga ctttaccatt taatatatat tagttgtcct 1380

atccagtcat gtactattta aggcaatatt aaaggtaact tagatttccc cacttacagt 1440

gatgcaaagc ccttcaataa tattctgttg tcttatttcc taaacatctg aataatacaa 1500

ctttatcaca t 1511

IDT Exhibit 1020.8270

<210> 49
<211> 835
<212> DNA
<213> Homo sapiens 3.D.60

<220>
<221> n
<222> (607)..(607)
<223> a or g or c or t

<220>
<221> n
<222> (675)..(675)
<223> a or g or c or t

<220>
<221> n
<222> (716)..(716)
<223> a or g or c or t

<220>
<221> n
<222> (731)..(731)
<223> a or g or c or t

<220>
<221> n
<222> (732)..(732)
<223> a or g or c or t

<220>
<221> n
<222> (741)..(741)
<223> a or g or c or t

<220>
<221> n
<222> (778)..(778)
<223> a or g or c or t

<220>
<221> n
<222> (805)..(805)
<223> a or g or c or t

<220>
<221> n
<222> (806)..(806)
<223> a or g or c or t

<400> 49
gcggccgccg cggagccggc gtccgcagcg gctgcgcatc tcgggcctgc agcggggcgc 60

ttggcgggcg ggggccgggg gagagcctgt ttgcgcagta cccccggagg gcggaaggcc 120

gccgaggtaa gagccgggac tcggccaggt gggagtgggc accttgggcc gggcctgcag 180

IDT Exhibit 1020.8271

ggcggtcccc gagcgtcccg gggtagggtg ggctccctgg ggacgatgcc cagggccccg 240

gccgcgctcc ggtcgcgccc caccccggct gcagcgcggc cttggggcgc tgctggcctc 300

gccgcggggg tgggagcggt cgcggcctgg agcagctccg ggcgggcccc aggctctggg 360

gccagggcca gctgcgcgca ggggtgagtg agcagccccc gggccctcaa gtgagccccc 420

gtccgctccc caccttgcat ttctcctctc cgcagtgggc gtggcgcccc tttgctgtat 480

agggggcgcc ccaaattgaa gaaggctggg ggggagaacg cataaacagg tgtttagggg 540

gcccaggcct gtgcgccaag ggttgaagaa taaagagtaa ttctttttttc ccccttttta 600

agggggnccg gagtcccccc ccccccggc cgtggtaagg gccccccctt gctccgtaag 660

gggccctcct ttggnaaaac aactcctttt ttctttttttt attttgtccc cccccnccca 720

ataatttaaa nncctccctg ntcgcccccg ccccccgctt ttttttttttt tttttctnaa 780

acccccccacc ccccccccccc cccttnnttt gtttccgctt ttattccaag aaaat 835

<210> 50
<211> 645
<212> DNA
<213> Homo sapiens 3.E.04

<400> 50
gcggccgccg gcttgacgtg tacggcgctg atcacctacg cttgctgggg gcagctgccg 60

ccgctgccct gggcgtcgcc aaccccgtcg cgaccggtgg gcgtgctgct gtggtgggag 120

cccttcgggg ggcgcgatag cgccccgagg ccgccccctg actgccggct gcgcttcaac 180

atcagcggct gccgcctgct caccgaccgc gcgtcctacg gagaggctca ggccgtgctt 240

ttccaccacc gcgacctcgt gaaggggccc cccgactggc ccccgccctg gggcatccag 300

gcgcacactg ccgaggaggt ggatctgcgc gtgttggact acgaggaggc agcggcggcg 360

gcagaagccc tggcgacctc cagccccagg ccccgggccc agcgctgggt ttggatgaac 420

ttcgagtcgc cctcgcactc cccggggctg cgaagcctgc aagtaacctc ttcaactgga 480

cgctctccta ccgggcggac tcggacgtct ttgtgcctta tggctacctc taccccagaa 540

gccacccccgg cgacccgcct cagcctggcc ccgcactgtc caggaaacaa gggctggtgg 600

catgggtggt gagccacttg ggacgagcgc caggcccggg tccgt 645

<210> 51
<211> 1021
<212> DNA
<213> Homo sapiens 3.E.50

<220>
<221> n
<222> (744)..(744)
<223> a or g or c or t

<400> 51
gcggccgcgg gacggggaga tgcggccccg gtattgatgt cgaaaatgat ggataacgcg 60

IDT Exhibit 1020.8272

ggaatggcaa atatactatt tgtctaatgg ctcggcaatt aaattcccct gtaaatgacc 120.
catgcctcat ttcatcctaa tctatggaat tttgattgaa ttcgtcagct ctaattgaaa 180
aatactgcac tttaatgtct gcattgcagt ttcaggacga gagtggtttt aatgagacag 240
tgcccccatg acccgggaat atttgagact tttattcgga atttaaagcc aggagattgc 300
tcgactgagc cctgagattt cctctcctgt atccacgtcc atccatctcc agacgcgatt 360
taataaacgc acttaaggat aaatgcgccc ccgaccctcg cgccaacgtg ttaccccacg 420
ggcgcccctc ctcggaataa gggacggcgg aggccgggga ggcgggggag ttgggggct 480
cagaaggtcc tggtccctcc ccggcccaag tttccctgcc ctccctgcca ccctggtccc 540
caggcactgt cgcggacccc agactccgcc ttccctaggc caaacctagg cgacctccct 600
ggactaggag gcctggctgc ctgccacccg cgcaccggaa gaagggactc gcgcactcgg 660
agaaggggcc gggccccgac gcgctttata tgcaaatggc gaggcgaagc catccctgag 720
aaatagctac ttgctgaagc tatntactag attgaaatga gttaagagaa acatttaagt 780
cgtgcaacga gataattggg ccgattaact ggggatgttt gctctttcaa aaaaaaaaaa 840
aaaaaaaccg ccgaggagga gagagcagta agccgcgttg attgagccca ctgtcaagac 900
cgaattccga tgcgggacgg tcctcgggac tcgaagagac ccacggagga ctgagaggct 960
ttcgccggcc gcgcatttct tttcaggcat ccaccggcca gggcctagaa gtccgaaagg 1020
c 1021

<210> 52
<211> 518
<212> DNA
<213> Homo sapiens 3.E.55

<400> 52
gcggccgcag gaaccacgat gagaggcagg agctgctcct ggctgagggg cttcaaccac 60
tcgccgagga ggagcagagg gcctaggagg accccgggcg tggaccaccc gccctggcag 120
ttgaatgggg cggcaattgc ggggcccacc ttagaccgaa ggggaaaacc cgctctctca 180
ggcgcatgtg ccagttgggg ccccgcgggt agatgccggc aggccttccg gaagaaaaag 240
agccattggt ttttgtagta ttggggccct cttttagtga tactggattg gcgttgtttg 300
tggctgttgc gcacatccct gccctcctac agcactccac cttgggacct gtttagagaa 360
gccggctctt caaagacaat ggaaactgta ccatacacat tggaaggctc cctaacacac 420
acagcgggga agctgggccg agtaccttaa tctgccataa agccattctt actcgggcga 480
ccccttaag tttagaaata attgaaagga aatgtttg 518

<210> 53
<211> 498
<212> DNA
<213> Homo sapiens 3.E.57

IDT Exhibit 1020.8273

<400> 53
gcggccgccc ccacggctcc accctctcgg cggggccgca gccatctggg gcccctgcca 60
gtagcggccg ccttccgctc agcctctggt cccaggcgag cctggcgagc cggcgaagca 120
ccggcgggga ggaggactag aacaggagga ggggcacggc ggattgaagc gagctgggct 180
gtgagcaagg gacacccaca gcctggagaa acagccccgc tctcttgcgc gctgtctgct 240
ccagccgcta ctgggggctc taagcagcgc gatgctgctt cgcttcttct aggcggcggc 300
cggcggaggc tttccgcagc cgcttggccg gcgccggccc ctattccgtt ggcaagtccc 360
ttgtctatcc cggagggcgc acccggacgc tcgagccgga gcgagcgcga agtccgaagt 420
ccgcccccag agccgccaac ttccctgtga gcccctctcc ccgccgcagc ctgcgccaga 480
cctgggagcg atgcgccc 498

<210> 54
<211> 471
<212> DNA
<213> Homo sapiens 3.E.59

<400> 54
gcggccgccc gggcccgcgg gcgggggggat cggcgggggg gacccgcggg gtgaccggcg 60
gcaggagccg ccaccatgga gttccgccag gaggagtttc ggaagctagc gggtcgtgct 120
ctcgggaagc tgcaccggtg agcctggcgg gggtcccggg agaagagtgg gaggatctga 180
ggaggatgct aattcccacc tgggcgcaga ctgacagatg aacgggcgat accccggcat 240
gggggtccac ccatctgtcc agttttctgc cgtgggctcc gacggcgctg ttctccctgg 300
tcgagccttg tccattatcc tgttcctttt tctgcacccc accccacccg gctccactct 360
ctctggtgct gtaaatgcct ctctcccggg tctctggctc ctcccccacc acttctgggt 420
ctctgtcccc gtctctttct ggatgtctct gccccttttc tctctgggtc t 471

<210> 55
<211> 971
<212> DNA
<213> Homo sapiens 3.F.16

<220>
<221> n
<222> (905)..(905)
<223> a or g or c or t

<400> 55
gcggccgccg tgggcctgca aaacttccaa agtagcagcc tgtttctcct cgtctccctt 60
ctcctgggta cccagcgccc cgccttcccc agaaagggcg aggggtgggg gcagggctcc 120
ctcgggaggt ggccaagcgc cgggacgcgc tcccagcgtt actcaggaca cttgggattt 180
ggcctgcagc ccccttcccc atccctggcc tggctgcggt gtcccttgct cccctctgct 240
gctgctcctg ccccatcaag tcgaaaatct gagggtggga tggggtgggg gaccaggggg 300
taccctccca ggccgctccg cagcaggccg aggtggagac cctgcccggt aggcgagtcc 360

IDT Exhibit 1020.8274

ttgtgcccac agctcggagc cagcagcgga gtgacaaaaa agataaagtt ggtgaatgat 420

aaagaccgta ttttccacgc tttgggtgcg ggaccagatg atctagaaaa tgagctgaaa 480

tggattcagc ctccgagcct gttgtgagag cagctgattc ccccatttcg ggccagatgg 540

ctgctgaaca cagatttgca ttcattttcg cttaatatcg tccaaaatag tggggcagct 600

gcatttgttg tcaaaaaggt ttaaaacccc ttttctttct ggggcaggat cgttacctta 660

tgtgatgggc ttatagaact tttttttcct ctttagtcaa cagtatcaga tttagaagga 720

tttgttttta aaccttctaa tttggtaatc agatttaaat cgccttggcg cgtgtaatct 780

gaattaaaga tactgtaaat gattntaagc atgatacttt cgttagcgca aggaaggggc 840

acctctagca caggctggac attttaggaa gtgtgctata aaggagcatt gttcctattt 900

caacttaatc ttccgaaaag gctttggtat tctgcataac gctgctggcg ttgcctggtg 960

agcccgagag t 971

<210> 56
<211> 550
<212> DNA
<213> Homo sapiens 3.F.2

<400> 56
gcggccgcac gcgggtgcta atttgcacac atcaagactg aagtgtagtg aggaaacgtt 60

gagtttctgt tttcaaacct ttaacttcgt aattagagat ttaacaactt gaaggggggc 120

ggggagaggc gggggaggag gtgggcagaa ggaataaaac tccatctaaa attcctaata 180

gcaattcctt agaattataa actgcgagat gatcagaagt gacatctttg ccttctttga 240

aggctctctt ctctaagtta ctaataatga taatgcacgt tcgggtacag aaatatgagc 300

caagaactca agtctgcaat gaaggagtgg acatgacagc gtaagaggga gcatcattgt 360

ttgatctatt ttaacctttt ccgtctcaaa gatacgatgg tgcttcctcc aggaagaaaa 420

gcctgtaagc tcaaacaaga gctcccctgg aacagaagac actggagacc gtaagaggtg 480

ggaggttgga agggggaaaa ggatagaaaa actgcctgtt gggtattatg ctcaccacat 540

gggtgacggg 550

<210> 57
<211> 870
<212> DNA
<213> Homo sapiens 3.F.50

<220>
<221> n
<222> (667)..(667)
<223> a or g or c or t

<220>
<221> n
<222> (684)..(684)
<223> a or g or c or t

IDT Exhibit 1020.8275

<220>
<221> n
<222> (725)..(725)
<223> a or g or c or t

<220>
<221> n
<222> (758)..(758)
<223> a or g or c or t

<220>
<221> n
<222> (776)..(776)
<223> a or g or c or t

<220>
<221> n
<222> (777)..(777)
<223> a or g or c or t

<220>
<221> n
<222> (789)..(789)
<223> a or g or c or t

<220>
<221> n
<222> (845)..(845)
<223> a or g or c or t

<220>
<221> n
<222> (870)..(870)
<223> a or g or c or t

<400> 57
ttagactctc actgggcagg tctgctgtcc cctctgctcc cgcaggactg gagccaccga 60

gctcgcgcct tcttctcggg gtgcgatttc tctcctcttt tggactcaag atcaatgctt 120

cccggccggc gcagatcaca cagcaggacc ccaggggaga ctgtggcctt cttcccgcct 180

cccaattccc caagaccgcc tctagaggct gctgtgtccg gagaactccg agcattttct 240

ggacacagat tgcctaacag aggaacaggg gttaggtggg gagcggctgg ccggcccaaa 300

cacagcagcc ccaagctggc tcccaagcct gggctctcca cccccgctcc catcctctct 360

tgagcacagt taggcccaac acccctgtcc cccaaaacac ctcctaccct ccctcccccc 420

cagcccccat cttcaggaac atcacagggc tcacactcac taaccgcgga gagcacatgc 480

aggccggagc cctcagcccg gcagctctcg gaccctgccc agctcgacgc ggactcatgc 540

agaagaggac attccgcagg taggtacaat cccagcgctg gggcctgggg cgtccggggg 600

gcggcctttg agcttccccg ataccgctcg cctgctcccg gagctgttcg gccgacggct 660

IDT Exhibit 1020.8276