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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 April 2009 (30.04.2009)

PCT

(10) International Publication Number
WO 2009/055732 A1

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

(21) International Application Number:
PCT/US2008/081206

(22) International Filing Date: 24 October 2008 (24.10.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/983,085 26 October 2007 (26.10.2007) US

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

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **RAYMOND, Christopher, K.** [US/US]; 2626 NE 86th Street, Seattle, WA 98115 (US). **ARMOUR, Christopher** [US/US]; 13531 104th Place NE, Kirkland, WA 98034 (US). **CASTLE, John** [US/US]; 7006 38th Avenue NE, Seattle, WA 98115 (US).

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

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

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

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*
— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

(54) **Title:** CDNA SYNTHESIS USING NON-RANDOM PRIMERS

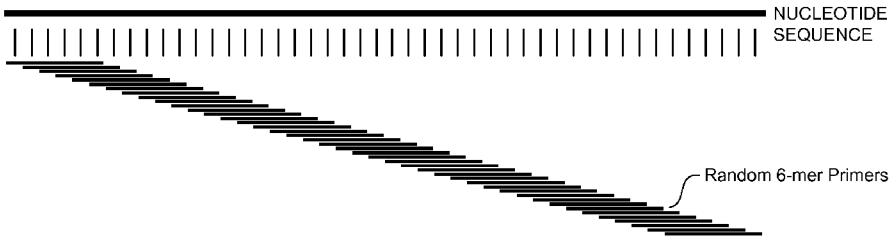

Fig.1A.

(57) **Abstract:** The present invention provides methods for selectively amplifying a target population of nucleic acid molecules in a population of RNA template molecules (e.g., all mRNA molecules expressed in a cell type except for the most highly expressed mRNA species). The present invention also provides a first population of oligonucleotides including the nucleic acid sequences set forth in SEQ ID NOS: 1-749 and a second population of oligonucleotides including the nucleic acid sequences set forth in SEQ ID NOS:750-1498. The first population of oligonucleotides can be used, for example, to prime the synthesis of first strand cDNA molecules complementary to mRNA molecules isolated from mammalian cells without priming the synthesis of cDNA molecules complementary to ribosomal RNA molecules. The second population of oligonucleotides can be used, for example, to prime the second strand synthesis of primer extension products (first strand cDNA) complementary to mRNA molecules isolated from mammalian cells without priming the second strand synthesis of primer extension products synthesized from ribosomal RNA molecules.

WO 2009/055732 A1

CDNA SYNTHESIS USING NON-RANDOM PRIMERS

FIELD OF THE INVENTION

The present invention relates to methods of selectively amplifying target nucleic acid molecules and oligonucleotides useful for priming the amplification of target nucleic acid molecules.

BACKGROUND

Gene expression analysis often involves amplification of starting nucleic acid molecules. Amplification of nucleic acid molecules may be accomplished by reverse transcription (RT), *in vitro* transcription (IVT) or the polymerase chain reaction (PCR), either individually or in combination. The starting nucleic acid molecules may be mRNA molecules, which are amplified by first synthesizing complementary cDNA molecules, then synthesizing second cDNA molecules that are complementary to the first cDNA molecules, thereby producing double stranded cDNA molecules. The synthesis of first strand cDNA is typically accomplished using a reverse transcriptase and the synthesis of second strand cDNA is typically accomplished using a DNA polymerase. The double stranded cDNA molecules may be used to make complementary RNA molecules using an RNA polymerase, resulting in amplification of the original starting mRNA molecules. The RNA polymerase requires a promoter sequence to direct initiation of RNA synthesis. Complementary RNA molecules may, for example, be used as a template to make additional complementary DNA molecules. Alternatively, the double stranded cDNA molecules may be amplified, for example, by PCR and the amplified PCR products may be used as sequencing templates or in microarray analysis.

Amplification of nucleic acid molecules requires the use of oligonucleotide primers that specifically hybridize to one or more target nucleic acid molecules in the starting material. Each oligonucleotide primer may include a promoter sequence that is located 5' to the hybridizing portion of the oligonucleotide that hybridizes to the target nucleic acid molecule(s). If the hybridizing portion of an oligonucleotide is too short, then the oligonucleotide does not stably hybridize to a target nucleic acid molecule and priming and subsequent amplification does not occur. Also, if the hybridizing portion of an oligonucleotide is too short, then the oligonucleotide does not specifically hybridize to

one or a small number of target nucleic acid molecules, but nonspecifically hybridizes to numerous target nucleic acid molecules.

Amplification of a complex mixture of different target nucleic acid molecules (e.g., RNA molecules) typically requires the use of a population of numerous oligonucleotides having different nucleic acid sequences. The cost of the oligonucleotides increases with the length of the oligonucleotides. In order to control costs, it is preferable to make oligonucleotide primers that are no longer than the minimum length required to ensure specific hybridization of an oligonucleotide to a target sequence.

It is often undesirable to amplify highly expressed RNAs (e.g., ribosomal RNAs). For example, in gene expression experiments that analyze expression of genes in blood cells, amplification of numerous copies of abundant globin mRNAs, or ribosomal RNAs, may obscure subtle changes in the levels of rare mRNAs. Thus, there is a need for populations of oligonucleotide primers that selectively amplify desired nucleic acid molecules within a population of nucleic acid molecules (e.g., oligonucleotide primers that selectively amplify all mRNAs that are expressed in a cell except for the most highly expressed RNAs). In order to reduce the cost of synthesizing the population of oligonucleotides, the hybridizing portion of each oligonucleotide should be no longer than necessary to ensure specific hybridization to a desired target sequence under defined conditions.

SUMMARY

In one aspect, the present invention provides methods for selectively amplifying a target population of nucleic acid molecules within a larger non-target population of nucleic acid molecules (e.g., all RNA molecules expressed in a cell type except for the most highly expressed RNA species). The methods of this aspect of the invention each include the steps of (a) providing a population of single-stranded primer extension products synthesized from a population of RNA template molecules in a sample isolated from a mammalian subject using reverse transcriptase enzyme and a first population of oligonucleotide primers, wherein each oligonucleotide in the first population of oligonucleotide primers comprises a hybridizing portion and a defined sequence portion located 5' to the hybridizing portion, wherein the population of RNA template molecules comprises a target population of nucleic acid molecules and a non-target population of

nucleic acid molecules; (b) synthesizing double-stranded cDNA from the population of single-stranded primer extension products according to step (a) using a DNA polymerase and a second population of oligonucleotide primers, wherein each oligonucleotide in the second population of oligonucleotides comprises a hybridizing portion, wherein the hybridizing portion consists of one of 6, 7, or 8 nucleotides and a defined sequence located 5' to the hybridizing portion wherein the hybridizing portion is selected from all possible oligonucleotides having a length of 6, 7, or 8 nucleotides that do not hybridize under the defined conditions to the non-target population of nucleic acid molecules in the synthesized single-stranded cDNA. In some embodiments, each oligonucleotide in the first population of oligonucleotide comprises a random hybridizing portion and a defined sequence located 5' to the hybridizing portion.

In another aspect, the present invention provides methods of selectively amplifying a target population of nucleic acid molecules within a larger non-target population of nucleic acid molecules. The methods of this aspect of the invention comprise the steps of (a) synthesizing single-stranded cDNA from a sample comprising total RNA isolated from a mammalian subject using reverse transcriptase enzyme and a first population of oligonucleotide primers, wherein each oligonucleotide within the first population of oligonucleotide primers comprises a hybridizing portion and a defined sequence portion located 5' to the hybridizing portion, wherein the hybridizing portion is a member of the population of oligonucleotides comprising SEQ ID NOS:1-749; and (b) synthesizing double-stranded cDNA from the single-stranded cDNA synthesized according to step (a) using a DNA polymerase and a second population of oligonucleotide primers, wherein each oligonucleotide within the second population of oligonucleotide primers comprises a hybridizing portion and a defined sequence portion located 5' to the hybridizing portion, wherein the hybridizing portion is a member of the population of oligonucleotides comprising SEQ ID NOS:750-1498.

In another aspect, the present invention provides methods for transcriptome profiling. The methods of this aspect of the invention comprise (a) synthesizing a population of single stranded primer extension products from a target population of nucleic acid molecules within a population of RNA template molecules in a sample isolated from a subject using reverse transcriptase enzyme and a first population of oligonucleotide primers comprising a hybridizing portion and a first PCR primer binding

site located 5' to the hybridizing portion; (b) synthesizing double-stranded cDNA from the population of single-stranded primer extension products generated according to step (a) using a DNA polymerase and a second population of oligonucleotide primers comprising a hybridizing portion and a second PCR primer binding site located 5' to the hybridizing portion; and (c) PCR amplifying the double-stranded cDNA generated according to step (b) using a first PCR primer that binds to the first PCR primer binding site and a second PCR primer that binds to the second PCR primer binding site, wherein the non-target population of nucleic acid molecules consists essentially of ribosomal RNA and mitochondrial ribosomal RNA of the same species as the mammalian subject.

In another aspect, the present invention provides populations of oligonucleotides comprising SEQ ID NOS:1-749. These oligonucleotides can be used, for example, to prime the synthesis of first-strand cDNA molecules complementary to RNA molecules isolated from a mammalian subject without priming the synthesis of first strand cDNA molecules complementary to ribosomal RNA (18S, 28S) or mitochondrial ribosomal RNA (12S, 16S) molecules. In some embodiments, each oligonucleotide in the population of oligonucleotides further comprises a defined sequence portion located 5' to the hybridizing portion. In one embodiment, the defined sequence portion comprises a transcriptional promoter, which may be used as a primer binding site in PCR amplification, or for *in vitro* transcription. In another embodiment, the defined sequence portion comprises a primer binding site that is not a transcriptional promoter. For example, in some embodiments the present invention provides populations of oligonucleotides wherein a transcriptional promoter, such as the T7 promoter (SEQ ID NO:1508), is located 5' to a member of the population of oligonucleotides having the sequences set forth in SEQ ID NOS:1-749. Thus, in some embodiments, the present invention provides populations of oligonucleotides wherein each oligonucleotide consists of the T7 promoter (SEQ ID NO:1508) located 5' to a different member of the population of oligonucleotides having the sequences set forth in SEQ ID NOS:1-749. In further embodiments, the present invention provides populations of oligonucleotides wherein the defined sequence portion comprises at least one primer binding site that is useful for priming a PCR synthesis reaction and that does not include an RNA polymerase promoter sequence. A representative example of a defined sequence portion for use in such embodiments is provided as 5'TCCGATCTCT3' (SEQ ID NO:1499),

which is preferably located 5' to a member of the population of oligonucleotides having the sequences set forth in SEQ ID NOS:1-749.

In another aspect, the present invention provides populations of oligonucleotides comprising SEQ ID NOS:750-1498. These oligonucleotides can be used, for example, to prime the synthesis of second strand cDNA molecules complementary to first strand cDNA molecules synthesized from RNA isolated from a mammalian subject without priming the synthesis of second strand cDNA molecules complementary to first strand cDNA reverse transcribed from ribosomal RNA (18S, 28S) or mitochondrial ribosomal RNA (12S, 16S) molecules. In some embodiments, each oligonucleotide in the population of oligonucleotides further comprises a defined sequence portion located 5' to the hybridizing portion. In one embodiment, the defined sequence portion comprises a transcriptional promoter, which may be used as a primer binding site in PCR amplification or for *in vitro* transcription. In another embodiment, the defined sequence portion comprises a primer binding site that is not a transcriptional promoter. For example, in some embodiments, the present invention provides populations of oligonucleotides wherein a transcriptional promoter, such as the T7 promoter (SEQ ID NO:1508), is located 5' to a member of the population of oligonucleotides having the sequences set forth in SEQ ID NOS:750-1498. Thus, in some embodiments, the present invention provides populations of oligonucleotides wherein each oligonucleotide consists of the T7 promoter (SEQ ID NO:1508) located 5' to a different member of the population of oligonucleotides having the sequences set forth in SEQ ID NOS:750-1498. In further embodiments, the present invention provides populations of oligonucleotides wherein the defined sequence portion comprises at least one primer binding site that is useful for priming a PCR synthesis reaction and that does not include an RNA polymerase promoter sequence. A representative example of a defined sequence portion for use in such embodiments is provided as 5'TCCGATCTGA3' (SEQ ID NO:1500), which is preferably located 5' to a member of the population of oligonucleotides having the sequences set forth in SEQ ID NOS:750-1498.

In another aspect, the present invention provides a reagent for selectively amplifying a target population of nucleic acid molecules in a larger population of non-target nucleic acid molecules. In one embodiment, the reagent comprises at least 10% of the oligonucleotides comprising SEQ ID NOS:1-749. In another embodiment,

the reagent comprises at least 10% of the oligonucleotides comprising SEQ ID NOS:750-1498.

In another aspect, the present invention provides a kit for selectively amplifying a target population of nucleic acid molecules. The kit of this aspect of the invention comprises a reagent comprising a first population of oligonucleotides for first strand cDNA synthesis, wherein each oligonucleotide in the first population of oligonucleotides comprises a hybridizing portion and a defined sequence portion located 5' to the hybridizing portion, wherein the hybridizing portion is a member of the population of oligonucleotides comprising SEQ ID NOS:1-749. In some embodiments, the kit further comprises a second population of oligonucleotides for second strand cDNA synthesis, wherein each oligonucleotide in the second population of oligonucleotides comprises a hybridizing portion and a defined sequence portion located 5' to the hybridizing portion, wherein the hybridizing portion is a member of the population of oligonucleotides comprising SEQ ID NOS:750-1498.

In another aspect, the present invention provides a population of selectively amplified nucleic acid molecules comprising a representation of a transcriptome of a mammalian subject comprising a 5' defined sequence, a population of amplified sequences corresponding to a nucleic acid expressed in the mammalian subject, a 3' defined sequence wherein the population of amplified sequences is characterized by having the following properties with reference to the particular mammalian species: (a) having greater than 75% polyadenylated and non-polyadenylated transcripts and having less than 10% ribosomal RNA.

DESCRIPTION OF THE DRAWINGS

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

FIGURE 1A shows the number of exact matches for random 6-mers (N6) oligonucleotides on nucleotide sequences in the human RefSeq transcript database as described in Example 1;

FIGURE 1B shows the number of exact matches for Not-So-Random (NSR) 6-mer oligonucleotides on nucleotide sequences in the human RefSeq transcript database as described in Example 1;

FIGURE 1C shows a representative embodiment of the methods of the invention for synthesizing a preparation of selectively amplified cDNA molecules using a mixture of random primers for first strand cDNA synthesis and a mixture of anti-NSR 6-mer oligonucleotides for second strand cDNA synthesis, as described in Example 2;

FIGURE 1D shows a representative embodiment of the methods of the invention for synthesizing a preparation of selectively amplified aDNA molecules using a mixture of NSR 6-mer oligonucleotides for first strand cDNA synthesis and a mixture of anti-NSR6-mer oligonucleotides for second strand cDNA synthesis, followed by PCR amplification, as described in Example 2 and Example 4;

FIGURE 2 is flow diagram illustrating a method of whole transcriptome analysis of a subject comprising selectively amplifying nucleic acid molecules from RNA isolated from the subject followed by sequence analysis or microarray analysis of the amplified nucleic acid molecules as described in Example 4 and Example 5;

FIGURE 3A is a histogram plot on a logarithmic scale showing the relative abundance of 18S, 28S, 12Sn and 16S (normalized to gene and N8) in a population of first strand cDNA molecules synthesized using various NSR-6 pools as compared to first strand cDNA generated using random primers (N8=100%) as described in Example 3;

FIGURE 3B graphically illustrates the relative levels of abundance of cytoplasmic rRNA (18S or 28S) in cDNA amplified using random primers (N7) in both first strand and second strand synthesis (N7>N7 = 100% 18S, 100% 28S) as compared to cDNA amplified using NSR primers (SEQ ID NOS:1-749) in the first strand followed by random primers (N7) in the second strand (NSR>N7 = 3.0% 18S, 3.4% 28S), and as compared to cDNA amplified using NSR primers (SEQ ID NOS:1-749) in the first strand followed by anti-NSR primers (SEQ ID NOS:750-1498) in the second strand (NSR>anti-NSR = 0.1% 18S, 0.5% 28S) as described in Example 3;

FIGURE 3C graphically illustrates the relative levels of abundance of mitochondrial rRNA (12S or 16S) in cDNA amplified using random primers (N7) in both first strand and second strand synthesis (N7>N7 =100% 12S, or 16S) as compared to cDNA amplified using NSR primers (SEQ ID NOS:1-749) in the first strand followed by

random primers (N7) in the second strand (NSR>N7 =27% 12S, 20.4% 16S), and as compared to cDNA amplified using NSR primers (SEQ ID NOS:1-749) in the first strand followed by anti-NSR primers (SEQ ID NOS:750-1498) in the second strand (NSR>anti-NSR =8.2% 12S, 3.5% 16S) as described in Example 3;

FIGURE 4A is a histogram plot showing the gene-specific polyA content of representative gene transcripts in cDNA synthesized using various NSR primers during first strand synthesis as described in Example 3;

FIGURE 4B is a histogram plot showing the relative abundance level of representative non-polyadenylated RNA transcripts in cDNA amplified from Jurkat-1 and Jurkat-2 total RNA using various NSR primers during first strand cDNA synthesis as described in Example 3;

FIGURE 5 graphically illustrates the log ratio of Jurkat/K562 mRNA expression data measured in cDNA generated using NSR-6mers (x-axis) versus the log ratio of Jurkat/K562 mRNA expression data measured in cDNA generated using random primers (N8), as described in Example 3;

FIGURE 6A graphically illustrates the proportion of rRNA to mRNA in total RNA typically obtained after polyA purification, demonstrating that even after 95% removal of rRNA from total RNA, the remaining RNA consists of a mixture of about 50% rRNA and 50% mRNA as described in Example 3;

FIGURE 6B graphically illustrates the proportion of rRNA to mRNA in a cDNA sample prepared using NSR primers during first strand cDNA synthesis and anti-NSR primers during second strand cDNA synthesis. As shown, in contrast to polyA purification, the use of NSR primers and anti-NSR primers to generate cDNA from total RNA is effective to remove 99.9% rRNA, resulting in a cDNA population enriched for greater than 95% mRNA as described in Example 3;

FIGURE 7A graphically illustrates the detection and positional distribution of polyA+ RefSeq mRNA in NSR-primed (dotted line) or expressed sequence tag (EST) (solid line) cDNAs across long transcripts (≥4kb), illustrating the combined read frequencies for 5,790 transcripts shown at each base position starting from the 5' termini, as described in Example 7;

FIGURE 7B graphically illustrates the detection and positional distribution of polyA+ RefSeq mRNA in NSR-primed (dotted line) or expressed sequence tag (EST)

(solid line) cDNAs across long transcripts (≥4kb), illustrating the combined read frequencies for 5,790 transcripts shown at each base position starting from the 3' termini, as described in Example 7; and

FIGURE 8 graphically illustrates the enrichment of small nucleolar RNAs (snoRNAs) encoded by the Chromosome 15 Prader-Willi neurological disease locus in NSR-primed cDNA generated from RNA isolated from whole brain relative to NSR-primed cDNA generated from RNA isolated from the Universal Human Reference (UHR) cell line, as described in Example 7.

DETAILED DESCRIPTION

Unless specifically defined herein, all terms used herein have the same meaning as they would to one skilled in the art of the present invention. Practitioners are particularly directed to Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2d ed., Cold Spring Harbor Press, Plainsview, New York; and Ausubel et al., *Current Protocols in Molecular Biology* (Supplement 47), John Wiley & Sons, New York, 1999, for definitions and terms of the art.

The use of Not-So-Random ("NSR") 6-mer primers for first strand cDNA synthesis is described in co-pending U.S. Patent Application Serial No. 11/589,322, filed October 27, 2006, incorporated herein by reference. In a particular embodiment, the NSR-6mers described in co-pending U.S. Patent Application Serial No. 11/589,322 comprise populations of oligonucleotides that hybridize to all mRNA molecules expressed in blood cells but that do not hybridize to globin mRNA (HBA1, HBA2, HBB, HBD, HBG1 and HBG2) or to nuclear ribosomal RNA (18S and 28S rRNA). In the present application, a different population of NSR primers (SEQ ID NOS:1-749) is provided that includes oligonucleotides that hybridize to all mRNA molecules expressed in mammalian cells, including globin mRNA, but that do not hybridize to nuclear ribosomal RNA (18S and 28S rRNA) and mitochondrial ribosomal RNAs (12S and 16S mt-rRNA). The present application further provides a second population of anti-NSR oligonucleotides (SEQ ID NOS:750-1498) for use during second strand cDNA synthesis. The anti-NSR oligonucleotides (SEQ ID NOS:750-1498) are selected to hybridize to all first strand cDNA molecules reverse transcribed from RNA templates expressed in mammalian cells, including globin mRNA, but that do not hybridize to first strand cDNA molecules transcribed from nuclear ribosomal RNA (18S and 28S rRNA) and

mitochondrial ribosomal RNAs (12S and 16S mt-rRNA). As described in Examples 1-4, the use of a first round of selective amplification using NSR primers (SEQ ID NOS:1-749) during first strand synthesis followed by a second round of selective amplification using anti-NSR primers (SEQ ID NOS:750-1498) during second strand synthesis results in a population of double stranded cDNA that represents substantially all of the polyA RNA and non-polyA RNA expressed in the cell, with a very low level (less than 10%) of nucleic acid molecules representing unwanted nuclear ribosomal RNA and mitochondrial ribosomal RNA. As shown in FIGURE 2, the invention also provides methods which analyze the products of the amplification methods of the invention, such as sequencing and gene expression profiling (e.g., microarray analysis).

In accordance with the foregoing, in one aspect, the present invention provides methods for selectively amplifying a target population of nucleic acid molecules within a larger non-target population of nucleic acid molecules (e.g., all RNA molecules expressed in a cell type except for the most highly expressed RNA species). The methods of this aspect of the invention each include the steps of (a) synthesizing single-stranded cDNA from RNA in a sample isolated from a mammalian subject using reverse transcriptase enzyme and a first population of oligonucleotide primers, wherein each oligonucleotide in the first population of oligonucleotide primers comprises a hybridizing portion and a defined sequence portion located 5' to the hybridizing portion, wherein the RNA comprises a target population of nucleic acid molecules within a larger non-target population of nucleic acid molecules; and (b) synthesizing double-stranded cDNA from the single-stranded cDNA synthesized according to step (a) using a DNA polymerase and a second population of oligonucleotide primers, wherein each oligonucleotide in the second population of oligonucleotides comprises a hybridizing portion, wherein the hybridizing portion consists of one of 6, 7, or 8 nucleotides and a defined sequence located 5' to the hybridizing portion wherein the hybridizing portion is selected from all possible oligonucleotides having a length of 6, 7, or 8 nucleotides that do not hybridize under the defined conditions to the non-target population of nucleic acid molecules in the synthesized single-stranded cDNA.

The second population of oligonucleotides may also include a defined sequence portion located 5' to the hybridizing portion. In one embodiment, the defined sequence

portion comprises a transcriptional promoter that can also be used as a primer binding site. Therefore, in certain embodiments of this aspect of the invention, each oligonucleotide of the second population of oligonucleotides comprises a hybridizing portion that consists of 6 nucleotides or 7 nucleotides or 8 nucleotides and a transcriptional promoter portion located 5' to the hybridizing portion. In another embodiment, the defined sequence portion of the second population of oligonucleotides includes a second primer binding site for use in a PCR amplification reaction and that may optionally include a transcriptional promoter. By way of example, the populations of anti-NSR oligonucleotides provided by the present invention are useful in the practice of the methods of this aspect of the invention.

For example, in one embodiment of the present invention, a population of oligonucleotides (SEQ ID NOS:750-1498), that each has a length of 6 nucleotides, was identified that can be used as primers to prime the second strand synthesis of all, or substantially all, first strand cDNA molecules synthesized from a target population of RNA molecules from mammalian cells but that do not prime the second strand synthesis of first strand cDNA reverse transcribed from non-target ribosomal RNA (rRNA) or mitochondrial rRNA (mt-rRNA) from mammalian cells. The identified second population of oligonucleotides (SEQ ID NOS:750-1498) is referred to as anti-Not-So-Random (anti-NSR) primers. Thus, this population of oligonucleotides (SEQ ID NOS:750-1498) can be used to prime the second strand synthesis of a population of first strand nucleic acid molecules (e.g., cDNAs) that are representative of a starting population of mRNA molecules isolated from mammalian cells but do not prime second strand synthesis of cDNA molecules that correspond to rRNA or mt-rRNAs.

In other embodiments, each oligonucleotide in the first population of oligonucleotides comprises a hybridizing portion, wherein the hybridizing portion consists of one of 6, 7, or 8 nucleotides and a defined sequence located 5' to the hybridizing portion wherein the hybridizing portion is selected from all possible oligonucleotides having a length of 6, 7, or 8 nucleotides that do not hybridize under the defined conditions to the non-target population of nucleic acid molecules in a sample comprising RNA from a mammalian subject.

The first population of oligonucleotides may also include a defined sequence portion located 5' to the hybridizing portion. In one embodiment, the defined sequence

portion comprises a transcriptional promoter that can also be used as a first primer binding site. Therefore, in certain embodiments of this aspect of the invention, each oligonucleotide of the first population of oligonucleotides comprises a hybridizing portion that consists of 6 nucleotides or 7 nucleotides or 8 nucleotides and a transcriptional promoter portion located 5' to the hybridizing portion. In another embodiment, the defined sequence portion of the first population of oligonucleotides includes a first primer binding site for use in a PCR amplification reaction and that may optionally include a transcriptional promoter. By way of example, the populations of NSR oligonucleotides provided by the present invention are useful in the practice of the methods of this aspect of the invention.

For example, in one embodiment of the present invention, a first population of oligonucleotides (SEQ ID NOS:1-749) wherein each has a length of 6 nucleotides, was identified that can be used as primers to prime the first strand synthesis of all, or substantially all, mRNA molecules from mammalian cells, but that do not prime the amplification of non-target ribosomal RNA (rRNA) or mitochondrial rRNA (mt-rRNA) from mammalian cells. The identified first population of oligonucleotides (SEQ ID NOS:1-749) is referred to as Not-So-Random (NSR) primers. Thus, this population of oligonucleotides (SEQ ID NOS:1-749) can be used to prime the first strand synthesis of a population of nucleic acid molecules (e.g., cDNAs) that are representative of a starting population of mRNA molecules isolated from mammalian cells but do not prime first strand synthesis of cDNA molecules that correspond to rRNA or mt-rRNAs.

The present invention also provides a first population of oligonucleotides for priming first strand cDNA synthesis, wherein a defined sequence, such as the T7 promoter (SEQ ID NO:1508) or a first primer binding site (SEQ ID NO:1499) is located 5' to a member of the population of oligonucleotides having the sequences set forth in SEQ ID NOS:1-749. Thus, each oligonucleotide may include a hybridizing portion (selected from SEQ ID NOS:1-749) that hybridizes to target nucleic acid molecules (e.g., mRNAs) and a defined sequence, such as a promoter sequence or first primer binding site, is located 5' to the hybridizing portion. The defined sequence portion may be incorporated into DNA molecules amplified using the oligonucleotides (that include the T7 promoter) as primers and can thereafter promote transcription from the DNA molecules.

Alternatively, the defined sequence portion, such as the transcriptional promoter or first primer binding site, may be covalently attached to the cDNA molecule, for example, by DNA ligase enzyme.

Useful transcription promoter sequences include the T7 promoter (5'AATTAATACGACTCACTATAGGGAGA3' (SEQ ID NO:1508)), the SP6 promoter (5'ATTTAGGTGACACTATAGAAGNG3' (SEQ ID NO:1509)), and the T3 promoter (5'AATTAACCCTCACTAAAGGGAGA3' (SEQ ID NO:1510)).

The target nucleic acid population can include, for example, all mRNAs expressed in a cell or tissue except for a selected group of non-target mRNAs such as, for example, the most abundantly expressed mRNAs. A non-target abundantly expressed mRNA typically constitutes at least 0.1% of all the mRNA expressed in the cell or tissue (and may constitute, for example, more than 50% or more than 60% or more than 70% of all the mRNA expressed in the cell or tissue). An example of an abundantly expressed non-target mRNA is ribosomal rRNA or mitochondrial rRNA in mammalian cells. Other examples of abundantly expressed non-target RNA that one could selectively eliminate using the methods of the invention include, for example, globin mRNA (from blood cells) or chloroplast rRNA (from plant cells).

The methods of the invention are useful for transcriptome profiling of total RNA in a biological cell sample in which it is desirable to reduce the presence of a group of RNAs (that do not hybridize to the NSR and/or anti-NSR primers) from an amplified sample, such as, for example, highly expressed RNAs (e.g., ribosomal RNAs). In some embodiments, the methods of the invention may be used to reduce the amount of a group of nucleic acid molecules that do not hybridize to the NSR primers and/or anti-NSR primers in amplified nucleic acid derived from an RNA sample by at least 2 fold up to 1000 fold, such as at least 10 fold, 50 fold, 100 fold, 500 fold or greater, in comparison to the amount of amplified nucleic acid molecules that do hybridize to the NSR and/or anti-NSR primers.

Populations of oligonucleotides used to practice the method of this aspect of the invention are selected from within a larger population of oligonucleotides, wherein the first population of oligonucleotides is selected based on its ability to hybridize under defined conditions to a target RNA population but not hybridize under the defined conditions to a non-target RNA population and the first population of oligonucleotides

comprises all possible oligonucleotides having a length of 6 nucleotides, 7 nucleotides, or 8 nucleotides.

The second population of oligonucleotides is selected based on its ability to hybridize under defined conditions to a target first strand cDNA population, but not hybridize under the defined conditions to a non-target first strand cDNA population and the second population of oligonucleotides comprises all possible oligonucleotides having a length of 6 nucleotides, 7 nucleotides, or 8 nucleotides. In one embodiment, the second population of oligonucleotides may be generated by synthesizing the reverse complement of the sequence of the first population of oligonucleotides.

Composition of First Population of Oligonucleotides. In some embodiments, the first population of oligonucleotides includes all possible oligonucleotides having a length of 6 nucleotides or 7 nucleotides or 8 nucleotides. The first population of oligonucleotides may include only all possible oligonucleotides having a length of 6 nucleotides or all possible oligonucleotides having a length of 7 nucleotides or all possible oligonucleotides having a length of 8 nucleotides. Optionally, the first population of oligonucleotides may include other oligonucleotides in addition to all possible oligonucleotides having a length of 6 nucleotides or all possible oligonucleotides having a length of 7 nucleotides or all possible oligonucleotides having a length of 8 nucleotides. Typically, each member of the first population of oligonucleotides is no more than 30 nucleotides long.

Sequences of First Population of Oligonucleotides. There are 4,096 possible oligonucleotides having a length of 6 nucleotides, 16,384 possible oligonucleotides having a length of 7 nucleotides, and 65,536 possible oligonucleotides having a length of 8 nucleotides. The sequences of the oligonucleotides that constitute the population of oligonucleotides can readily be generated by a computer program such as Microsoft Word®.

Selection of Subpopulation of First Oligonucleotides. The subpopulation of first oligonucleotides is selected from the population of oligonucleotides based on the ability of the members of the subpopulation of first oligonucleotides to hybridize under defined conditions to a population of target nucleic acids but not hybridize under the same defined conditions to a non-target population. A sample of amplified includes target nucleic acid molecules (e.g., RNA or DNA molecules) that are to be amplified (e.g.,

using reverse transcription) and also includes non-target nucleic acid molecules that are not to be amplified. The subpopulation of first oligonucleotides is made up of oligonucleotides that each hybridize under defined conditions to target sequences distributed throughout the population of the nucleic acid molecules that are to be amplified but that do not hybridize under the same defined conditions to most (or any) of the non-target nucleic acid molecules that are not to be amplified. The subpopulation of first oligonucleotides hybridizes under defined conditions to target nucleic acid sequences other than those that have been intentionally avoided (non-target sequences).

For example, the cell sample may include a population of all mRNA molecules expressed in mammalian cells including many ribosomal RNA molecules (e.g., 5S, 18S, and 28S ribosomal RNAs) and mitochondrial rRNA molecules (e.g., 12S and 16S ribosomal RNAs). It is typically undesirable to amplify the ribosomal RNAs. For example, in gene expression experiments that analyze expression of genes in cells, amplification of numerous copies of abundant ribosomal RNAs may obscure subtle changes in the levels of less abundant mRNAs. Consequently, in the practice of the present invention, a subpopulation of first oligonucleotides is selected that does not hybridize under defined conditions to most (or any) non-target ribosomal RNAs but that does hybridize under the same defined conditions to most (preferably all) of the other target mRNA molecules expressed in the cells.

In order to select a subpopulation of first oligonucleotides that hybridizes under defined conditions to a target nucleic acid population but does not hybridize under the defined conditions to a non-target nucleic acid population, it is necessary to know the complete or substantially complete nucleic acid sequences of the member(s) of the non-target nucleic acid population. Thus, for example, it is necessary to know the nucleic acid sequences of the 5S, 18S, and 28S ribosomal RNAs (or a representative member of each of the foregoing classes of ribosomal RNA) and the nucleic acid sequences of the 12S and 16S ribosomal mitochondrial RNAs. The sequences for the ribosomal RNAs for the mammalian species from which the cell sample is obtained can be found in a publically accessible database. For example, the NCBI Genbank identifiers are provided in TABLE 1 for human 12S, 16S, 18S, and 28S ribosomal RNA, as accessed on September 5, 2007.

A suitable software program is then used to compare the sequences of all of the oligonucleotides in the population of first oligonucleotides (e.g., the population of all possible 6 nucleic acid oligonucleotides) to the sequences of the ribosomal RNAs to determine which of the oligonucleotides will hybridize to any portion of the ribosomal RNAs under defined hybridization conditions. Only the oligonucleotides that do not hybridize to any portion of the ribosomal RNAs under defined hybridization conditions are selected. Perl script may easily be written that permits comparison of nucleic acid sequences and identification of sequences that hybridize to each other under defined hybridization conditions.

Thus, for example, as described more fully in Example 1, the subpopulation of all possible 6 nucleic acid oligonucleotides that were not exactly complementary to any portion of any ribosomal RNA sequence was identified. In general, the subpopulation of oligonucleotides (that hybridizes under defined conditions to a target nucleic acid population but does not hybridize under the defined conditions to a non- target nucleic acid population) must contain enough different oligonucleotide sequences to hybridize to all or substantially all nucleic acid molecules in the RNA sample. Example 1 herein shows that the population of oligonucleotides having the nucleic acid sequences set forth in SEQ ID NOS:1-749 hybridizes to all or substantially all nucleic acid sequences within a population of gene transcripts stored in the publicly accessible database called RefSeq.

Additional Defined Nucleic Acid Sequence Portions. The selected subpopulation of first oligonucleotides (e.g., SEQ ID NOS:1-749) can be used to prime the reverse transcription of a target population of RNA molecules to generate first strand cDNA. Alternatively, a population of first oligonucleotides can be used as primers wherein each oligonucleotide includes the sequence of one member of the selected subpopulation of oligonucleotides and also includes an additional defined nucleic acid sequence. The additional defined nucleic acid sequence is typically located 5' to the sequence of the member of the selected subpopulation of oligonucleotides. Typically, the population of oligonucleotides includes the sequences of all members of the selected subpopulation of oligonucleotides (e.g., the population of oligonucleotides can include all of the sequences set forth in SEQ ID NOS:1-749).

The additional defined nucleic acid sequence is selected so that it does not affect the hybridization specificity of the oligonucleotide to a complementary target sequence.

For example, as shown in FIGURE 1D, each first oligonucleotide can include a transcriptional promoter sequence or first primer binding site (PBS#1) located 5' to the sequence of the member of the selected subpopulation of oligonucleotides. The promoter sequence may be incorporated into the amplified nucleic acid molecules which can, therefore, be used as templates for the synthesis of RNA. Any RNA polymerase promoter sequence can be included in the defined sequence portion of the population of oligonucleotides. Representative examples include the T7 promoter (SEQ ID NO:1508), the SP6 promoter (SEQ ID NO:1509), and the T3 promoter (SEQ ID NO:1510).

In some embodiments of this aspect of the invention, as shown in FIGURE 1C, each oligonucleotide in the first population of oligonucleotides comprises a random hybridizing portion and a defined sequence located 5' to the hybridizing portion. As shown in FIGURE 1C, each first oligonucleotide can include a defined sequence comprising a primer binding site located 5' to the random hybridizing portion. The primer binding site is incorporated into the amplified nucleic acids, which can then be used as a PCR primer binding site for the generation of double-stranded amplified DNA products from the cDNA. The primer binding site may be a portion of a transcriptional promoter sequence.

Sequences of Second Population of Oligonucleotides. The selection process for the second population of oligonucleotides is similar to the process described above for the selection of the first population of oligonucleotides with the difference being that the hybridizing portion consisting of 6 nucleotides, 7 nucleotides, or 8 nucleotides is selected to hybridize to the first strand cDNA reverse transcribed from the target RNA under defined conditions and not hybridize to the first strand cDNA reverse transcribed from the non-target RNA under defined conditions. The second population of oligonucleotides can be selected using the methods described above, for example, using the publicly available sequences for ribosomal RNA. The second population of oligonucleotides can also be generated as the reverse-complement of the first population of oligonucleotides (anti-NSR).

Thus, for example, as described more fully in Example 1, the second population was selected based on all possible 6 nucleic acid oligonucleotides that were not exactly complementary to any portion of any ribosomal RNA sequence was identified. Example 1 herein shows that the population of oligonucleotides having the nucleic acid

sequences set forth in SEQ ID NOS:1-749 hybridizes to all or substantially all nucleic acid sequences within a population of gene transcripts stored in the publicly accessible database called RefSeq. A second population SEQ ID NOS:750-1498 (anti-NSR) was then generated that was the reverse complement of the first population of oligonucleotides (SEQ ID NOS:1-749, NSR).

Additional Defined Nucleic Acid Sequence Portions. The selected subpopulation of second oligonucleotides (e.g., SEQ ID NOS:750-1498) can be used to prime the second strand cDNA synthesis of a target population of first strand cDNA molecules. Alternatively, a population of second oligonucleotides can be used as primers wherein each oligonucleotide includes the sequence of one member of the selected subpopulation of oligonucleotides and also includes an additional defined nucleic acid sequence. The additional defined nucleic acid sequence is typically located 5' to the sequence of the member of the selected subpopulation of oligonucleotides. Typically, the population of oligonucleotides includes the sequences of all members of the selected subpopulation of oligonucleotides (e.g., the population of oligonucleotides can include all of the sequences set forth in SEQ ID NOS:750-1498).

The additional defined nucleic acid sequence is selected so that it does not affect the hybridization specificity of the oligonucleotide to a complementary target sequence. For example, as shown in FIGURE 1D, each first oligonucleotide can include a transcriptional promoter sequence or second primer binding site (PBS#2) located 5' to the sequence of the member of the selected subpopulation of oligonucleotides. The promoter sequence may be incorporated into the amplified nucleic acid molecules that can, therefore, be used as templates for the synthesis of RNA. Any RNA polymerase promoter sequence can be included in the defined sequence portion of the population of oligonucleotides. Representative examples include the T7 promoter (SEQ ID NO:1508), the SP6 promoter (SEQ ID NO:1509), and the T3 promoter (SEQ ID NO:1510).

In another aspect, the present invention provides a population of first oligonucleotides wherein each oligonucleotide of the population includes (a) a sequence of a 6 nucleic acid oligonucleotide that is a member of a subpopulation of oligonucleotides (SEQ ID NOS:1-749), wherein the subpopulation of oligonucleotides hybridizes to all or substantially all RNAs expressed in mammalian cells but does not hybridize to ribosomal RNAs; and (b) a primer binding site (PBS#1) sequence

(SEQ ID NO:1499) located 5' to the sequence of the 6 nucleic acid oligonucleotide. In one embodiment, the population of first oligonucleotides includes all of the 6 nucleotide sequences set forth in SEQ ID NOS:1-749. In another embodiment, the population of first oligonucleotides includes at least 10% (such as at least 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, or 99%) of the 6 nucleotide sequences set forth in SEQ ID NOS:1-749.

Optionally, a spacer portion is located between the defined sequence portion and the hybridizing portion in the first population of oligonucleotides. The spacer portion is typically from 1 to 12 nucleotides long (e.g., from 1 to 6 nucleotides long) and can include any combination of random nucleotides (N=A, C, T, or G). The spacer portion can, for example, be composed of a random selection of nucleotides. All or part of the spacer portion may or may not hybridize to the same target nucleic acid sequence as the hybridizing portion. If all or part of the spacer portion hybridizes to the same target nucleic acid sequence as the hybridizing portion, then the effect is to enhance the efficiency of cDNA synthesis primed by the oligonucleotide that includes the hybridizing portion and the hybridizing spacer portion. In some embodiments, the population of first oligonucleotides further comprises a spacer region consisting of from 1 to 10 random nucleotides (A, C, T, or G) located between the primer binding site and the hybridizing portion. In another embodiment, the population of first oligonucleotides includes all of the six nucleotide sequences set forth in SEQ ID NOS:1-749 wherein each nucleotide sequence further comprises at least one spacer nucleotide at the 5' end.

In another aspect, the present invention provides a population of second oligonucleotides wherein each oligonucleotide of the population includes (a) a sequence of a 6 nucleic acid oligonucleotide that is a member of a subpopulation of oligonucleotides (SEQ ID NOS:750-1498), wherein the subpopulation of oligonucleotides hybridizes to all or substantially all first strand cDNAs reverse transcribed from RNAs expressed in mammalian cells but does not hybridize to first strand cDNAs reverse transcribed from ribosomal RNAs; and (b) a primer binding site (PBS#2) sequence (SEQ ID NO:1500) located 5' to the sequence of the 6 nucleic acid oligonucleotide. In one embodiment, the population of first oligonucleotides includes all of the 6 nucleotide sequences set forth in SEQ ID NOS:750-1498. In another embodiment, the population of first oligonucleotides includes at least 10% (such as at

least 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, or 99%) of the 6 nucleotide sequences set forth in SEQ ID NOS:750-1498.

Optionally, a spacer portion is located between the defined sequence portion and the hybridizing portion in the second population of oligonucleotides. The spacer portion is typically from 1 to 12 nucleotides long (e.g., from 1 to 6 nucleotides long) and can include any combination of random nucleotides (N=A, C, T, or G). The spacer portion can, for example, be composed of a random selection of nucleotides. All or part of the spacer portion may or may not hybridize to the same target nucleic acid sequence as the hybridizing portion. If all or part of the spacer portion hybridizes to the same target nucleic acid sequence as the hybridizing portion, then the effect is to enhance the efficiency of cDNA synthesis primed by the oligonucleotide that includes the hybridizing portion and the hybridizing spacer portion. In some embodiments, the population of first oligonucleotides further comprises a spacer region consisting of from 1 to 10 random nucleotides (A, C, T, or G) located between the primer binding site and the hybridizing portion. In another embodiment, the population of first oligonucleotides includes all of the six nucleotide sequences set forth in SEQ ID NOS:750-1498, wherein each nucleotide sequence further comprises at least one spacer nucleotide at the 5' end.

In some embodiments, the defined sequence portion of the first population of oligonucleotides and the defined sequence portion of the second population of oligonucleotides each consists of a length ranging from at least 10 nucleotides up to 30 nucleotides, such as from 10 to 12 nucleotides, from 10 to 14 nucleotides, from 10 to 16 nucleotides, from 10 to 18 nucleotides, and from 10 to 20 nucleotides. In some embodiments, the defined sequence portion of each of the first and second population of oligonucleotides consists of 10 nucleotides, wherein the defined sequence portion comprises a PCR primer binding site, and wherein at least 8 consecutive nucleotides in the PCR binding site in each member of the first population of oligonucleotides have an identical sequence with at least 8 nucleotides in the PCR binding site in each member of the second population of oligonucleotides. In a further embodiment, the defined sequence portion of each of the first and second population of oligonucleotides consists of 10 nucleotides, wherein the defined sequence portion comprises a PCR primer binding site, and wherein at least 8 consecutive nucleotides in the PCR binding site in each member of the first population of oligonucleotides have an identical sequence with at

least 8 nucleotides in the PCR binding site in each member of the second population of oligonucleotides, and wherein the remaining two nucleotides at the 3' end of the defined sequence portion in the first population of oligonucleotides are different (e.g., C, T) from the two nucleotides at the 3' end of the defined sequence portion in the second population of oligonucleotides (e.g., G, A), thereby allowing for the identification of the transcript strand (sense or antisense) after sequence analysis prior to alignment of the sequence reads.

In a further embodiment, hybrid RNA/DNA oligonucleotides are provided wherein the defined sequence portion of the first population of oligonucleotides comprises an RNA portion and a DNA portion, wherein the RNA portion is 5' with respect to the DNA portion. In one embodiment, the 5' RNA portion of the hybrid primer consists of at least 11 RNA nucleotide defined sequence portions and the 3' DNA portion of the hybrid primer consists of at least three DNA nucleotides. In a specific embodiment, the hybrid RNA/DNA oligonucleotides comprise SEQ ID NO:1558 covalently attached to the 5' end of the NSR primers (SEQ ID NOS:1-749). The cDNA generated using the hybrid RNA/DNA oligonucleotides may be used as a template for generating single-stranded amplified DNA using the methods described in U.S. Patent No. 6,946,251, hereby incorporated by reference, as further described in Example 6.

For example, a first population of oligonucleotides for first strand cDNA synthesis comprising a hybrid RNA/DNA defined sequence portion (SEQ ID NO:1558) and a hybridizing portion (SEQ ID NOS:1-749) forms the basis for replication of the target nucleic acid molecules in template RNA. The first population of oligonucleotides comprising the hybrid RNA/DNA primer portion hybridize to the target RNA in the RNA templates and the hybrid RNA/DNA primer is extended by an RNA-dependent DNA polymerase to form a first primer extension product (first strand cDNA). After cleavage of the template RNA, a second strand cDNA is formed in a complex with the first primer extension product. In accordance with this embodiment, the double-stranded complex of first and second primer extension products is composed of an RNA/DNA hybrid at one end due to the presence of the hybrid primer in the first primer extension product. The double-stranded complex is then used to generate single-stranded DNA amplification products with an agent such as an enzyme which cleaves RNA from the RNA/DNA hybrid (such as RNAseH) which cleaves the RNA sequence from the hybrid, leaving a

sequence on the second primer extension product available for binding by another hybrid primer, which may or may not be the same as the first hybrid primer. Another first primer extension product is produced by a highly processive DNA polymerase, such as phi29, which displaces the previously bound cleaved first primer extension product, resulting in displaced cleaved first primer extension product.

In an alternative embodiment, a double-stranded complex for single-stranded DNA amplification is generated by modifying a double-stranded cDNA product (all DNA), generated using either random primers or NSR and anti-NSR primers, or a combination thereof. The double-stranded cDNA product is denatured and an RNA/DNA hybrid primer is annealed to a pre-determined primer sequence at the 3' end portion of the second strand cDNA. The DNA portion of the hybrid primer is then extended using reverse transcriptase to form a double-stranded complex with an RNA hybrid portion. The double-stranded complex is then used as a template for single-stranded DNA amplification by first treating with RNAseH to remove the RNA portion of the complex, adding the RNA/DNA hybrid primer, and adding a highly processive DNA polymerase, such as phi29 to generate single-stranded DNA amplification products.

Hybridization Conditions. In the practice of the present invention, a population of first oligonucleotides is selected from a population of oligonucleotides based on the ability of the members of the population of oligonucleotides to hybridize under defined conditions to a target nucleic acid population but not hybridize under the same defined conditions to a non-target nucleic acid population. The defined hybridization conditions permit the first oligonucleotides to specifically hybridize to all nucleic acid molecules that are present in the sample except for ribosomal RNAs. Typically, hybridization conditions are no more than 25°C to 30°C (for example, 10°C) below the melting temperature (Tm) of the native duplex. Tm for nucleic acid molecules greater than about 100 bases can be calculated by the formula $Tm = 81.5 + 0.41\%(G+C) - \log(Na+)$, wherein (G+C) is the guanosine and cytosine content of the nucleic acid molecule. For oligonucleotide molecules less than 100 bases in length, exemplary hybridization conditions are 5°C to 10°C below Tm. On average, the Tm of a short oligonucleotide duplex is reduced by approximately (500/oligonucleotide length)°C. In some embodiments of the present invention, the hybridization temperature is in the range of

from 40°C to 50°C. The appropriate hybridization conditions may also be identified empirically without undue experimentation.

In one embodiment of the present invention, the first population of oligonucleotides hybridizes to a target population of nucleic acid molecules at a temperature of about 40°C.

In one embodiment of the present invention, the second population of oligonucleotides hybridizes to a target population of nucleic acid molecules in a population of single-stranded primer extension products at a temperature of about 37°C.

Amplification Conditions. In the practice of the present invention, the amplification of the first subpopulation of a target nucleic acid population occurs under defined amplification conditions. Hybridization conditions can be chosen as described, *supra*. Typically, the defined amplification conditions include first strand cDNA synthesis using a reverse transcriptase enzyme. The reverse transcription reaction is performed in the presence of defined concentrations of deoxynucleotide triphosphates (dNTPs). In some embodiments, the dNTP concentration is in a range from about 1000 to about 2000 microMolar in order to enrich the amplified product for target genes, as described in co-pending U.S. Patent Application Serial No. 11/589,322, filed October 27, 2006, incorporated herein by reference.

Composition and Synthesis of Oligonucleotides. An oligonucleotide primer useful in the practice of the present invention can be DNA, RNA, PNA, chimeric mixtures, or derivatives or modified versions thereof, as long as it is still capable of priming the desired reaction. The oligonucleotide primer can be modified at the base moiety, sugar moiety, or phosphate backbone and may include other appending groups or labels, so long as it is still capable of priming the desired amplification reaction.

For example, an oligonucleotide primer may comprise at least one modified base moiety that is selected from the group including but not limited to 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-

D-mannosylqueosine, 5′-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, and 2,6-diaminopurine.

Again by way of example, an oligonucleotide primer can include at least one modified sugar moiety selected from the group including, but not limited to, arabinose, 2-fluoroarabinose, xylulose, and hexose.

By way of further example, an oligonucleotide primer can include at least one modified phosphate backbone selected from the group consisting of a phosphorothioate, a phosphorodithioate, a phosphoramidothioate, a phosphoramidate, a phosphordiamidate, a methylphosphonate, an alkyl phosphotriester, and a formacetal or analog thereof.

An oligonucleotide primer for use in the methods of the present invention may be derived by cleavage of a larger nucleic acid fragment using non-specific nucleic acid cleaving chemicals or enzymes, or site-specific restriction endonucleases, or by synthesis by standard methods known in the art, for example, by use of an automated DNA synthesizer (such as are commercially available from Biosearch, Applied Biosystems, etc.) and standard phosphoramidite chemistry. As examples, phosphorothioate oligonucleotides may be synthesized by the method of Stein et al. (*Nucl. Acids Res.* *16*:3209-3221, 1988) and methylphosphonate oligonucleotides can be prepared by use of controlled pore glass polymer supports (Sarin et al., *Proc. Natl. Acad. Sci. U.S.A.* *85*:7448-7451, 1988).

Once the desired oligonucleotide is synthesized, it is cleaved from the solid support on which it was synthesized and treated by methods known in the art to remove any protecting groups present. The oligonucleotide may then be purified by any method known in the art, including extraction and gel purification. The concentration and purity of the oligonucleotide may be determined by examining an oligonucleotide that has been separated on an acrylamide gel or by measuring the optical density at 260 nm in a spectrophotometer.

The methods of this aspect of the invention can be used, for example, to selectively amplify coding regions of mRNAs, introns, alternatively spliced forms of a gene, and non-coding RNAs that regulate gene expression.

In another aspect, the present invention provides populations of oligonucleotides comprising at least 10% (such as at least 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, or 99%) of the nucleic acid sequences set forth in SEQ ID NOS:1-749. These oligonucleotides (SEQ ID NOS:1-749) can be used, for example, to prime the first strand synthesis of cDNA molecules complementary to RNA molecules isolated from a mammalian subject without priming the first strand synthesis of cDNA molecules complementary to ribosomal RNA molecules. Indeed, these oligonucleotides (SEQ ID NOS:1-749) can be used, for example, to prime the synthesis of cDNA using any population of RNA molecules as templates, without amplifying a significant amount of ribosomal RNAs or mitochondrial ribosomal RNAs. For example, the present invention provides populations of oligonucleotides wherein a defined sequence portion, such as a transcriptional promoter such as the T7 promoter (SEQ ID NO:1508), or a primer binding site (PBS#1) (SEQ ID NO:1499) is located 5' to a member of the population of oligonucleotides having the sequences set forth in SEQ ID NOS:1-749. Thus, in some embodiments, the present invention provides populations of oligonucleotides wherein each oligonucleotide consists of the T7 promoter (SEQ ID NO:1508) located 5' to a different member of the population of oligonucleotides having the sequences set forth in SEQ ID NOS:1-749. In some embodiments, the present invention provides populations of oligonucleotides wherein each oligonucleotide consists of the primer binding site SEQ ID NO:1499 and a random spacer nucleotide (A, C, T, or G) is located 5' to a different member of the population of oligonucleotides having the sequences set forth in SEQ ID NOS:1-749. In some embodiments, the population of oligonucleotides includes at least 10% (such as 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 99%) of the six nucleotide sequences set forth in SEQ ID NOS:1-749.

In another aspect, the present invention provides populations of oligonucleotides comprising at least 10% (such as at least 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, or 99%) of the nucleic acid sequences set forth in SEQ ID NOS:750-1498. These oligonucleotides (SEQ ID NOS:750-1498) can be used, for example, to prime the second strand synthesis of single-stranded primer extension products complementary to RNA molecules isolated from a mammalian subject without priming the second strand synthesis of cDNA molecules complementary to ribosomal RNA molecules. Indeed, these oligonucleotides (SEQ ID NOS:750-1498) can be used, for example, to prime the

synthesis second strand cDNA using any population of single stranded primer extension molecules as templates, without amplifying a significant amount of single-stranded primer extension molecules that are complementary to ribosomal RNAs or mitochondrial ribosomal RNAs. For example, the present invention provides populations of oligonucleotides wherein a defined sequence portion, such as a transcriptional promoter such as the T7 promoter (SEQ ID NO:1508), or a primer binding site (PBS#2) (SEQ ID NO:1500) is located 5' to a member of the population of oligonucleotides having the sequences set forth in SEQ ID NOS:750-1498. Thus, in some embodiments, the present invention provides populations of oligonucleotides wherein each oligonucleotide consists of the T7 promoter (SEQ ID NO:1508) located 5' to a different member of the population of oligonucleotides having the sequences set forth in SEQ ID NOS:750-1498. In some embodiments, the present invention provides populations of oligonucleotides wherein each oligonucleotide consists of the primer binding site (PBS#2) SEQ ID NO:1500 and a random spacer nucleotide (A, C, T, or G) is located 5' to a different member of the population of oligonucleotides having the sequences set forth in SEQ ID NOS:750-1498. In some embodiments, the population of oligonucleotides includes at least 10% (such as 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 99%) of the six nucleotide sequences set forth in SEQ ID NOS:750-1498.

In another aspect, the present invention provides a reagent for selectively synthesizing single-stranded primer extension products (first strand cDNA) from a population of RNA template molecules. The reagent can be used, for example, to prime the synthesis of first strand cDNA molecules complementary to target RNA template molecules in a sample isolated from a mammalian subject without priming the synthesis of first strand cDNA molecules complementary to ribosomal RNA molecules. The reagent of the present invention comprises a population of oligonucleotides comprising at least 10% of the nucleic acid sequences set forth in SEQ ID NOS:1-749. In some embodiments, the present invention provides a reagent comprising a population of oligonucleotides that includes at least 10% (such as 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, or 99%) of the six nucleotide sequences set forth in SEQ ID NOS:1-749. In some embodiments, the population of oligonucleotides is selected to hybridize to substantially all nucleic acid molecules that are present in a sample except for ribosomal RNAs and mitochondrial rRNAs. In other embodiments, the population of

oligonucleotides is selected to hybridize to a subset of nucleic acid molecules that are present in a sample, wherein the subset of nucleic acid molecules does not include ribosomal RNAs.

In another aspect, the present invention provides a reagent for selectively synthesizing double-stranded cDNA from a population of single-stranded primer extension products (first strand cDNA). The reagent can be used, for example, to prime the synthesis of second strand cDNA molecules that are complementary to target RNA template molecules in a sample isolated from a mammalian subject without priming the synthesis of second-strand cDNA molecules complementary to ribosomal RNA molecules. The reagent in accordance with this aspect of the invention may be used to prime the synthesis of first strand cDNA generated using random primers, or may be used to prime the synthesis of first strand cDNA generated using NSR primers, such as SEQ ID NO:1-749, in order to provide an additional step of selectivity of target molecules. The reagent according to this aspect of the present invention comprises a population of oligonucleotides comprising at least 10% of the nucleic acid sequences set forth in SEQ ID NOS:750-1498. In some embodiments, the present invention provides a reagent comprising a population of oligonucleotides that includes at least 10% (such as 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, or 99%) of the six nucleotide sequences set forth in SEQ ID NOS:750-1498. In some embodiments, the population of oligonucleotides is selected to hybridize to substantially all first strand cDNA molecules that are present in a sample except for first strand cDNA synthesized from ribosomal RNAs and mitochondrial rRNAs. In other embodiments, the population of oligonucleotides is selected to hybridize to a subset of first strand cDNA molecules that are present in a sample, wherein the subset of first strand cDNA molecules does not include cDNA molecules synthesized from ribosomal RNAs.

In another embodiment, the present invention provides a reagent that comprises a population of oligonucleotides wherein a defined sequence portion comprising a transcriptional promoter such as the T7 promoter is located 5' to a member of the population of oligonucleotides having the sequences set forth in SEQ ID NOS:1-749. Thus in some embodiments, the present invention provides a reagent comprising populations of oligonucleotides wherein each oligonucleotide consists of the T7 promoter (SEQ ID NO:1508) located 5' to a different member of the population of oligonucleotides

having the sequences set forth in SEQ ID NOS:1-749. In another embodiment, the present invention provides a reagent that comprises a population of oligonucleotides wherein a defined sequence portion comprising a primer binding site (e.g., PBS#1) is located 5' to a member of the population of oligonucleotides having the sequences set forth in SEQ ID NOS:1-749. Thus, in some embodiments, the present invention provides a reagent comprising populations of oligonucleotides wherein each oligonucleotide consists of the primer binding site (PBS#1) (SEQ ID NO:1499) located 5' to a different member of the population of oligonucleotides having the sequences set forth as SEQ ID NOS:1-749. In some embodiments, the present invention provides a reagent the further comprises a spacer region of at least one random nucleotide located between the primer binding site and a different member of the population of oligonucleotides having the sequences set forth as SEQ ID NOS:1-749.

In another embodiment, the present invention provides a reagent that comprises a population of oligonucleotides wherein a defined sequence portion comprising a transcriptional promoter such as the T7 promoter is located 5' to a member of the population of oligonucleotides having the sequences set forth in SEQ ID NOS:750-1498. Thus in some embodiments, the present invention provides a reagent comprising populations of oligonucleotides wherein each oligonucleotide consists of the T7 promoter (SEQ ID NO:1508) located 5' to a different member of the population of oligonucleotides having the sequences set forth in SEQ ID NOS:750-1498. In another embodiment, the present invention provides a reagent that comprises a population of oligonucleotides wherein a defined sequence portion comprising a primer binding site (e.g., PBS#2) is located 5' to a member of the population of oligonucleotides having the sequences set forth in SEQ ID NOS:750-1498. Thus in some embodiments, the present invention provides a reagent comprising populations of oligonucleotides wherein each oligonucleotide consists of the primer binding site (PBS#2) (SEQ ID NO:1500) located 5' to a different member of the population of oligonucleotides having the sequences set forth as SEQ ID NOS:750-1498. In some embodiments, the present invention provides a reagent the further comprises a spacer region of at least one random nucleotide located between the primer binding site and a different member of the population of oligonucleotides having the sequences set forth as SEQ ID NOS:750-1498.

The reagents of the present invention can be provided as an aqueous solution or an aqueous solution with the water removed or a lyophilized solid.

In a further embodiment, the reagent of the present invention may include one or more of the following components for the production of double-stranded cDNA: a reverse transcriptase, a DNA polymerase, a DNA ligase, an RNase H enzyme, a Tris buffer, a potassium salt, a magnesium salt, an ammonium salt, a reducing agent, deoxynucleoside triphosphates (dNTPs), [beta]-nicotinamide adenine dinucleotide (β-NAD+), and a ribonuclease inhibitor. For example, the reagent may include components optimized for first strand cDNA synthesis, such as a reverse transcriptase with reduced RNase H activity and increased thermal stability (e.g., SuperScript™ III Reverse Transcriptase, Invitrogen), and a final concentration of dNTPs in the range of from 50 to 5000 microMolar or, more preferably, in the range of from 1000 to 2000 microMolar.

In another aspect, the present invention provides kits for selectively amplifying a target population of nucleic acid molecules within a population of RNA template molecules in a sample obtained from a mammalian subject. In some embodiments, the kits comprise (a) a first reagent that comprises a first population of oligonucleotide primers wherein a defined sequence portion such as a primer binding site (PBS#1) is located 5' to a hybridizing portion consisting of 6 nucleotides selected from all possible oligonucleotides having a length of 6 nucleotides that do not hybridize under defined conditions to the non-target population of nucleic acid molecules in the population of RNA template molecules, wherein the non-target population of nucleic acid molecules consists essentially of the most abundant nucleic acid molecules in the population of RNA template molecules, (b) a second reagent that comprises a second population of oligonucleotide primers wherein a defined sequence portion such as a primer binding site (PBS#2), is located 5' to a hybridizing portion consisting of 6 nucleotides selected from the reverse complement of the nucleotide sequence of the hybridizing portions of the first population of oligonucleotide primers, and (c) a first PCR primer that binds to the first defined sequence portion of the first population of oligonucleotides and a second PCR primer that binds to the second defined sequence portion of the second population of oligonucleotides.

In some embodiments, the first reagent comprises a member of the population of oligonucleotides having the sequences set forth in SEQ ID NOS:1-749. Thus in some

embodiments, the present invention provides kits containing a first reagent comprising a first population of oligonucleotides wherein each oligonucleotide consists of a first primer binding site (PBS#1) (SEQ ID NO:1499) located 5' to a different member of the population of oligonucleotides having the sequences set forth in SEQ ID NOS:1-749. In some embodiments, the present invention provides kits containing a second reagent comprising a second population of oligonucleotides wherein each oligonucleotide consists of a second primer binding site (PBS#2) (SEQ ID NO:1500) located 5' to a different member of the population of oligonucleotides having the sequences set forth in SEQ ID NOS:750-1498. In some embodiments, the invention provides kits containing a first PCR primer comprising at least 10 consecutive nucleotides that hybridize to the defined sequence portion in the first oligonucleotide population, and optionally comprises an additional sequence tail that does not hybridize to the first oligonucleotide population and a second PCR primer comprising at least 10 consecutive nucleotides that hybridize to the defined sequence portion in the second oligonucleotide population, and optionally comprises an additional sequence tail that does not hybridize to the second oligonucleotide population. In one embodiment, the first PCR primer consists of SEQ ID NO:1501, and the second PCR primer consists of SEQ ID NO:1502. The kits according to this embodiment are useful for producing amplified PCR products from cDNA generated using the Not-So-Random primers (SEQ ID NOS:1-749) and the anti-NSR (SEQ ID NOS:750-1498) primers of the invention.

The kits of the invention may be designed to detect any target nucleic acid population, for example, all RNAs expressed in a cell or tissue except for the most abundantly expressed RNAs, in accordance with the methods described herein. Nonlimiting examples of exemplary oligonucleotide primers include SEQ ID NOS:1-749. Nonlimiting examples of primer binding regions are set forth as SEQ ID NOS:1499 and 1500.

The spacer portion may include any combination of nucleotides including nucleotides that hybridize to the target RNA.

In certain embodiments, the kit comprises a reagent comprising oligonucleotide primers with hybridizing portions of 6, 7, or 8 nucleotides.

In certain embodiments, the kit comprises a reagent comprising a population of oligonucleotide primers that may be used to detect a plurality of mammalian mRNA targets.

In certain embodiments, the kit comprises oligonucleotides that hybridize in the temperature range of from 40ºC to 50ºC.

In another embodiment, the kit comprises a subpopulation of oligonucleotides that do not detect rRNA or mitochondrial rRNA. Exemplary oligonucleotides for use in accordance with this embodiment of the kit are provided in SEQ ID NOS:1-749 and SEQ ID NOS:750-1498.

In some embodiments, the kits comprises a reagent comprising a population of oligonucleotides comprising at least 10% (such as at least 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, or 99%) of the six nucleotide sequences set forth in SEQ ID NOS:1-749.

In some embodiments, the kits comprise a reagent comprising a population of oligonucleotides comprising at least 10% (such as at least 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, or 99%) of the six nucleotide sequences set forth in SEQ ID NOS:750-1498.

In certain embodiments, the kit includes oligonucleotides wherein the transcription promoter comprises the T7 promoter (SEQ ID NO:1508), the SP6 promoter (SEQ ID NO:1509), or the T3 promoter (SEQ ID NO:1510).

In another embodiment, the kit may comprise oligonucleotides with a spacer portion of from 1 to 12 nucleotides that comprises any combination of nucleotides.

In some embodiments of the present invention, the kit may further comprise one or more of the following components for the production of cDNA: a reverse transcriptase enzyme a DNA polymerase enzyme, a DNA ligase enzyme, an RNase H enzyme, a Tris buffer, a potassium salt (e.g., potassium chloride), a magnesium salt (e.g., magnesium chloride), an ammonium salt (e.g., ammonium sulfate), a reducing agent (e.g., dithiothreitol), deoxynucleoside triphosphates (dNTPs), [beta]-nicotinamide adenine dinucleotide (β-NAD+), and a ribonuclease inhibitor. For example, the kit may include components optimized for first strand cDNA synthesis, such as a reverse transcriptase with reduced RNase H activity and increased thermal stability (e.g., SuperScript™ III Reverse Transcriptase, Invitrogen), and a dNTP stock solution to

provide a final concentration of dNTPs in the range of from 50 to 5000 microMolar or, more preferably, in the range of from 1000 to 2000 microMolar.

In various embodiments, the kit may include a detection reagent such as SYBR green dye or BEBO dye that preferentially or exclusively binds to double-stranded DNA during a PCR amplification step. In other embodiments, the kit may include a forward and/or reverse primer that includes a fluorophore and quencher to measure the amount of the PCR amplification products.

A kit of the invention can also provide reagents for *in vitro* transcription of the amplified cDNAs. For example, in some embodiments the kit may further include one or more of the following components: a RNA polymerase enzyme, an IPPase (Inositol polyphosphate 1-phosphatase) enzyme, a transcription buffer, a Tris buffer, a sodium salt (e.g., sodium chloride), a magnesium salt (e.g., magnesium chloride), spermidine, a reducing agent (e.g., dithiothreitol), nucleoside triphosphates (ATP, CTP, GTP, UTP), and amino-allyl-UTP.

In another embodiment, the kit may include reagents for labeling the *in vitro* transcription products with Cy3 or Cy5 dye for use in hybridizing the labeled cDNA samples to microarrays.

In another embodiment, the kit may include reagents for labeling the double-stranded PCR products. For example, the kit may include reagents for incorporating a modified base, such as amino-allyl dUTP, during PCR which can later be chemically coupled to amine-reactive Cy dyes. In another example, the kit may include reagents for direct chemical linkage of Cy dyes to guanine residues for labeling PCR products.

In another embodiment, the kit may include one or more of the following reagents for sequencing the double-stranded PCR products: Taq DNA Polymerase, T4 Polynucleotide kinase, Exonuclease I (*E. coli*), sequencing primers, dNTPs, termination (deaza) mixes (mix G, mix A, mix T, mix C), DTT solution, and sequencing buffers.

The kit optionally includes instructions for using the kit in the selective amplification of mRNA targets. The kit can also be optionally provided with instructions for *in vitro* transcription of the amplified cDNA molecules and with instructions for labeling and hybridizing the *in vitro* transcription products to microarrays. The kit can also be provided with instructions for labeling and/or sequencing. The kit can also be

provided with instructions for cloning the PCR products into an expression vector to generate an expression library representative of the transcriptome of the sample at the time the sample was taken.

In another aspect, the present invention provides methods of selectively amplifying a target population of nucleic acid molecules to generate selectively amplified cDNA molecules. The method according to this aspect of the invention comprises (a) providing a first population of oligonucleotides, wherein each oligonucleotide comprises a hybridizing portion and first PCR primer binding site located 5' to the hybridizing portion, (b) annealing the first population of oligonucleotides to a sample comprising RNA templates isolated from a mammalian subject; (c) synthesizing cDNA from the RNA using a reverse transcriptase enzyme; (d) synthesizing double-stranded cDNA using a DNA polymerase and a second population of oligonucleotides, wherein each oligonucleotide comprises a hybridizing portion and a second PCR binding site located 5' to the hybridizing portion, wherein the hybridizing portion is a member of the population of oligonucleotides comprising SEQ ID NOS:750-1498; and (e) purifying the double-stranded cDNA molecules. In some embodiments, the method further comprises PCR amplifying the double-stranded cDNA molecules. FIGURE 1C shows a representative embodiment of the methods according to this aspect of the invention. As shown in FIGURE 1C, in one embodiment of the method, the first primer mixture comprises a first PCR primer binding site (PBS#1) located 5' to a hybridizing portion, wherein the hybridizing portion comprises a population of random 9mers.

In another embodiment, the present invention provides methods of selectively amplifying a target population of nucleic acid molecules to generate selectively amplified aDNA molecules. FIGURE 1D shows a representative embodiment of the methods according to this aspect of the invention. As shown in FIGURE 1D, the first primer mixture comprises a first PCR primer binding site (PBS#1) located 5' to the hybridizing portion, wherein the hybridizing portion is a member of the population of oligonucleotides comprising SEQ ID NOS:1-749. The method further comprises PCR amplifying the double-stranded cDNA using thermostable DNA polymerase, a first PCR primer that binds to the first PCR primer binding site and a second PCR primer that binds to the second PCR primer binding site to generate amplified double-stranded DNA

(aDNA). As shown in FIGURE 1D, in some embodiments, the method further comprises the step of sequencing at least a portion of the aDNA.

The methods and reagents described herein are useful in the practice of this aspect of the invention. In accordance with this aspect of the invention, any DNA-dependent DNA polymerase may be utilized to synthesize second-strand DNA molecules from the first strand cDNA. For example, the Klenow fragment of DNA Polymerase I can be utilized to synthesize the second strand DNA molecules. The synthesis of second strand DNA molecules is primed using a second population of oligonucleotides comprising a hybridizing portion consisting of from 6 to 9 nucleotides and further comprising a defined sequence portion 5' to the hybridizing portion.

The defined sequence portion may include any suitable sequence, provided that the sequence differs from the defined sequence contained in the first population of oligonucleotides. Depending on the choice of primer sequence, these defined sequence portions can be used, for example, to selectively direct DNA-dependent RNA synthesis from the second DNA molecule and/or to amplify the double-stranded cDNA template via DNA-dependent DNA synthesis.

Purification of Double-Stranded DNA Molecules. Synthesis of the second DNA molecules yields a population of double-stranded DNA molecules wherein the first DNA molecules are hybridized to the second DNA molecules, as shown in FIGURE 1D. Typically, the double-stranded DNA molecules are purified to remove substantially all nucleic acid molecules shorter than 50 base pairs, including all or substantially all (i.e., typically more than 99%) of the second primers. Preferably, the purification method selectively purifies DNA molecules that are substantially double-stranded and removes substantially all unpaired, single-stranded nucleic acid molecules such as single-stranded primers. Purification can be achieved by any art-recognized means, such as by elution through a size-fractionation column. The purified second DNA molecules can then, for example, be precipitated and redissolved in a suitable buffer for the next step of the methods of this aspect of the invention.

Amplification of the Double-Stranded DNA Molecules. In the practice of the methods of this aspect of the invention, the double-stranded DNA molecules are utilized as templates that are enzymatically amplified using the polymerase chain reaction. Any suitable primers can be used to prime the polymerase chain reaction. Typically, two

primers are used—one primer hybridizes to the defined portion of the first primer sequence (or to the complement thereof), and the other primer hybridizes to the defined portion of the second primer sequence (or to the complement thereof).

PCR Amplification Conditions. In general, the greater the number of amplification cycles during the polymerase chain reaction, the greater the amount of amplified DNA that is obtained. On the other hand, too many amplification cycles may result in randomly-biased amplification of the double-stranded DNA. Thus, in some embodiments, a desirable number of amplification cycles is between 5 and 40 amplification cycles, such as from 5 to 35, such as from 10 to 30 amplification cycles.

With regard to temperature conditions, typically a cycle comprises a melting temperature such as 95°C, an annealing temperature that varies from about 40°C to 70°C, and an elongation temperature that is typically about 72°C. With regard to the annealing temperature, in some embodiments the annealing temperature is from about 55°C to 65°C, more preferably about 60°C.

In one embodiment, amplification conditions for use in this aspect of the invention comprise 10 cycles of (95°C, 30 sec; 60°C, 30 sec; 72°C, 60 sec) then 20 cycles of (95°C, 30 sec; 60°C, 30 sec, 72°C, 60 sec (+ 10 sec added to the elongation step with each cycle)).

With regard to PCR reaction components for use in the methods of this aspect of the invention, dNTPs are typically present in the reaction in a range from 50 μM to 2000 μM dNTPs and, more preferably, from 800 to 1000 μM. $MgCl_2$ is typically present in the reaction in a range from 0.25 mM to 10 mM, and more preferably about 4 mM. The forward and reverse PCR primers are typically present in the reaction from about 50 nM to 2000 nM, and more preferably present at a concentration of about 1000 nM.

DNA Labeling. Optionally, the amplified DNA molecules can be labeled with a dye molecule to facilitate use as a probe in a hybridization experiment, such as a probe used to screen a DNA chip. Any suitable dye molecules can be utilized, such as fluorophores and chemiluminescers. An exemplary method for attaching the dye molecules to the amplified DNA molecules is provided in Example 5.

The methods according this aspect of the invention may be used, for example, for transcriptome profiling in a biological sample containing total RNA. In some embodiments, the amplified aDNA generated from cDNA using NSR priming in the first

strand cDNA and anti-NSR priming in the second-strand synthesis produced in accordance with the methods of this aspect of the invention is labeled for use in gene expression experiments, thereby providing a hybridization based reagent that typically produces a lower level of background than amplified RNA generated from NSR-primed cDNA.

In some embodiments of this aspect of the invention, the defined sequence portion of the first and/or second primer binding regions further includes one or more restriction enzyme sites, thereby generating a population of amplified double-stranded DNA products having one or more restriction enzyme sites flanking the amplified portions. These amplified products may be used directly for sequence analysis or may be released by digestion with restriction enzymes and subcloned into any desired vector, such as an expression vector for further analysis. Sequence analysis of the PCR products may be carried out using any DNA sequencing method, such as, for example, the dideoxy chain termination method of Sanger, dye-terminator sequencing methods, or a high throughput sequencing method as described in U.S. Patent No. 7,232,656 (Solexa), hereby incorporated by reference.

In another aspect, the invention provides a population of selectively amplified nucleic acid molecules comprising a representation of a target population of nucleic acid molecules within a population of RNA template molecules is a sample isolated from a mammalian subject, each amplified nucleic acid molecule comprising: a 5' defined sequence portion flanking a member of the population of amplified nucleic acid sequences, and a 3' defined sequence, wherein the population of selectively amplified sequences comprises amplified nucleic acid sequence corresponding to a target RNA molecule expressed in the mammalian subject, and is characterized by having the following properties with reference to the particular mammalian species: (a) having greater than 75% poly-adenylated and non-polyadenylated transcripts and having less than 10% ribosomal RNA (e.g., rRNA (18S or 28S) and mt-RNA).

The populations of selectively amplified nucleic acid molecules in accordance with this aspect of the invention can be generated using the methods of the invention described herein. The population of selectively amplified nucleic acid molecules may be cloned into an expression vector to generate a library. Alternatively, the population of selectively amplified nucleic acid molecules may be immobilized on a substrate to make

a microarray of the amplification products. The microarray may comprise at least one amplification product immobilized on a solid or semi-solid substrate fabricated from a material selected from the group consisting of paper, glass, ceramic, plastic, polystyrene, polypropylene, nylon, polyacrylamide, nitrocellulose, silicon, metal, and optical fiber. An amplification product may 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.

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

EXAMPLE 1

This Example describes the selection of a first population (Not-So-Random, "NSR") of 749 6-mer oligonucleotides (SEQ ID NOS:1-749) that hybridizes to all or substantially all RNA molecules expressed in mammalian cells but that does not hybridize to nuclear ribosomal RNA (18S and 28S rRNA) or mitochondrial ribosomal RNA (12S and 16S mt-rRNA). A second population of anti-NSR oligonucleotides (SEQ ID NOS:750-1498) was also generated that is the reverse complement of the NSR oligos. The NSR oligo population may be used to prime first strand cDNA synthesis and the anti-NSR oligo population may be used to prime second strand cDNA synthesis.

Rationale:

Random 6-mers (N6) can anneal at every nucleotide position on a transcript sequence from the RefSeq database (represented as "nucleotide sequence"), as shown in FIGURE 1A. After subtracting out the 6-mers whose reverse complements are a perfect match to nuclear ribosomal RNAs (18S and 28S rRNA) and mitochondrial ribosomal RNAs (12S and 16S mt-rRNA), the remaining NSR oligonucleotides (SEQ ID NOS:1-749) show a perfect match to every 4 to 5 nucleotides on nucleic acid sequences within the RefSeq database (represented as "nucleotide sequence"), as shown in FIGURE 1B.

Methods:

All 4,096 possible 6-mer oligonucleotides were computed, wherein each nucleotide was A, T (or U), C, or G. The reverse complement of each 6-mer oligonucleotide was compared to the nucleotide sequences of 18S and 28S rRNAs, and to

the nucleotide sequences of 12S and 16S mitochondrial rRNAs, as shown below in TABLE 1.

TABLE 1: RIBOSOMAL RNA

Gene Symbol	*NCBI Reference Sequence Transcript Identifier, accessed September 5, 2007*	*Nucleotide Coordinates*
12S	Genbank Ref # bJ01415.2	nt648-1601
16S	Genbank Ref # bJ01415.2	nt1671-3229
18S	Genbank Ref # bU13369.1	nt3657-5527
28S	Genbank Ref # bU13369.1	nt7935-12969

Reverse-complement 6-mer oligonucleotides having perfect matches to any of the human nuclear rRNA transcript sequences shown in TABLE 1, (which totaled 2,781) were eliminated. The reverse complements of 749 6-mers (SEQ ID NOS:1-749) did not perfectly match any portion of the rRNA transcripts. Matches to mitochondrial rRNA were also eliminated (566), leaving a total of 749 oligo 6-mers (4096(all 6mers) - 2782(matches to euk-rRNAs) - 566(matches to mito-rRNA)) = 749 total.

The 749 6-mer oligonucleotides (SEQ ID NOS:1-749) that do not have a perfect match to any portion of the rRNA genes and mt-rRNA genes are referred to as "Not-So-Random" ("NSR") primers. Thus the population of 749 6-mers (SEQ ID NOS:1-749) is capable of amplifying all transcripts except 18S, 28S, and mitochondrial rRNA (12S and 16S).

The population of NSR oligos (SEQ ID NO:1-749) may be used to prime first strand cDNA synthesis, as described in EXAMPLE 2, which may then be followed by second strand synthesis using either random primers, or anti-NSR primers.

As further described in EXAMPLE 2, a population of anti-NSR oligos (SEQ ID NOS:750-1498) may be used to prime second strand cDNA synthesis. As shown in FIGURE 1C, first strand cDNA synthesis may be carried out using random primers, followed by second strand cDNA synthesis using anti-NSR primers. Alternatively, as shown in FIGURE 1D, first strand cDNA synthesis may be carried out using NSR primers, followed by second strand cDNA synthesis using anti-NSR primers.

Applications to Other Types of RNA Samples. For gene profiling of mammalian cells other than human (e.g., rat, mouse), a similar approach may be carried out by

subtracting out ribosomal nuclear rRNA of the genes corresponding to 18S and 28S, as well as subtracting out ribosomal mitochondrial rRNA of the genes corresponding to 12S and 16S from the respective mammalian species.

Gene profiling of plant cells may also be carried out by generating a population of Not-So-Random (NSR) primers that exclude chloroplast ribosomal RNA.

EXAMPLE 2

This Example shows that amplification of total RNA using NSR primers and anti-NSR primers selectively reduces priming of unwanted, non-target ribosomal sequences.

Methods:

To construct new primer libraries, primers were synthesized individually as follows:

A first population of NSR-6mer primers (SEQ ID NOS:1-749) and a second population of anti-NSR-6mer primers (SEQ ID NOS:750-1498) were generated as described in Example 1.

NSR for First Strand cDNA Synthesis. In some embodiments, the first primer set of NSR primers for use in first strand cDNA synthesis (SEQ ID NOS:1-749) further comprises the following 5' primer binding sequence:

> PBS#1: 5' TCCGATCTCT 3' (SEQ ID NO:1499) covalently attached at the 5' end (otherwise referred to as "tailed"),

resulting in a population of oligonucleotides having the following configuration:

> 5' PBS#1 (SEQ ID NO:1499) + NSR-6mer (SEQ ID NOS:1-749) 3'

In another embodiment, a population of oligonucleotides was generated wherein each NSR-6mer optionally included at least one spacer nucleotide (N) (where each N = A, G, C, or T) where (N) was located between the 5'PBS#1 and the NSR-6mer. The spacer region may comprise from one nucleotide up to ten or more nucleotides (N = 1 to 10), resulting in a population of oligonucleotides having the following configuration:

5' PBS#1 (SEQ ID NO:1499) + (N_{1-10}) + NSR-6mer (SEQ ID NOS:1-749) 3'

Anti-NSR for Second Strand cDNA Synthesis. In some embodiments, the population of anti-NSR-6mer primers for use in second strand cDNA synthesis (SEQ ID NOS:750-1498) further comprises the following 5' primer binding sequence:

PBS#2: 5'TCCGATCTGA 3'(SEQ ID NO:1500) covalently attached at the 5' end of the anti-NSR-6mer primers (otherwise referred to as "tailed"),

resulting in the following configuration:

5' PBS#2 (SEQ ID NO:1500) + anti-NSR-6mer (SEQ ID NOS:750-1498) 3'

In another embodiment, a population of oligonucleotides was generated wherein each anti-NSR-6mer optionally included at least one spacer nucleotide (N) (where each N = A, G, C, or T) where (N) was located between the 5'PBS#2 and the anti-NSR-6mer. The spacer region may comprise from one nucleotide up to ten or more nucleotides (N = 1 to 10), resulting in a population of oligonucleotides having the following configuration:

5' PBS#2 (SEQ ID NO:1500) + (N_{1-10}) + anti-NSR-6mer (SEQ ID NOS:750-1498) 3'

Forward and Reverse Primers (for PCR Amplification). The following forward and reverse primers were synthesized to amplify double-stranded cDNA generated using NSR-6mers tailed with PBS#1 (SEQ ID NO:1499) and anti-NSR-6mers tailed with PBS#2 (SEQ ID NO:1500).

NSR_F_SEQprimer 1: 5' $N_{(10)}$TCCGATCTCT-3' (SEQ ID NO:1501), where each N = G, A, C, or T.

NSR_R_SEQprimer 1: 5' $N_{(10)}$TCCGATCTGA-3' (SEQ ID NO:1502), where each N = G, A, C, or T.

In the embodiment described above, the 5' most region of the forward primer (SEQ ID NO:1501) and reverse primer (SEQ ID NO:1502) each include a 10mer sequence of (N) nucleotides. In another embodiment, the 5' -most region of the forward primer (SEQ ID NO:1501) and reverse primer (SEQ ID NO:1502) each include more than 10 (N) nucleotides, such as at least 20 (N) nucleotides, at least 30 (N) nucleotides, or at least 40 (N) nucleotides to facilitate DNA sequencing of the amplified PCR products.

Control Primers. The following primers were used to amplify the control reactions amplified with random primer pools:

The following primer binding sites were added to random primers:

Y4F: 5' CCACTCCATTTGTTCGTGTG 3' (SEQ ID NO:1506)

Y4R: 5' CCGAACTACCCACTTGCATT 3' (SEQ ID NO:1507)

The following primer binding sites with random primers (N=7 or N=9), or NSR primers:

Y4R-N7 (1st strand cDNA):

5' CCGAACTACCCACTTGCATTNNNNNNN 3' (SEQ ID NO:1503) [where N = A, G, C, or T]

Y4R-NSR (1st strand cDNA):

5' CCGAACTACCCACTTGCATTN 3' (SEQ ID NO:1504) covalently attached to NSR primers that include the core set of 6-mer NSR oligos with no perfect match to globin (alpha or beta), no perfect match to rRNA (18S, 28S).

Y4F-N9 (2nd strand cDNA synthesis):

5' CCACTCCATTTGTTCGTGTGNNNNNNNNN 3' (SEQ ID NO:1505) [where N= A, G, C, or T]

Y4F 5' CCACTCCATTTGTTCGTGTG 3' (SEQ ID NO:1506)

Y4R 5' CCGAACTACCCACTTGCATT 3' (SEQ ID NO:1507)

Other Optional Primer Pool Configurations. Additional primers that could be used as primer binding sites covalently attached to the NSR pool in order to add transcriptional promoters to the amplified cDNA product:

T7: 5' AATTAATACGACTCACTATAGGGAGA 3' (SEQ ID NO:1508)

SP6: 5' ATTTAGGTGACACTATAGAAGNG 3' (SEQ ID NO:1509)

T3: 5'AATTAACCCTCACTAAAGGGAGA 3' (SEQ ID NO:1510)

Primer Pool Configurations Used to Amplify RNA. Primers were synthesized individually as described above and pooled in the following configuration, then the primer pools were used to generate libraries of amplified nucleic acids from total RNA as described below.

TABLE 2: PRIMER POOL CONFIGURATIONS

Reference ID	Pool Components (includes all expressed RNA except for those listed)	Number of individual sequences in Pool	Description of Pool	SEQ ID NO:	5' Primer Binding Sequence (covalently attached)
saNSR#1 pool	NSR-6mers -(R,M,G)	510	core set of 6-mer NSR oligos with no perfect match to rRNA (18S, 28S), mt-RNA (12S, 16S) or globin (alpha or beta)	SEQ ID NO: 1-510, with a spacer (N=A, G, C, or T) located between PBS#1 and NSR-6mer	PBS#1 (SEQ ID NO:1499)
saNSR#2 pool	NSR-6mers - (G, R)	403	core set of 6-mer NSR oligos with perfect match to mt-rRNA, but not globin or rRNA	control set, (sequences not provided)	SEQ ID NO:1499
saNSR#3 pool	NSR-6mers - (M, R)	239	core set of 6-mer NSR oligos with perfect match to globins, but not mt-rRNA or rRNA	SEQ ID NO: 511-749 with a spacer (N=A, G, C, or T) located between PBS#1 and NSR-6mer	PBS#1 (SEQ ID NO:1499)
saNSR #4 pool	NSR-6mers - (R)	163	core set of 6-mer NSR oligos with perfect match to mt-rRNA and globin, but not to rRNA	control set, (sequences not shown)	SEQ ID NO:1499

Reference ID	Pool Components (includes all expressed RNA except for those listed)	Number of individual sequences in Pool	Description of Pool	SEQ ID NO:	5' Primer Binding Sequence (covalently attached)
sa-antiNSR#5 pool	anti-NSR-6mers - (R, M, G)	510	core set of 6-mer NSR oligos with no perfect match to rRNA (18S, 28S), mt-RNA (12S, 16S) or globin (alpha or beta);	SEQ ID NO: 750-1259 with a spacer (N=A, G, C, or T) located between PBS#2 and anti-NSR-6mer	PBS#2 (SEQ ID NO:1500)
sa-antiNSR#6 pool	anti-NSR-6mers - (G, R)	403	core set of 6-mer anti-NSR oligos with perfect match to mt-rRNA, but not globin or rRNA	control set, (sequences not shown)	SEQ ID NO:1500
sa-antiNSR#7 pool	anti-NSR-6mers - (M, R)	239	core set of 6-mer antiNSR oligos with perfect match to globins, but not mt-rRNA or rRNA	SEQ ID NO: 1260-1499 with a spacer (N=A, G, C, or T) located between PBS#2 and anti-NSR-6mer	PBS#2 (SEQ ID NO:1500)
sa-antiNSR#8 pool	anti-NSR-6mers - (R)	163	core set of 6-mer anti-NSR oligos with perfect match to mt-rRNA and globin, but not to rRNA	control set, (sequences not shown)	SEQ ID NO: 1500

PM= perfect match at 3'-most 6nt of primer
R=rRNA (18S or 28S)
M=mt-rRNA (12S or 16S)
G=globin (HBA1, HBA2, HBB, HBD, HBG1, HBG2)

5

TABLE 3: PRIMER SETS FOR USE IN RNA AMPLIFICATION EXPERIMENT

Reference ID	Process	Amount (µL)	Description	SEQ ID NO:
saNSR#1 pool	1st strand cDNA synthesis	510 µL total	510 µL of saNSR#1 pool only	SEQ ID NOS: 1-510, with a spacer (N=A, G, C, or T) located between PBS#1 and NSR-6mer
saNSR#1 pool + saNSR#2 pool	1st strand cDNA synthesis	913 µL total	510 µL of saNSR#1 pool combined with 403 µL of saNSR#2 pool	control set

Reference ID	Process	Amount (µL)	Description	SEQ ID NO:
saNSR#1 pool + saNSR#3 pool	1st strand cDNA synthesis	749 µL total	510 µL of saNSR#1 pool combined with 239 µL of NSR#3 pool	SEQ ID NOS: 1-749, with a spacer (N=A, G, C, or T) located between PBS#1 and NSR-6mer
saNSR#1 pool + saNSR#4 pool	1st strand cDNA synthesis	673 µL total	510 µL of saNSR#1 pool combined with 163 µL of saNSR#4 pool	control set
sa-anti-NSR#5 pool	2nd strand cDNA synthesis	510 µL total	510 µL of sa-antiNSR#5 pool only	SEQ ID NOS: 750-1259 with a spacer (N=A, G, C, or T) located between PBS#2 and anti-NSR-6mer
sa-anti-NSR#5 pool + sa-anti-NSR#6 pool	2nd strand cDNA synthesis	913 µL total	510 µL of sa-anti-NSR#5 pool combined with 403 µL of sa-anti-NSR#6 pool	control set
sa-anti-NSR#5 pool + sa-anti-NSR#7 pool	2nd strand cDNA synthesis	749 µL total	510 µL of sa-anti-NSR#5 pool combined with 239 µL of sa-anti-NSR#7 pool	SEQ ID NOS: 750-1499 with a spacer (N=A, G, C, or T) located between PBS#2 and anti-NSR-6mer
sa-anti-NSR#5 pool + sa-anti-NSR#8 pool	2nd strand cDNA synthesis	673 µL total	510 µL of sa-anti-NSR#5 pool combined with 163 µL of sa-anti-NSR#8 pool	control set

cDNA Synthesis and PCR Amplification. The protocol involved a three-step amplification approach as follows: (1) first strand cDNA was generated from RNA using reverse transcription that was primed with NSR primers comprising a first primer binding site (PBS#1) to generate NSR primed first strand cDNA; (2) second strand cDNA synthesis was primed with anti-NSR primers comprising a second primer binding site (PBS#2); and (3) the synthesized cDNA was PCR amplified using forward and reverse primers that bind to the first and second primer binding sites to generate amplified DNA (aDNA).

TABLE 4: PRIMERS USED FOR FIRST AND SECOND STRAND SYNTHESIS

Reaction ID	1st Strand Primer Pool (+Reverse Transcriptase) 100 µM	2nd Strand Primer Pool (+Klenow)	RNA Template (1 µL of 1 µg/uL Total RNA)	Method
1	saNSR#1 pool	sa-anti-NSR#5 pool	Jurkat-1	RT-PCR
2	saNSR#1 pool + saNSR#2 pool	sa-anti-NSR#5 pool + sa-anti-NSR#6 pool	Jurkat-1	RT-PCR
3	saNSR#1 pool + saNSR#3 pool	sa-anti-NSR#5 pool + sa-anti-NSR#7 pool	Jurkat-1	RT-PCR
4	saNSR#1 pool + saNSR#4 pool	sa-anti-NSR#5 pool + sa-anti-NSR#8 pool	Jurkat-1	RT-PCR
5	Y4R-NSR	Y4F-N9	Jurkat-1	RT-PCR
6	Y4R-NSR	Y4F-N9	Jurkat-1	RT-PCR
7	Y4-N7	Y4F-N9	Jurkat-1	RT-PCR
8	N8	None	Jurkat-1	RT
9	saNSR#1 pool	sa-anti-NSR#5 pool	Jurkat-2	RT-PCR
10	saNSR#1 pool + saNSR#2 pool	sa-anti-NSR#5 pool + sa-anti-NSR#6 pool	Jurkat-2	RT-PCR
11	saNSR#1 pool + saNSR#3 pool	sa-anti-NSR#5 pool + sa-anti-NSR#7 pool	Jurkat-2	RT-PCR
12	saNSR#1 pool + saNSR#4 pool	sa-anti-NSR#5 pool + sa-anti-NSR#8 pool	Jurkat-2	RT-PCR
13	Y4R-NSR	Y4F-N9	Jurkat-2	RT-PCR
14	Y4R-NSR	Y4F-N9	Jurkat-2	RT-PCR
15	Y4-N7	Y4F-N9	Jurkat-2	RT-PCR
16	N8	None	Jurkat-2	RT
17	saNSR#1 pool	sa-antiNSR#5 pool	K562	RT-PCR
18	saNSR#1 pool + saNSR#2 pool	sa-anti-NSR#5 pool + sa-anti-NSR#6 pool	K562	RT-PCR
19	saNSR#1 pool + saNSR#3 pool	sa-anti-NSR#5 pool + sa-anti-NSR#7 pool	K562	RT-PCR
20	saNSR#1 pool + saNSR#4 pool	sa-anti-NSR#5 pool + sa-anti-NSR#8 pool	K562	RT-PCR

Reaction ID	1st Strand Primer Pool (+Reverse Transcriptase) 100 µM	2nd Strand Primer Pool (+Klenow)	RNA Template (1 µL of 1 µg/uL Total RNA)	Method
21	Y4R-NSR	Y4F-N9	K562	RT-PCR
22	Y4R-NSR	Y4F-N9	K562	RT-PCR
23	Y4-N7	Y4F-N9	K562	RT-PCR
24	N8	None	K562	RT

Reaction Conditions:

Total RNA was obtained from Ambion, Inc. (Austin, Texas), for the cell lines Jurkat (T lymphocyte, ATCC No. TIB-152) and K562 (chronic myelogenous leukemia, ATCC No. CCL-243).

First Strand Reverse Transcription:

First strand reverse transcription was carried out as follows:

Combine:

- 1 µl of 1 µg/µl Jurkat total RNA template (obtained from Ambion, Inc. (Austin, Texas)).
- 2 µl of 100 µM stock NSR primer pool (as described in Table 2)
- 7 µl H_2O to a final volume of 10 µl.

Mixed and incubated at 70°C for 5 minutes, snap chilled on ice.

Added 10 µl of RT cocktail (prepared on ice) containing:

- 4 ul 5X First Strand Buffer (250 mM Tris-HCL, pH 8.3, 375 mM KCl, 15 mM $MgCl_2$)
- 1.6 µl 25 mM dNTP (high) or 1.0 ul 10 mM dNTP (low)
- 1 µl H_2O
- 1 µl 0.1 M DTT
- 1 µl RNAse OUT (Invitrogen)
- 1 µl MMLV reverse transcriptase (200 units/µl) (SuperScript III™ (SSIII), Invitrogen Corporation, Carlsbad, California)

The sample was mixed, incubated at 23°C for 10 minutes, transferred to a 40°C pre-warmed thermal cycler (to provide a "hot start"), and the sample was then incubated at 40°C for 30 minutes, 70°C for 15 minutes, and chilled to 4°C.

1 µl of RNAse H (1-4 units/µl) was then added and the sample was incubated at 37°C for 20 minutes, then heated to 95°C for 5 minutes, and snap-chilled at 4°C.

Second Strand Synthesis:

A second strand synthesis cocktail was prepared as follows:

- 10 µl 10X Klenow Buffer
- 4 µl anti-NSR Primer (100 µM)
- 5.0 µl 10 mM dNTPs
- 56.7 µl H_2O
- 0.33 µl Klenow enzyme (5 U/µl)

80 µl of the second strand synthesis cocktail was added to the 20 µl first strand template reaction mixture, mixed and incubated at 37°C for 30 minutes, then snap-chilled at 4°C.

cDNA Purification:

The resulting double-stranded cDNA was purified using Spin Cartridges obtained from Ambion (Message Amp™ II aRNA Amplification Kit, Ambion Cat #AM1751) and buffers supplied in the kit according to the manufacturer's directions. A total volume of 30 µl was eluted from the column, of which 20 µl was used for follow-on PCR.

PCR Amplification:

The following mixture was added to 1 µl of purified cDNA template (diluted 1:5):

- 10 µl 5X Roche Expand Plus PCR Buffer
- 2.5 µl 10 mM dNTPS
- 2.5 µl Forward PCR Primer (10 µM stock) (SEQ ID NO:1501)
- 2.5 µl Reverse PCR Primer (10 µM stock) (SEQ ID NO:1502)
- 0.5 µl Taq DNA polymerase enzyme
- 27 µl H_2O
- 4 µl 25 mM $MgCl_2$

PCR Amplification Conditions:

PCR Program #1:

94°C for 2 minutes

94°C for 10 seconds

8 cycles of:

- 60°C for 10 sec
- 72°C for 60 sec
- 72°C for 60 sec

94°C for 15 sec

17 cycles of:

- 60°C for 30 sec
- 72°C for 60 sec + 10 sec/cycle

72°C for 5 minutes to polish and chilled at 4°C.

PCR program #2:

94°C for 2 minutes

94°C for 10 seconds

2 cycles of:

- 40°C for 10 sec
- 72°C for 60 sec
- 72°C for 60 sec
- 94°C for 10 seconds

8 cycles of:

- 60°C for 30 sec
- 72°C for 60 sec
- 72°C for 60 sec
- 94°C for 15 sec

15 cycles of:

- 60°C for 30 sec
- 72°C for 60 sec + 10 sec/cycle

72°C for 5 minutes to polish and chilled at 4°C.

Results of cDNA Synthesis:

The results were analyzed in terms of (1) measuring amplified DNA "aDNA" yield; (2) evaluation of an aliquot of the aDNA on an agarose gel to confirm that the population of species in the cDNA was equally represented; and (3) measuring the level of amplification of selected reporter genes by qPCR (as described in Example 3).

The PCR products were analyzed on 2% agarose gels. A DNA smear between 100-1000 bp was observed for both control reactions and test conditions using the PCR amplification program #2, indicating successful cDNA synthesis of a plurality of RNA species and PCR amplification. With PCR amplification program #1, the control reactions were successful as determined by the presence of a DNA smear in the

100-100 bp range; however, none of the test conditions amplified into a DNA smear. Instead, a low molecular weight fragment was observed that likely resulted from primer dimers (unpurified PCR product). Therefore, these results indicate that low temperature annealing (40°C) is important for PCR amplification with short (10 nt) amplification tails.

It was also determined that high dNTP concentration (25 mM) during first strand cDNA synthesis increased specificity of the cDNA product as compared to low dNTP concentration (10 mM) dNTP (data not shown).

It was further determined that RNAse H treatment reduced the amount of contamination from amplified rRNA if the NSR primer pool was used only for first strand cDNA synthesis followed by random primed second strand synthesis. However, when NSR primers were used to prime the first strand synthesis, followed by the use of anti-NSR primers to prime the second strand synthesis, then RNAse treatment was not found to affect specificity of the resulting cDNA product. Although not important for increasing specificity, RNAse may be added to second strand cDNA synthesis using anti-NSR primers to improve efficiency of the reaction by making the cDNA more available as a template during the Klenow reaction.

In summary, it was found that the use of anti-NSR primers during second strand synthesis provided several unexpected advantages for selective amplification of target nucleic acid molecules. For example, it was unexpectedly found that the magnitude of rRNA depletion during second strand synthesis using anti-NSR primers was nearly identical to the magnitude of rRNA depletion observed using NSR primers during reverse transcription. In addition, it was an unexpected result that priming specificity during second strand synthesis was achieved under standard reaction conditions using Klenow enzyme. These results indicate that short oligonucleotides can be used to specifically prime DNA synthesis using a variety of polymerases and nucleic acid templates, however, the reaction conditions that dictate priming specificity may be enzyme-specific.

EXAMPLE 3

This Example shows that the 749 NSR 6-mers (SEQ ID NOS:1-749) (that each have PBS#1 (SEQ ID NO:1499 plus N spacer) covalently attached at the 5' end) for first strand cDNA synthesis followed by the 749 anti-NSR 6-mers (SEQ ID NOS:750-1498) (that each have PBS#2 (SEQ ID NO:1500 plus N spacer) covalently attached at the

5' end) prime the amplification of a substantial fraction of the transcriptome present in a sample containing total RNA.

Methods:

Following PCR amplification as described in Example 2, each PCR reaction was purified using the Qiagen MinElute spin column. The column was washed with 80% ethanol and eluted with 20 μL of elution buffer. The yield was quantitated with UV/VIS spectrometer using the NanoDrop instrument. Samples were then diluted and characterized by quantitative PCR (qPCR) using the following assays:

Duplicate measurements of 2 μl of cDNA were made in 10 μl final reaction volumes by quantitative PCR (qPCR) in a 384-well optical PCR plate using a 7900 HT PCR instrument (Applied Biosystems, Foster City, CA). qPCR was performed using ABI TaqMan® assays using the probes shown below in TABLE 5 and TABLE 6 using the manufacturer's recommended conditions.

TABLE 5: REPORTER GENE ASSAYS FOR JURKAT CELLS

Target	ABI Assay probe	Forward Primer	Reverse Primer	FAM reporter primer
STMN1 stathmin 1/ oncoprotein 18	Hs01027516_g1	Not Relevant (NR)	NR	NR
PPIA peptidylprolyl isomerase A (cyclophilin A)	Hs99999904_m1	NR	NR	NR
EIF3S3 eukaryotic translation initiation factor 3, subunit 3 gamma, 40 kDa	Hs00186779_m1	NR	NR	NR
NUCB2 nucleobindin 2	Hs00172851_m1	NR	NR	NR
SRP14 signal recognition particle 14 kDa (homologous Alu RNA binding protein)	Hs01923965_u1	NR	NR	NR
TRIM63	Hs00761590	NR	NR	NR
DBN1	Hs00365623	NR	NR	NR
CDCA7	Hs00230589_m1	NR	NR	NR
GAPDH	Hs99999905	NR	NR	NR
Actin (ACTB)	Hs99999903	NR	NR	NR
18s rRNA	Hs99999901_s1	NR	NR	NR

Target	ABI Assay probe	Forward Primer	Reverse Primer	FAM reporter primer
R28S_3-ANY	custom	GGTTCGCCCCGAGAGA (SEQ ID NO:1511)	GGACGCCGCCGGAA (SEQ ID NO:1512)	CCGCGACGCTTTCCAA (SEQ ID NO:1513)
28S.4-JUN	custom	GTAGCCAAATGCCTCGTCATC (SEQ ID NO:1514)	CAGTGGGAATCTCGTTCATCC ATT (SEQ ID NO:1515)	ATGCGCGTCACTAATTA (SEQ ID NO:1516)
28S-7-ANY	custom	CCGAAACGATCTCAACCTATT CTCA (SEQ ID NO:1517)	GCTCCACGCCAGCGA (SEQ ID NO:1518)	CCGGGCTTCTTACCC (SEQ ID NO:1519)
28S-8-ANY	custom	GCGGGTGGTAAACTCCATCTA AG (SEQ ID NO:1520)	CCCTTACGGTACTTGTTGACT ATCG (SEQ ID NO:1521)	TCGTGCCGGTATTTAG (SEQ ID NO:1522)
18S-1-ANY	custom	GGTGACCACGGGTGACG (SEQ ID NO:1523)	GGATGTGGTAGCCGTTTCTCA (SEQ ID NO:1524)	TCCCTCTCCGGAATCG (SEQ ID NO:1525)
16S-1-ANY	custom	ACCAAGCATAATATAGCAAG GACTAACC (SEQ ID NO:1526)	TGGCTCTCCTTGCAAAGTTAT TTCT (SEQ ID NO:1527)	CCTTCTGCATAATGAATTAA (SEQ ID NO:1528)
12S-1-ANY	custom	GACAAGCATCAAGCACGCA (SEQ ID NO:1529)	CTAAAGGTTAATCACTGCTGT TTCCC (SEQ ID NO:1530)	CAATGCAGCTCAAAACG (SEQ ID NO:1531)
12S-2-ANY	custom	GTCGAAGGTGGATTTAGCAGT AAAC (SEQ ID NO:1532)	TGTACGCGCTTCAGGGC (SEQ ID NO:1533)	CCTGTTCAACTAAGCACTCTA (SEQ ID NO:1534)
hs16S-2	custom	AAGCGTTCAAGCTCAACACC (SEQ ID NO:1535)	GGTCCAATTGGGTATGAGGA (SEQ ID NO:1536)	
hs16S-3	custom	GCATAAGCCTGCGTCAGATT (SEQ ID NO:1537)	GGTTGATTGTAGATATTGGGC TGT (SEQ ID NO:1538)	
hsHST1_H2AH	custom	TACCTGACCGCTGAGATCCT (SEQ ID NO:1539)	AGCTTGTTGAGCTCCTCGTC (SEQ ID NO:1540)	
hsNC_7SK	custom	GACATCTGTCACCCCATTGA (SEQ ID NO:1541)	CTCCTCTATCGGGGATGGTC (SEQ ID NO:1542)	
hsNC_7SL1	custom	GGAGTTCTGGGCTGTAGTGC (SEQ ID NO:1543)	GTTTTGACCTGCTCCGTTTC (SEQ ID NO:1544)	

Target	ABI Assay probe	Forward Primer	Reverse Primer	FAM reporter primer
hsNC_BC200	custom	GCTAAGAGGCGGGAGGATAG (SEQ ID NO:1545)	GGTTGTTGCTTTGAGGGAAG (SEQ ID NO:1546)	
hsNC_HY1	custom	GCTGGTCCGAAGGTAGTGAG (SEQ ID NO:1547)	ATGCCAGGAGAGTGGAAACT (SEQ ID NO:1548)	
hsNC_HY3	custom	TCCGAGTGCAGTGGTGTTTA (SEQ ID NO:1549)	GTGGGAGTGGAGAAGGAACA (SEQ ID NO:1550)	
hsNC_HY4	custom	GGTCCGATGGTAGTGGGTTA (SEQ ID NO:1551)	AAAAAGCCAGTCAAATTTAG CA (SEQ ID NO:1552)	
hsNC_U4B1	custom	TGGCAGTATCGTAGCCAATG (SEQ ID NO:1553)	CTGTCAAAAATTGCCAATGC (SEQ ID NO:1554)	
hsNC_U6A	custom	CGCTTCGGCAGCACATATAC (SEQ ID NO:1555)	AAAATATGGAACGCTTCACG A (SEQ ID NO:1556)	

TABLE 6: REPORTER GENE PROBES

Assay Name	REPORTER FAM	REPORTER SYBR	1/df
NUCB2	+		10
18s (Hs99999901_s1)	+		1000
18S-1	+		1000
18S-4	+		1000
28S-3	+		1000
28S-4	+		1000
28S-7	+		1000
28S-8	+		1000
12S-1	+		1000
12S-2	+		1000
16S-1	+		1000
hs16S-2		+	1000
hs16S-3		+	1000
hsHST1_H2AHfwd		+	1000
hsNC_7SKfwd		+	1000
hsNC_7SL1fwd		+	1000
NUCB2	+		10
PPIA	+		10
SRP14	+		10
STMN1	+		10
TRIM63	+		10
ACTB	+		10
CDCA7	+		10
DBN1	+		10
EIF3S3	+		10
GAPDH	+		10
hsNC_BC200fwd		+	10
hsNC_HY1fwd		+	10
hsNC_HY3fwd		+	1000
hsNC_HY4fwd		+	1000
hsNC_U4B1fwd		+	10
hsNC_U6Afwd		+	10

Following qPCR, the results table was exported to Excel (Microsoft Corp., Redmond, WA) and quantitative analysis for samples was regressed from the raw data (abundance = $10^{[(Ct-5)/-3.4]}$).

Results:

FIGURE 3A is a histogram plot on a logarithmic scale showing the relative abundance of 18S, 28S, 12S and 16S (normalized to gene and N8) for first strand cDNA synthesis generated using various NSR pools as shown in TABLE 4 as compared to

unamplified cDNA generated using random primers (N8 = 100%). As shown in FIGURE 3A, the cDNA generated using the primer pool with NSR#1 + NSR#3 (NSR-6mers that do not hybridize to mt-rRNA or rRNA) for first strand cDNA synthesis and the primer pool anti-NSR#5 and anti-NSR#7 for second strand synthesis showed a substantial reduction in abundance of rRNA (0.086% 18S; 0.673% 28S) and a reduced abundance of mt-rRNA (1.807% 12S; and 8.512% 16S) as compared to cDNA generated with random 8-mers.

FIGURE 3B graphically illustrates the relative levels of abundance of nuclear ribosomal RNA (18S or 28S) in control cDNA amplified using random primers (N7) in both first strand and second strand synthesis (N7>N7 = 100% 18S, 100% 28S) as compared to cDNA amplified using NSR-6mer primers (SEQ ID NOS:1-749) in the first strand followed by random primers (N7) in the second strand (NSR-6mer>N7 = 3.0% 18S, 3.4% 28S), and as compared to cDNA amplified using NSR-6mer primers (SEQ ID NOS:1-749) in the first strand followed by anti-NSR-6mer primers (SEQ ID NOS:750-1498) in the second strand (NSR-6mer>anti-NSR-6mer = 0.1% 18S, 0.5% 28S). The results in FIGURE 3C show a similar trend when measuring mitochondrial rRNA, with N7>N7 = 100% 12S, or 16S; NSR-6mer>N7 = 27% 12S, 20.4% 16S; and NSR-6mer>anti-NSR-6mer = 8.2% 12S, 3.5% 16S.

In order to determine if the PCR amplified aDNA generated from the cDNA synthesized using the various NSR and anti-NSR pools preserved the target gene expression profiles present in the corresponding cDNA, quantitative PCR analysis was conducted with nine randomly chosen TaqMan reagents, detecting the following genes: PPIA, SRP14, STMN1, TRIM63, ACTB, DBN1, EIFS3, GAPDH, and NUCB2. As shown in TABLE 7 and FIGURE 4A, measurable signal was measured for the nine genes assayed in both NSR and anti-NSR primed cDNA and aDNA generated therefrom (as determined from 10 µl cDNA template input).

TABLE 7: QUANTITATIVE PCR ANALYSIS

					Input Adjusted Abundance									
Sample ID	**ng/ul**	**1st strand Primer Pool (+Reverse Transcriptase)**	**2nd strand Primer Pool (+ Klenow)**	**RNA**	**NUCB2[1]**	**18S[3]**	**18S-1[2]**	**28S-3[2]**	**28S-4[2]**	**28S-7[2]**	**28S-8[2]**	**12S-1[2]**	**12S-2[2]**	**16S-1[2]**
1	76.5	saNSR.1 pool	sa.anti-NSR#5 pool	Jurkat 1	11.4	52.9	195.0	349.1	800.8	989.2	612.5	798.8	216.0	108.1
2	73.1	saNSR.1pool+2 pool	sa.anti-NSR#5 pool +sa.anti-NSR#6 pool	Jurkat 1	5.0	55.9	238.2	335.5	616.0	1066.5	715.2	1478.0	3671.0	863.7
3	72.8	saNSR.1pool+3pool	sa.anti-NSR#5pool + sa.anti-NSR#7 pool	Jurkat 1	17.6	29.2	125.6	169.3	551.5	964.3	1310.5	312.9	159.0	80.5
4	78.2	saNSR.1pool+4pool	sa.anti-NSR#5pool +sa.anti-NSR#8 pool	Jurkat 1	12.6	55.3	155.5	272.9	538.2	964.1	610.4	639.8	1041.1	787.1
5	77.1	saNSR.1	sa.anti-NSR#5 pool	Jurkat 2	11.5	51.0	183.5	331.2	922.5	1228.1	609.5	1210.9	221.1	126.6
6	46.2	saNSR.1+2	sa.anti-NSR#5 pool +sa.anti-NSR#6 pool	Jurkat 2	7.4	34.7	180.6	405.1	364.3	1560.1	410.9	1799.2	4385.0	1007.9
7	45.2	saNSR.1+3	sa.anti-NSR#5pool + sa.anti-NSR#7 pool	Jurkat 2	20.9	30.6	107.6	234.1	378.8	1581.6	771.5	310.6	276.1	142.5
8	81.7	saNSR.1+4	sa.anti-NSR#5pool +sa.anti-NSR#8 pool	Jurkat 2	9.7	71.9	182.1	249.9	820.5	1059.7	886.2	933.7	1192.8	1075.4
9	72.5	saNSR.1	sa.anti-NSR#5 pool	K562	0.6	36.2	143.9	219.3	769.3	930.1	545.8	1275.9	152.3	279.2
10	69.1	saNSR.1+2	sa.anti-NSR#5 pool +sa.anti-NSR#6 pool	K562	0.3	46.5	139.9	146.6	492.9	691.6	602.0	1562.6	3291.7	889.2
11	73.5	saNSR.1+3	sa.anti-NSR#5pool + sa.anti-NSR#7 pool	K562	1.1	24.1	108.4	138.1	586.9	914.5	1480.4	481.7	150.1	224.2

					Input Adjusted Abundance									
Sample ID	**ng/ul**	**1st strand Primer Pool (+Reverse Transcriptase)**	**2nd strand Primer Pool (+ Klenow)**	**RNA**	**NUCB2[1]**	**18S[3]**	**18S-1[2]**	**28S-3[2]**	**28S-4[2]**	**28S-7[2]**	**28S-8[2]**	**12S-1[2]**	**12S-2[2]**	**16S-1[2]**
12	75.9	saNSR.1+4	sa.anti-NSR#5pool +sa.anti-NSR#8 pool	K562										
13	43.6	Y4R-NSR	Y4F-N9	Jurkat 1	6.7	126.1	1830.6	3675.6	874.0	5637.9	904.2	293.6	1437.9	1644.5
14	59.0	Y4-N7	Y4F-N9	Jurkat 1	7.0	562.9	5317.4	19201.8	2489.9	23678.1	2463.8	355.5	1243.7	1751.5
15	47.5	Y4R-NSR	Y4F-N9	Jurkat 2	7.7	253.5	2669.7	6898.6	1716.2	7254.4	1396.9	457.5	2184.7	3482.8
16	59.0	Y4-N7	Y4F-N9	Jurkat 2	7.1	286.6	2948.3	11437.4	1977.7	18794.7	1857.7	282.7	1119.2	1528.5
17	50.2	Y4R-NSR	Y4F-N9	K562	0.4	139.2	1939.0	3940.1	939.7	4801.4	614.6	420.6	1423.4	3997.5
18	54.1	Y4-N7	Y4F-N9	K562	0.5	517.5	4292.3	14486.7	1673.4	15459.0	1590.5	285.6	849.2	1870.3
19	44.8	N8	None- RT only, no second strand synthesis	Jurkat 1	0.4	648.0	3626.8	341.3	1778.6	7321.5	1183.5	299.8	323.8	95.4
20	46.5	N8	None- RT only, no second strand synthesis	Jurkat 2	0.4	758.9	4521.8	513.6	2302.5	9776.5	1396.9	321.6	327.5	104.3
21	44.6	N8	None- RT only, no second strand synthesis	K562	0.0	734.6	3460.3	496.4	2191.6	8023.3	1344.0	286.5	298.8	139.1

1= FAM 10
2= FAM1000
3= Hs99999901

FIGURE 4A graphically illustrates the gene-specific polyA content of cDNA amplified using various NSR primers during first strand synthesis and anti-NSR primers or random primers during second strand synthesis as determined using a set of representative gene-specific assays for PPIA, SRP14, STMN1, TRIM63, ACTB, DBN1, EIF3S3, GAPDH, and NUCB2.

Relative abundance of the polyA content shown in FIGURE 4A was calculated by first combining the input adjusted raw abundance values of individual rRNA assays by transcript. The collapsed rRNA transcript abundance values were normalized to NUCB2 gene levels measured within each sample preparation such that gene content was equal to 1.0. The rRNA/gene ratios calculated for amplified samples were then normalized to that obtained for the unamplified control (N8) such that N8 was equal to 100 for each rRNA transcript. Therefore, the N8 was used as the standard value for the abundance level of each gene.

With regard to the figure legend for FIGURE 4A and FIGURE 4B, with reference to TABLE 2 and TABLE 3, saNSR.1 refers to cDNA amplified using NSR#1 primer pool in the first strand synthesis and anti-NSR#5 primer pool in the second strand synthesis (i.e., depleted for rRNA, mt-rRNA and globin in first and second strand synthesis). saNSR.1+2 refers to cDNA amplified using NSR#1+#2 primer pools in the first strand synthesis and anti-NSR#5+#6 primer pools in the second strand synthesis (i.e., depleted for rRNA and globin, but not depleted for mt-rRNA in both first and second strand synthesis). saNSR.1+3 refers to cDNA amplified using NSR#1+#3 primer pools in the first strand synthesis and anti-NSR #5 +#7 primer pools in the second strand synthesis (i.e., depleted for rRNA and mt-rRNA, but not depleted for globin in both first and second strand synthesis). saNSR.1+4 refers to cDNA amplified using NSR#1+#4 primer pools in the first strand synthesis and anti-NSR#5+#8 primer pools in the second strand synthesis (i.e., depleted for rRNA, but not depleted for mt-rRNA and globin in both first and second strand synthesis). Y4R-NSR refers to cDNA amplified using NSR primers including the core set of 6-mer NSR oligos with no perfect match to globin (alpha or beta), no perfect match to rRNA (18S, 28S).for first strand synthesis, and random 9-mer primers for the second strand synthesis (i.e., depleted for globin and rRNA, but not depleted for mt-rRNA in the first strand synthesis, but not depleted for any sequences in the second strand synthesis). Y4-N7 refers to cDNA amplified using random 7-mer

primers during first and second strand synthesis. Finally, N8 refers to first strand synthesis using random 8mers (no second strand synthesis).

As shown in FIGURE 4A, the NSR priming for first strand synthesis amplified gene-specific transcripts at least as efficiently as random primers, with the exception of the gene TRIM63.

FIGURE 4B graphically illustrates the relative abundance level of non-polyadenylated RNA transcripts in cDNA amplified from Jurkat-1 and Jurkat-2 total RNA using various NSR primers during first strand cDNA synthesis. As shown in FIGURE 4B, gene specific content in the cDNA amplified using NSR and anti-NSR primers is enriched as the rRNA and mt-rRNA content is decreased. This demonstrates that NSR-dependent rRNA depletion is not a general effect, but rather is specific to the transcripts targeted for removal. These results also demonstrate that both polyA minus and polyA plus transcripts are reproducibly amplified using NSR-PCR.

FIGURE 5 graphically illustrates the log ratio of Jurkat/K562 mRNA expression data measured in cDNA generated using the primer pool NSR#1+#3 (x-axis) versus the log ratio of Jurkat/K562 mRNA expression data measured in cDNA generated using the random primer pool N8 (no amplification). This result shows that the relative abundance of messenger RNA in different samples is preserved through NSR priming and PCR amplification.

FIGURE 6A graphically illustrates the proportion of rRNA to mRNA in total RNA that is typically obtained after polyA purification using conventional methods. As shown in FIGURE 6A, prior to polyA purification, total RNA isolated from a mammalian cell includes approximately 98% rRNA and approximately 2% mRNA and other (non-polyA RNA). As shown, even after 95% removal of rRNA from total RNA using polyA purification, the remaining RNA consists of a mixture of about 50% rRNA and 50% mRNA.

FIGURE 6B graphically illustrates the proportion of rRNA to mRNA in a cDNA sample prepared using NSR primers during first strand cDNA synthesis and anti-NSR primers during second strand cDNA synthesis. As shown in FIGURE 6B the use of NSR primers and anti-NSR primers to generate cDNA from total RNA is effective to remove 99.9% rRNA (including nuclear and mitochondrial rRNA), resulting in a cDNA population enriched for greater than 95% mRNA. This is a very significant result for

several reasons. First, the use of polyA purification or strategies that rely on primer binding to the polyA tail of mRNA exclude non-polyA containing RNA molecules such as, for example, miRNA and other molecules of interest, and therefore exclude nucleic acid molecules that contribute to the richness of the transcriptome. In contrast, the methods of the present invention that include the use of NSR primers and anti-NSR primers during cDNA synthesis do not require polyA selection and therefore preserve the richness of the transcriptome. Second, the use of NSR and anti-NSR primers during cDNA synthesis is effective to generate cDNA with removal of 99.9% rRNA, resulting in cDNA with less than 10% rRNA contamination, as shown in FIGURE 6B. This is in contrast to polyA purified mRNA and cDNA synthesis using random primers that only removes 98% rRNA, resulting in cDNA with approximately 50% mRNA and 50% rRNA contamination, as shown in FIGURE 6A.

Conclusion:

These results demonstrate that the NSR #1+#3 primer pool (SEQ ID NOS:1-749) and anti-NSR primer pool (SEQ ID NOS:750-1498) works remarkably well for first strand and second strand cDNA synthesis, respectively, resulting in a double-stranded cDNA product that is substantially enriched for target genes (including poly-adenylated and non-polyadenylated RNA) with a low level (less than 10%) of unwanted rRNA and mt-rRNA.

EXAMPLE 4

This Example shows that the use of the 749 NSR-6mers (SEQ ID NOS:1-749) (each has a spacer N and the PBS#1 (SEQ ID NO:1499) covalently attached at the 5' end) for first strand cDNA synthesis and the use of the 749 anti-NSR-6mers (SEQ ID NOS:750-1498) (that each have a spacer N and the PBS#2 (SEQ ID NO:1500) covalently attached at the 5' end) prime the amplification of a substantial fraction of the transcriptome (both polyA+ and polyA-) and do not prime unwanted non-target sequences present in total RNA, as determined by sequence analysis of the amplified cDNA.

Methods:

cDNA was generated using 749 NSR-6mers (SEQ ID NOS:1-749) (each has a spacer N and the PBS#1 (SEQ ID NO:1499) covalently attached at the 5' end) for first strand cDNA synthesis and the use of the 749 anti-NSR-6mers (SEQ ID NOS:750-1498)

(each has a spacer N and the PBS#2 (SEQ ID NO:1500) covalently attached at the 5' end), with the various primer pools shown in TABLE 8, using the methods described in Example 2.

TABLE 8: Protocols Used to Selectively Amplify cDNA

Protocol Reference Number	First Strand cDNA Primers	Second Strand cDNA Synthesis Primers	Comments	Number of Exp
NSR-V1	NSR primers (no perfect match to rRNA, no globin, + mt rRNA)	N7 random	Reaction conditions: RT run with Y4 primer tails (SEQ ID NO:1504) high dNTP (25 mM), 2 hrs at 40°C, 30 min RNAsH treatment and a 95°C denaturation step	n=170
NSR-V2	NSR primers (no perfect match to rRNA, no globin, + mt rRNA)	N7 random	Reaction conditions: primers and conditions the same as above for NSR-V1 except RNAse treatment for 10 minutes and 95°C denaturation step was eliminated	n=130
NSR-V3	NSR primers (no perfect match to rRNA, no globin, + mt rRNA)	N7 random	Reaction Conditions: primers and conditions the same as above for NSR-V2 except RNAse treatment was eliminated	n=187
NSR-V4	NSR primers (no perfect match to rRNA, no mt-RNA + globin) (SEQ ID NOS: 1-749)	anti-NSR (SEQ ID NOS:750-1499)	Reaction Conditions: primers (SEQ ID NO:1501) were used; reaction conditions as described in Example 2.	n=187
NSR-V5	NSR (no perfect match to rRNA, no mt-RNA + globin) (SEQ ID NOS: 1-749)	anti-NSR (SEQ ID NOS:750-1499)	Reaction conditions: primers and conditions- same as NSR-V4 with additional cleanup step between 1st and 2nd strand synthesis	n=187
N7	N7 Random	N7 Random	Reaction Conditions: same conditions as NSR-V5 with random N7 primers	n=171

The cDNA products were PCR amplified and column purified as described in Example 2. The column-purified PCR products were then cloned into TOPO vectors

using the pCR-XL TOPO kit (Invitrogen). The TOPO ligation reaction was carried out with 1 µl PCR product, 4 µl water and 1 µl of vector. Chemically competent TOP10 One Shot cells (Invitrogen) were transformed and plated onto LB+Kan (50 µg/mL) and grown overnight at 37°C. Colonies were screened for inserts using PCR amplification. It was determined by 2% agarose gel analysis that all clones had inserts of at least 100 bp (data not shown).

The clones were then used as templates for DNA sequence analysis. Resulting sequences were run against a public database for determining homology to rRNA species and the genome.

Results:

TABLE 9 provides the results of sequence analysis of the PCR products generated from cDNA synthesized using the various primer pools shown in TABLE 8.

TABLE 9: Results of DNA Sequence Analysis of aDNA Generated From Selectively Amplified cDNA

Primers Used for cDNA Synthesis	rRNA (% of Total) (18S or 28S rRNA)	mt-RNA (% of Total) (12S or 16S rRNA)	Gene-Specific RNA[1] (% of Total)	Other[2] (% of Total)
N7	77.2	8.2	13.5	1.2
NSR-V1	44.7	19.4	28.8	7.1
NSR-V2	17.0	20.0	51.0	12.0
NSR-V3	2.0	17.0	64.0	17.0
NSR-V4	10.7	5.3	67.4	16.6
NSR-V5	3.7	3.2	78.6	14.4

1 = determined to overlap with any known gene or mRNA including exon, intron, and UTR regions as determined by sequence alignment with public databases.

2 = determined to overlap with repeat elements or alignment to intergenic regions as determined by sequence alignment with public databases.

Conclusion:

These results demonstrate that aDNA (PCR products) amplified from double-stranded cDNA templates generated using the NSR 6-mers (SEQ ID NOS:1-749), and anti-NSR6-mers (SEQ ID NOS:750-1498) as described in Example 2, preserved the enrichment of target genes relative to nuclear ribosomal RNA and mitochondrial ribosomal RNA.

EXAMPLE 5

This Example describes methods that are useful to label the aDNA (PCR products) for subsequent use in gene expression monitoring applications.

1. Direct Chemical Coupling of Fluorescent Label to the PCR Product.

Cy3 and Cy5 direct label kits were obtained from Mirus (Madison, Wisconsin, kit MIR Product Numbers 3625 and 3725).

10 µg of PCR product (aDNA), obtained as described in Example 2, was incubated with labeling reagent as described by the manufacturer. The labeling reagents covalently attach Cy3 or Cy5 to the nucleic acid sample, which can then be used in almost any molecular biology application, such as gene expression monitoring. The labeled aDNA was then purified and its fluorescence was measured relative to the starting label.

Results:

Four aDNA samples were labeled as described above and fluorescence was measured. A range of 0.9 to 1.5% of retained label was observed across the four labeled aDNA samples (otherwise referred to as a labeling efficiency of 0.9 to 1.5%). These results fall within the 1% to 3% labeling efficiency typically observed for aaUTP labeled, *in vitro* translated, amplified RNA.

2. Incorporation of aminoallyl modified dUTP (aadUTP) during PCR with an aDNA template using one primer (forward or reverse) to yield aa-labeled, single-stranded aDNA.

Methods:

1 µg of the aDNA PCR product, generated using the NSR and anti-NSR primer pool as described in Example 2, is added to a PCR reaction mix as follows:

- 100 to 1000 µM aadUTP+dCTP+cATP+dGTP+dUTP (the optimal balance of aadUTP to dUTP may be empirically determined using routine experimentation)
- 4 mM $MgCl_2$
- 400-1000 nM of only the forward or reverse primer, but not both.

PCR Reaction: 5 to 20 cycles of PCR (94°C 30 seconds, 60°C 30 seconds, 72°C 30 seconds), during which time only one strand of the double-stranded PCR template is synthesized. Each cycle of PCR is expected to produce one copy of the aa-labeled,

single-stranded aDNA. This PCR product is then purified and a Cy3 or Cy5 label is incorporated by standard chemical coupling.

3. Incorporation of aminoallyl modified dUTP (aadUTP) during PCR with an aDNA template using forward and reverse primers to yield aa-labeled, double-stranded aDNA.

Methods:

1 µg of the aDNA PCR product generated using the NSR7 primer pool as described in Example 11 is added to a PCR reaction mix as follows:

- 100 to 1000 µM aadUTP+dCTP+cATP+dGTP+dUTP (the optimal balance of aadUTP to dUTP may be empirically determined using routine experimentation)
- 4 mM $MgCl_2$
- 400-1000 nM of the forward and reverse primer (e.g., Forward: SEQ ID NO:1501; or Reverse: SEQ ID NO:1502)

PCR Reaction: 5 to 20 cycles of PCR (94°C 30 seconds, 60°C 30 seconds, 72°C 30 seconds), during which time both strands of the double-stranded PCR template are synthesized. The double-stranded, aa-labeled aDNA PCR product is then purified and a Cy3 or Cy5 label is incorporated by standard chemical coupling.

EXAMPLE 6

This Example describes the use of a hybrid RNA/DNA primer covalently linked to NSR-6mers to generate amplified nucleic acid templates useful for generating single-stranded DNA molecules for gene expression analysis.

Rationale: In one embodiment of the selective amplification methods of the invention, the defined sequence portion (e.g., PBS#1) of a first oligonucleotide population for first strand cDNA synthesis, and/or the defined sequence portion (e.g., PBS#2) of a second oligonucleotide population for second strand cDNA synthesis comprises an RNA portion to generate an amplified nucleic acid template suitable for generating multiple copies of DNA products using strand displacement, as described in U.S. Patent No. 6,946,251, hereby incorporated by reference. A hybrid NSR primer (PBS#1(RNA/DNA)/NSR) may be used to synthesize first strand cDNA, thereby generating products suitable for use as templates for synthesis of single-stranded DNA

having a sequence complementary to template RNA. Alternatively, an RNA/DNA hybrid primer tail may be added after second strand synthesis, as described in more detail below.

One advantage provided by this method is the ability to generate a plurality of single-stranded amplification products of the original cDNA sequence, and not the amplification of the product of the amplification itself.

Methods:

1. RNA:DNA hybrid NSR for First Strand cDNA Synthesis:

In some embodiments, the population of NSR primers for use in first strand cDNA synthesis (SEQ ID NOS:1-749) may further comprise a 5' primer binding sequence (RNA), such as hybrid PBS#1:

Hybrid PBS#1(RNA) 5' GACGGAUGCGGUCU 3' (SEQ ID NO:1557) covalently attached at the 5' end of the NSR primers.

Resulting in a population of RNA:DNA hybrid oligonucleotides having an RNA defined sequence portion located 5' to the DNA hybridizing portion with the following configuration:

5' hybrid PBS#1(RNA) (SEQ ID NO:1557) + NSR6-mer (DNA) (SEQ ID NOS:1-749) 3'

In another embodiment, a population of oligonucleotides may be generated wherein each NSR6-mer optionally includes at least one DNA spacer nucleotide (N) (where each N = A, G, C, or T) where (N) is located between the 5' hybrid PBS#1 (RNA) and the NSR6mer (DNA). The spacer region may comprise from one nucleotide up to ten or more nucleotides (N = 1 to 10), resulting in a population of oligonucleotides having the following configuration:

5' Hybrid PBS#1(RNA) (SEQ ID NO:1557) + (N_{1-10}) (DNA) + NSR6-mer (SEQ ID NOS:1-749) (DNA)3'

The process of preparing the first strand cDNA is carried out essentially as described in Example 2, with the substitution of the hybrid PBS#1 (SEQ ID NO:1557) (RNA) for the PBS#1 (SEQ ID NO:1499) (DNA), with the use of an RNAseH- reverse

transcriptase and without the addition of RNAseH prior to second strand cDNA synthesis, to generate a double-stranded substrate for amplification of single-stranded DNA products

The substrate for single stranded amplification preferably consists of a double stranded template with the first strand consisting of an RNA/DNA hybrid molecule and the second strand consisting of all DNA. In order to construct this double-stranded template, second strand synthesis is carried out using an RNAseH- reverse transcriptase. Alternatively, the second strand synthesis may be carried out using Klenow followed by a polished step with RNAseH- reverse transcriptase, since Klenow will not use RNA as a template.

Second strand cDNA synthesis may be carried out using either random primers, or using anti-NSR primers. The use of the RNA hybrid/NSR primer population during first strand cDNA synthesis results in the incorporation of a unique sequence of the RNA portion of the hybrid primer into the synthesized single-stranded cDNA product.

Single-stranded DNA amplification products that are identical to the target RNA sequence may then be generated from the double-stranded template described above by denaturing and RNAseH treating the denatured substrate to remove the RNA portion of the substrate, and adding a hybrid RNA/DNA single-stranded amplification primer, e.g., 5' GACGGAUGCGGTGT 3' (SEQ ID NO:1558), where the 5' portion of the primer consists of at least eleven RNA nucleotides (underlined) that hybridize to a predetermined sequence on the first strand cDNA and the 3' portion consists of at least three DNA nucleotides to the substrate in the presence of a highly processive strand displacing DNA polymerase, such as, for example, phi29.

In alternative embodiment, the substrate for single-stranded DNA amplification may be prepared by preparing first strand cDNA synthesis using DNA primers (e.g., NSR or random primers), followed by second strand synthesis with Klenow also using DNA primers (e.g., anti-NSR or random primers). The double-stranded DNA template is then modified to produce a substrate for single-stranded DNA amplification by denaturing and annealing an RNA/DNA hybrid oligonucleotide that hybridizes to the second strand cDNA and extending the hybrid RNA/DNA oligonucleotide with Reverse Transcriptase, to generate a double stranded template with one strand consisting of an RNA/DNA hybrid molecule and the other strand consisting of all DNA.

Single stranded DNA amplification products that are complementary to the target RNA sequence may then be generated from the double-stranded substrate by denaturing and RNAseH treating the denatured substrate to remove the RNA portion of the substrate. A hybrid RNA/DNA single-stranded amplification primer is then annealed to the second strand, wherein the 5' portion of the hybrid primer consists of at least eleven RNA nucleotides that hybridize to a pre-determined sequence on the second strand cDNA and the 3' portion of the hybrid primer consists of at least three DNA nucleotides. A highly processive strand displacing DNA polymerase, such as, for example, phi29. is then used to generate single-stranded DNA products.

EXAMPLE 7

This Example describes the robust detection of poly A+ and poly **A**- transcripts in cDNA amplified from total RNA using NSR primers .

Rationale:

The whole transcriptome, that is, the entire collection of RNA molecules present within cells and tissues at a given instant in time, carries a rich signature of the biological status of the sample at the moment the RNA was collected. However, the biochemical reality of total RNA is that an overwhelming majority of it codes for structural subunits of cytoplasmic and mitochondrial ribosomes, which provide relatively little information on cellular activity. Consequently, molecular techniques that enrich for more informative low copy transcripts have been developed for large-scale transcriptional studies, such as the exploitation of 3' polyadenylation sequences as an affinity tag for non-ribosomal RNA. Targeted sequencing of polyA+ RNA transcripts has provided a rich foundation of cDNA fragments that form the basis of current gene models (see e.g., Hsu F. et al., *Bioinformatics* 22:1036-1046 (2006)). Priming of cDNA synthesis from polyA sequences has also been used for the most commonly practiced, genome-wide RNA profiling methods.

Although these methods have been very successful for analysis of messenger RNA expression, methods that strictly focus on polyA+ transcripts present an incomplete view of global transcriptional activity. PolyA priming often fails to capture information distal to 3' polyA sites, such as alternative splicing events and alternative transcriptional start sites. Conventional methods also fail to monitor expression of non-poly-adenylated transcripts including those that encode protein subunits of histone deacetylase and many

non-coding RNAs. Although alternative methods have been developed to specifically target many of these RNA sub-populations (Johnson J.M. et al., *Science 302*:2141-2144 (2003); Shiraki T. et al., *PNAS* 100:15776-15781 (2003); Vitali P. et al., *Nucleic Acids Res. 31*:6543-6551 (2003)), only a few studies have attempted to monitor all transcriptional events in parallel. The most comprehensive analysis of whole transcriptome content has been carried out using genome tiling arrays (Cheng J. et al., *Science 308*:1149-1154 (2005); Kapranov P. et al., *Science 316*:1484-1488 (2007)). However, the complexity of these experiments and the need for subsequent validation by complementary methods has limited the use of tiling arrays for routine whole transcriptome profiling applications. Recent advances in DNA sequencing present an opportunity for new approaches to expression analysis, allowing both the quantitative assessment of RNA abundance and experimentally-verified transcript discovery on a single platform (Mortazavi A. et al., *Nat. Methods 5*:621-628 (2008)). Therefore, there is a need for a method that provides an unbiased survey of both known and novel transcripts that can utilize high-throughput profiling of numerous samples.

Methods:

Overview:

In accordance with the foregoing, the inventors have developed a sample preparation procedure that relies on the "not-so-random" ("NSR") priming libraries in which all hexamers with perfect matches to ribosomal RNA (rRNA) sequences have been removed. For NSR selective priming to be useful as a whole transcriptome profiling technology, it must faithfully detect non-ribosomal RNA transcripts. To test the performance of NSR-priming, a whole transcriptome cDNA library was constructed. Antisense NSR hexamers ("NSR" primers) were synthesized to prime first strand synthesis, with a universal tail sequence to facilitate PCR amplification and downstream sequencing using the Illumina 1G Genome Analyzer. A second set of tailed NSR hexamers complementary to the first set of NSR primers ("anti-NSR" primers) was generated to prime 2nd strand synthesis. The unique tail sequences used for first and second strand NSR primers enabled the preservation of strand orientation during amplification and sequencing. For this study, all sequencing reads were oriented in a 3' to 5' direction with respect to the template RNA, although opposite strand reads can be easily generated by modifying the universal PCR amplification primers.

To evaluate whole transcriptome content in NSR-primed libraries, a survey was conducted of NSR-primed cDNA libraries generated from the RNA isolated from whole brain and RNA isolated from the Universal Human Reference (UHR) cell line (Stratagene) by sequencing, as described below.

Oligonucleotides Used to Generate Libraries:

A first population of NSR-6mer primers 5' (SEQ ID NO:1499) covalently attached to each of (SEQ ID NOS:1-749) was used for amplification of the first strand and a second population of anti-NSR-6mer primers (SEQ ID NO:1500) covalently attached to each of (SEQ ID NOS:750-1498) for use in second strand cDNA synthesis, as described in Example 1. Oligos were desalted and resuspended in water at 100 uM before pooling.

A collection of random hexamers were also synthesized with the tail sequences SEQ ID NO:1499 and SEQ ID NO:1500 for generation of control libraries.

Library Generation:

Overview: NSR-priming selectively captures the non-ribosomal RNA fraction including poly A+ and poly A- transcripts. Two rounds of NSR priming selectivity were applied during library construction. First, NSR oligonucleotides (antisense) initiate reverse transcription at not-so-random template sites. Following ribonuclease treatment to remove the RNA template, anti-NSR oligonucleotides (sense) anneal to single-stranded cDNA at not-so-random template sites and direct Klenow-mediated second strand synthesis. PCR amplification with asymmetric forward and reverse primers preserves strand orientation and adds terminal sites for downstream end sequencing. Antisense tag sequencing is then carried out from the 3' end of cDNA fragments using a portion of the forward amplification primer. Pairwise alignments are then used to map the reverse complements of tag sequences to the human genome.

Methods:

Total RNA from whole brain was obtained from the FirstChoice® Human Total RNA Survey Panel (Ambion, Inc.). Universal Human Reference (UHR) cell line RNA was purchased from Stratagene Corp. Total RNA was converted into cDNA using Superscript™ III reverse transcription kit (Invitrogen Corp). Second strand synthesis was

carried out with 3'-5' exo-Klenow Fragment (New England Biolabs Inc.). DNA was amplified using Expand High Fidelity^{PLUS} PCR System (Roche Diagnostics Corp.).

For NSR primed cDNA synthesis, 2 µl of 100 µM NSR primer mix (SEQ ID NO:1499 plus SEQ ID NOS:1-749) was combined with 1 µl template RNA and 7 µl of water in a PCR-strip-cap tube (Genesee Scientific Corp.). The primer-template mix was heated at 65°C for 5 minutes and snap-chilled on ice before adding 10 µl of high dNTP reverse transcriptase master mix (3 µl of water, 4 µl of 5X buffer, 1 µL of 100 mM DTT, 1 µl of 40 mM dNTPs and 1.0 µl of SuperScript™ III enzyme). The 20 µl reverse transcriptase reaction was incubated at 45°C for 30 minutes, 70°C for 15 minutes and cooled to 4°C. RNA template was removed by adding 1 µl of RNAseH (Invitrogen Corp.) and incubated at 37°C for 20 minutes, 75°C for 15 minutes and cooled to 4°C. DNA was subsequently purified using the QIAquick® PCR purification kit and eluted from spin columns with 30 µl elution buffer (Qiagen, Inc. USA).

For second strand synthesis, 25 µl of purified cDNA was added to 65 µl Klenow master mix (46 µl of water, 10 µl of 10X NEBuffer 2, 5 µl of 10 mM dNTPs, 4 µl of 5 units/µL exo-Klenow Fragment, New England Biolabs, Inc.) and 10 µL of 100 µM anti-NSR primer mix (SEQ ID NO:1500 plus SEQ ID NOS:750-1498). The 100 µl reaction was incubated at 37°C for 30 minutes and cooled to 4°C. DNA was purified using QIAquick spin columns and eluted with 30 µl elution buffer (Qiagen, Inc. USA).

For PCR amplification, 25 µL of purified second strand synthesis reaction was combined with 75 µL of PCR master mix (19 µl of water, 20 µl of 5X Buffer 2, 10 µl of 25 mM $MgCl_2$, 5 ul of 10 mM dNTPs, 10 µl of 10 µM forward primer, 10 µL of 10 µM reverse primer, 1 µL of ExpandPLUS enzyme, Roche Diagnostics Corp.).

Forward PCR primer:

(5'ATGATACGGCGACCACCGACACTCTTTCCCTACACGACGCTCTTCCG ATCTCT3' (SEQ ID NO:1559))

Reverse PCR primer:

(5'CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTGA3' (SEQ ID NO:1560))

Samples were denatured for 2 minutes at 94°C and followed by 2 cycles of 94°C for 10 seconds, 40°C for 2 minutes, 72°C for 1 minute, 8 cycles of 94°C for 10 seconds, 60°C for 30 seconds, 72°C for 1 minute, 15 cycles of 94°C for 15 seconds, 60°C for

30 seconds, 72°C for 1 minute with an additional 10 seconds added at each cycle; and 72°C for 5 minutes to polish ends before cooling to 4°C. Double-stranded DNA was purified using QIAquick spin columns.

A control library was generated using the same methods with the use of random primers, expect for the concentration of dNTPs was 0.5 mM (rather than 2.0 mM) in the final reverse transcription reaction. The random primed control library was amplified using the PCR primers SEQ ID NO:1559 and SEQ ID NO:1560.

Quantitative PCR:

Individual rRNA and mRNA transcripts were quantified by qPCR using TaqMan® Gene Expression Assays (Applied Biosystems). qPCR Assays were carried out using the reagents shown below in TABLE 10.

TABLE 10: Primers for qPCR Assay

Target	**ABI Assay Probe**	**Forward Primer**	**Reverse Primer**	**FAM reporter primer**
PPIA peptidylprolyl isomerase A (cyclophilin A)	Hs99999904_m1	NR	NR	NR
STMN1 stathmin 1/ oncoprotein 18	Hs01027516_g1	NR	NR	NR
EIF3S3 eukaryotic translation initiation factor 3, subunit 3 gamma, 40 kDa	Hs00186779_m1	NR	NR	NR
18s rRNA	Hs99999901_s1	NR	NR	NR
12S rRNA	custom	SEQ ID NO:1532	SEQ ID NO:1533	SEQ ID NO:1534
16S rRNA	custom	SEQ ID NO:1526	SEQ ID NO:1527	SEQ ID NO:1528
28S rRNA	custom	SEQ ID NO:1511	SEQ ID NO:1512	SEQ ID NO:1513

Triplicate measurements of diluted library DNA were made for each assay in 10 µl final reaction volumes in a 384-well optical PCR plate using a 7900 HT PCR instrument (Applied Biosystems). Following PCR, the results table was exported to Excel (Microsoft Corp.), standard curves were generated, and quantitative analysis for

samples was regressed from the raw data. Abundance levels were then normalized to input cDNA mass.

Results of qPCR Analysis:

Comparison of cDNA libraries generated from whole brain total RNA using either NSR-priming or a nonselective priming control of random sequence, tailed heptamers revealed a significant depletion of rRNA and a concomitant enrichment of target mRNA in NSR-primed libraries. Specifically, a >95% reduction was observed in the abundance of all four of the rRNA transcripts included in the computational filter used for NSR primer design (data not shown).

Sequence and Read Classification:

In order to obtained a detailed view of rRNA depletion in NSR primed libraries, tag sequences were generated as 36 nucleotide antisense reads from NSR-primed (2.6 million) and random-primed (3.8 million) cDNA libraries using the Illumina 1G Genome Analyzer (Illumina, Inc.). To characterize sequence tags, the dinucleotide barcode (CT) at the 5' end of each read was removed and the reverse complement of bases 2-34 was aligned to several sequence databases using the ELAND mapping program, which allows up to 2 mismatches per 32 nt alignment (Illumina, Inc.).

To generate expression profiles of RefSeq mRNA and non-coding RNA transcripts, each tag sequence was permitted to align to multiple transcripts. Read counts were then converted to expression values by calculating frequency per 1000 nucleotides from transcript length. A sample normalization factor (nf) was applied to adjust for the total number of reads generated from each library. This was derived from the total number of non-ribosomal RNA reads mapping to the genome for each library (brain 1:17.7 million reads, 1.0 nf; brain 2:19.3 million reads, 1.087 nf; UHR:17.6 million reads, 0.995 nf).

For global classification, sequencing reads were first aligned to the non-coding RNA and repeat databases with alignments to multiple reference sequences permitted. The remaining tag sequences were then mapped to the March 2006 hg18 assembly of the human genome sequence (http:genome.ucsd.edu/). Reads mapping to single genomic sites were classified into mRNA, intron and intergenic categories using coordinates defined by UCSC Known Genes (http://genome.ucsc.edu). Sequences that mapped to multiple genomic sequences that did not include repeats or non-coding RNAs made up

the "other" category. Ribosomal RNA sequences were obtained from RepeatMasker (http://www.repeatmasker.org/) and Genbank (NC_001807). Non-coding RNA sequences were collected from Sanger RFAM (http://www.sanger.ac.uk/Software/Rfam/), Sanger miRBASE (http://microrna.sanger.ac.uk), snoRNABase (http://www-snorna.biotoul.fr) and RepeatMasker. Repetitive elements were obtained from RepeatMasker.

Results: More than 54 million high quality 32-nucleotide tag sequence reads that aligned to non-rRNA genomic regions were obtained from two independently prepared whole brain libraries and a single UHR library. Seventy-seven percent of these reads mapped to single genomic sites. Among 22,785 model transcripts in the RefSeq mRNA database (Pruitt K.D. et al., *Nucleic Acids Res. 33*:D501-504 (2005)), over 87% were represented by 10 or more sequence tag reads in at least some of the samples queried, and 69% were represented by 10 or more reads in all three libraries.

TABLE 11: Results of alignment of 32 nucleotide tag sequence reads from NSR-primed (2.6 million) and random-primed (3.8 million) libraries.

Target	NSR Primed Library (1st and 2nd strand NSR)	Random-primed library
large subunit rRNA (includes 5S, 5.8S and 28S rRNA transcripts)	10.3%	47.2%
small subunit rRNA (includes 18S rRNA transcript)	0.8%	18.0%
mitochondrial rRNA (includes 12S and 16S rRNA)	2.2%	12.6%
non-ribosomal RNA (includes all other sequences that mapped to one or more genomic sites)	86.7%	22.2%

As shown above in TABLE 11, only 13% of sequence tags from NSR primed libraries mapped to the human genome corresponded to ribosomal RNA, whereas 78% of random-primed cDNA matched rRNA sequences. These results demonstrate that NSR-

priming resulted in a nearly complete depletion of small subunit 18S rRNA and a dramatic reduction in mitochondrial rRNA transcripts. Although the reduction of large subunit rRNA abundance was less efficient than other rRNA transcripts, relatively modest depletion of 28S RNA can have a large impact on final library composition, owing to its high initial molar concentration and transcript length. In addition, over 86% of NSR-primed sequences mapped to non-rRNA genomic regions compared to 22% of random-primed cDNA. Only 5% of all sequence reads from either library did not map to any genomic sequence, indicating that the library construction process generated very little template-independent artifacts. Similar results were observed from NSR-primed and random-primed libraries generated from UHR total RNA, isolated from a diverse mixture of cell lines (data not shown).

In order to detect polyA+ RefSeq mRNA in NSR-primed libraries, quantitative analysis of sequencing alignments within RefSeq transcripts was used to produce sequence-based digital expression profiles. Excellent reproducibility of NSR-primed cDNA amplification was observed between two separate NSR libraries prepared from the same whole brain total RNA, with a log10 ratio of transcripts represented by at least 10 NSR tag sequences in replicate #1 versus replicate #2 with a correlation coefficient of r=0.997 for n=17,526.

To assess the accuracy of mRNA profiles obtained from NSR libraries, a comparison was made between the NSR-primed brain profile and the UHR expression profile to the "gold-standard" TaqMan® qPCR profile created for the MicroArray Quality Control Study (MAQC Consortium) (Shi L. et al., *Nat. Biotechnol. 24*:1151-1161 (2006)),

Correlation of gene expression profiles obtained by NSR tag sequencing and TaqMan® quantitative PCR was also assessed. The log 10 ratios of transcript levels in brain and UHR obtained by NSR tag sequencing were plotted against TaqMan® measurements obtained from the MAQC Consortium with a correlation coefficient of r=0.930 for n=609.

Detection of poly A+ Ref Seq mRNA in NSR-primed libraries was carried out as follows. The positional distribution of NSR tag sequences was examined across transcript lengths. FIGURE 7A shows the combined read frequencies for 5,790 transcripts shown at each base position starting from the 5' termini, with NSR (dotted

line) or EST (solid line) cDNAs across long transcripts (≥4kb). FIGURE 7B shows the combined read frequencies for 5,790 transcripts shown at each base position starting from the 3' termini, with NSR (dotted line) or EST (solid line) cDNAs across long transcripts (≥4kb). Data shown in FIGURES 7A and 7B were normalized to the maximal value within each dataset. As shown in FIGURES 7A and 7B, NSR-primed cDNA fragments show full-length coverage of large transcripts with higher representation of internal sites than conventional ESTs. This is an important feature of whole transcriptome profiling because the technology preferably captures alternative splicing information. The sequencing coverage exhibited a modest deficit at the extreme 5' ends of known transcripts owing to the fact that all of the sequencing reads were generated from the 3' ends of cDNA fragments. This effect may be alleviated if sequencing is directed at both ends of NSR cDNA products. Taken together, these results demonstrate the robustness of NSR-based selective priming as a technology for whole transcriptome expression profiling.

Another requirement of whole transcriptome profiling is that it must effectively capture poly A- transcripts. The representation of poly A- non-coding RNAs in NSR-primed cDNA was determined as follows. Sequence tags from NSR-primed libraries were aligned to a comprehensive database of known poly A- non-coding RNA (ncRNA) sequences. Transcripts representing diverse functional classes were widely detected with a substantial fraction of small nucleolar RNAs ("snoRNAs") (286/665) and small nuclear RNAs ("snRNAs") (7/19) present at 5 or more copies in at least one sample. Interestingly, only a small portion of miRNA hairpins and tRNA species were observable at detectable levels. As shown below in TABLE 12, individual transcripts were observed over a broad range of expression levels with members of the snRNA and snoRNA families among the most highly abundant.

TABLE 12: Rank-ordered Expression Levels of non-coding (ncRNA) transcripts represented by at least two NSR tag sequences in whole brain

ncRNA Transcript/Type	Log10 Expression Level	Brain Expression Rank (out of a total of 200)
HBII-52 (brain-specific C/D box snoRNA)	6.5	1st
HBII-85 (brain-specific C/D box snoRNA)	6	2nd

U2 (snRNA)	5.8	3rd
U1 (snRNA)	5.3	5th
U3 (snRNA)	5	8th
U4 (snRNA)	4.8	10th
U13 (snRNA)	3.7	28th
U6 (snRNA)	3.5	33rd
HBII-436 (brain-specific C/D box snoRNA)	3.4	40th
HBII-437 (brain-specific C/D box snoRNA)	3.1	60th
HBII-438A (brain-specific C/D box snoRNA)	2.8	85th
HBII-13 (brain-specific C/D box snoRNA)	2.7	90th
U5 (snRNA)	2.3	105th
U8 (snRNA)	2	140th

As shown below in TABLE 13, the NSR-primed libraries containing poly A-transcripts included members of the snRNA and snoRNA families, as well as RNAs corresponding to other well-known transcripts such as 7SK, 7SL and members of the small cajal body-specific RNA family.

TABLE 13: Representation of Major non-coding (ncRNA) Classes in NSR primed library generated from Whole Brain Total RNA

polyA- Transcript in NSR primed library	% of library
snoRNA	60.4%
snRNA	22.1%
7SL	13.8%
7SK	4.7%
scRNA	1.3%
miRNA	0.7%
tRNA	0.1%

Many transcripts were found to be enriched in the NSR primed library generated from the whole brain total RNA, as compared to the NSR primed library generated from UHR, including the cluster of C/D box snoRNAs located in the q11 region of chromosome 15 that has been implicated in the Prader-Willi neurological syndrome (Cavaile J. et al., *J. Biol. Chem. 276*:26374-26383 (2001); Cavaile J. et al., *PNAS 97*:14311-14316 (2000)). FIGURE 8 graphically illustrates the enrichment of snoRNAs

encoded by the Chromosome 15 Prader-Willi neurological disease locus in whole brain NSR primed library relative to the UHR NSR primed library.

It is interesting to note that a significant proportion of known ncRNA transcripts detected in this study were less than 100 nucleotides in length and were predicted to have extensive secondary structure, thereby also demonstrating that NSR-priming is capable of capturing templates considered problematic to capture using conventional methods.

Global Overview of Transcriptional Activity

The collection of whole transcriptome cDNA sequences generated using NSR priming may be assembled into a global expression map for whole brain and UHR. In order to assemble such a global expression map, all non-ribosomal RNA tag sequences were assigned to one of six non-overlapping categories based on current genome annotations as shown in TABLE 14 below.

TABLE 14: Classification of Whole Transcriptome Expression in NSR-primed cDNA tags mapping to non-ribosomal RNA genomic regions

Category	NSR-primed whole Brain library	NSR-primed UHR library
mRNA	46%	35%
intron	19%	30%
intergenic	12%	13%
ncRNA	4%	1%
repeats	3%	6%
other	16%	15%

The mRNA, intron and intergenic categories shown above in TABLE 14 were defined by the genomic coordinates of UCSC Known Genes and include only cDNAs that map to unique locations. Sequencing tag reads overlapping any part of a coding exon or UTR were considered mRNA. Sequencing tag reads mapping to multiple genomic sites were binned into the ncRNA, repeats or other categories.

As shown above in TABLE 14, it was determined that tissue and cell line RNA populations exhibited similar overall expression patterns. For example, 65% of tag sequences occurred within the boundaries of known protein-coding genes, whereas only 12-13% of tag sequences mapped to intergenic regions, which is considerably lower than previously reported (Cheng J. et al., *Science 308*:1149-1154 (2005)). The fraction of

cDNAs corresponding to pseudogenes and other redundant sequences, such as motifs shared within gene families (the "other" category in TABLE 14), was also similar in both samples. However, the representation of some categories was notably different in whole brain and UHR. Although intronic expression was substantial in both RNA populations, transcriptional activity in introns was 60% higher in UHR than in whole brain. Expression of repetitive elements was also higher in UHR than in whole brain. In contrast, the cumulative abundance of known ncRNAs was 4-fold higher in brain than UHR. While not wishing to be bound by any particular theory, these results may reflect general differences in splicing activity between cell lines and tissues. Alternatively, these findings may indicate that transcription is generally more pervasive in cell lines and may be a result of relaxed regulatory constraints.

In order to assess the number of unique transcription sites ascribed to unannotated regions, overlapping NSR tag sequences were assembled into contiguous transcription units. Multiple sequencing reads mapping to single genomic sites were collapsed into single transcripts when at least one nucleotide overlapped on either strand. Overall, over 2.5 million transcriptionally active regions were identified that were not covered by current transcript models. Of these, only 21% were supported by sequences in public EST databases (Benson, D.A. et al., *Nucleic Acids Res 32*:D23-26 (2004)). Unannotated transcription sites averaged 36.9 nucleotides in length and ranged from 32 to 1003 bp, with nearly 5% exceeding 100 bp. Many of the transcriptional elements identified here may represent novel non-coding RNAs. They may also be previously unidentified segments of known genes including alternatively spliced exons and extensions of untranslated regions.

Next, the strand specificity of NSR priming was examined by aligning sequence tags to functional elements of known protein-coding genes. Over 99% of cDNA sequences mapping to protein-coding exons were oriented in the sense orientation, demonstrating the discrimination power of this method for monitoring strand-specific expression. This discrimination power allowed us to determine the orientation of novel transcripts and to assess the prevalence of antisense transcription among the functional elements of known genes. As shown below in TABLE 15, antisense transcription was detected at particularly high levels in 5'UTRs and introns, constituting about 20% of transcription events in those regions.

TABLE 15: The relative frequency ratio of NSR tag sequences oriented in the sense or antisense direction for sequencing reads obtained from NSR primed whole brain and UHR libraries

Element of Known genes	Relative frequency ratio of Sense Reads	Relative frequency ratio of Antisense Reads
5' UTR	0.80	0.20
coding exon	0. 99	0.01
3' UTR	0.95	0.05
intron	0.80	0.20

The sequencing categories shown above in TABLE 15 were defined by the genomic coordinates of non-coding and coding regions of UCSC known genes.

It is interesting to note that other groups have also documented widespread antisense expression in humans and several model organisms (Katayama S. et al., *Science 309*:1564-1566 (2005); Ge X. et al., *Bioinformatics 22*:2475-2479 (2006); Zhang Y. et al., *Nucleic Acid Res 34*:3465-3475 (2006)). The complex patterns of sense and antisense expression observed in many genes suggest that at least some of the intronic and UTR transcriptional events have functional significance.

Discussion:

As demonstrated in this Example, the application of ultra-high throughput sequencing to NSR-primed cDNA libraries allows for the unbiased interrogation of global transcriptional content that surpasses the scope of information produced by conventional methods. Transcript discovery by sequencing provides information with a level of specificity that cannot be achieved with genomic tiling arrays, which are prone to adverse cross-hybridization effects that necessitate significant data processing and subsequent experimental validation (see. e.g., Royce T.E. et al., *Trends Genet 21*:466-475 (2005)). However, the depth of sampling needed to obtain sufficient coverage of rare transcripts in highly complex whole transcriptome libraries limits the capacity of sequencing to rapidly survey large numbers of tissues. In contrast, expression profiling microarrays facilitate the quantitative analysis of transcript levels in many samples, provided there is quality sequence information to direct probe selection.

NSR selective priming provides several advantages over conventional methods. For example, NSR selective priming provides a direct link between informative

sequencing and high throughput array experiments. The sequence information obtained using NSR selective primed cDNA libraries allows for the identification of unannotated transcriptional features. The functional characterization of the unannotated transcriptional features identified using the NSR-primed libraries will shed light on a wide range of biological processes and disease states.

The information obtained from high-throughput sequencing may used to inform the design of whole transcriptome arrays for hybridization with NSR-primed cDNA. For example, custom designed whole transcriptome profiling arrays may be used to assess the expression patterns of novel features in relation to one another and in the context of known transcripts. Large scale profiling studies may also be used to implicate individual transcripts in human pathological states and expand the repertoire of biomarkers available for clinical studies (see, e.g., van't Veer, L.J. et al., *Nature 415*:530-536 (2002)). In addition, the integration of whole transcriptome expression profiling data with genetic linkage analysis may be used to reveal biological activities that are modulated by novel transcriptional elements.

Variations of the tag sequencing method described in this example may be utilized for whole transcriptome analysis in accordance with various embodiments of the invention. In one embodiment, paired-end sequencing is utilized for whole transcriptome analysis. Paired-end sequencing provides a direct physical link between the 5' and 3' termini of individual cDNA fragments (Ng P. et al., *Nucleic Acids Res 34* e84 (2006); and Campbell, P.J. et al., *Nat Genet 40*:722-729 (2008)). Therefore, pair-end sequencing allows spliced exons from distal sites to be unambiguously assigned to a single transcript without any additional information. Once whole transcript structures are defined, large-scale computational analysis can be applied to determine whether these genes represent protein-coding or non-coding RNA entities (Frith M.C. et al., *RNA Biol. 3*:40-48 (2006)).

As described above, NSR priming is an elementary form of cDNA subtraction with the advantage that it can be simply and reproducibly applied to a wide variety of samples. NSR primer pools may be designed to avoid any population of confounding, hyper-abundant transcripts. For example, an NSR primer pool may be designed to avoid the mRNAs encoding the alpha and beta subunits of globin proteins, which constitute up to 70% of whole blood total RNA mass, and can adversely affect both the sensitivity and accuracy of blood profiling experiments (see Li L. et al., *Physiol. Genomics 32*:190-197

(2008)). NSR primer pools may also be designed to reduce rRNA content in other organisms, allowing cross-species comparisons of whole transcriptome expression patterns. This approach may be utilized for routine expression profiling experiments in prokaryotic species, where polyA selection of RNA sub-populations is not useful.

In summary, analysis of over 54 million 32-nucleotide tag sequences demonstrated that NSR-priming in the first and second strand cDNA synthesis produces cDNA libraries with broad representation of known poly A+ and poly A- transcripts and dramatically reduced rRNA content when compared to conventional random-priming. The sequencing of NSR-primed libraries provides a global overview of transcription which includes evidence of widespread antisense expression and transcription from previously unannotated genomic sequences. Thus, the simplicity and flexibility of NSR priming technology makes it an ideal companion for ultra-high-throughput sequencing in transcriptome research across a wide range of experimental settings.

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

CLAIMS

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

1. A method of selectively amplifying a target population of nucleic acid molecules within a population of RNA template molecules, the method comprising the steps of:

(a) providing a population of single-stranded primer extension products synthesized from a population of RNA template molecules in a sample isolated from a mammalian subject using reverse transcriptase enzyme and a first population of oligonucleotide primers, wherein each oligonucleotide in the first population of oligonucleotide primers comprises a hybridizing portion and a defined sequence portion located 5' to the hybridizing portion, wherein the population of RNA template molecules comprises a target population of nucleic acid molecules and a non-target population of nucleic acid molecules; and

(b) synthesizing double-stranded cDNA from the population of single-stranded primer extension products according to step (a) using a DNA polymerase and a second population of oligonucleotide primers, wherein each oligonucleotide in the second population of oligonucleotides comprises: a hybridizing portion consisting of 6, 7, or 8 nucleotides and a defined sequence portion located 5' to the hybridizing portion, wherein the hybridizing portion is selected from all possible oligonucleotides having a length of 6, 7, or 8 nucleotides that hybridize under defined conditions to the target population of nucleic acid molecules and do not hybridize under defined conditions to the non-target population of nucleic acid molecules in the population of single-stranded primer extension products.

2. The method of Claim 1, wherein the hybridizing portion of the second population of oligonucleotide primers is selected to include all possible oligonucleotides having a length of 6 nucleotides that do not hybridize under the defined conditions to the non-target nucleic acid population in the population of single-stranded primer extension products.

3. The method of Claim 1, wherein the non-target population of nucleic acid molecules consists essentially of the most abundant nucleic acid molecules in the population of RNA template molecules.

4. The method of Claim 3, wherein the most abundant nucleic acid molecules are selected from the group consisting of ribosomal RNA, mitochondrial ribosomal RNA, and a combination thereof.

5. The method of Claim 1, wherein the hybridizing portion of the first population of oligonucleotides consists of one of 6, 7, 8, or 9 random nucleotides and the defined sequence portion comprises a first primer binding site for PCR amplification.

6. The method of Claim 1, wherein the population of hybridizing portions in the first population of oligonucleotide primers is selected from all possible oligonucleotides having a length of 6 nucleotides that do not hybridize under defined conditions to the non-target nucleic acid molecules in the population of RNA template molecules.

7. The method of Claim 1, wherein the sample comprises total RNA.

8. The method of Claim 1, wherein the defined sequence portion of each oligonucleotide in the first and second population of oligonucleotides consists of a primer binding site for PCR amplification ranging in length from 10 nucleotides to 20 nucleotides.

9. The method of Claim 8, wherein at least one of the first or second primer binding sites comprises a transcriptional promoter.

10. The method of Claim 1, wherein each oligonucleotide in the second population of oligonucleotides further comprises a spacer sequence portion consisting of from 1 to 10 random nucleotides, wherein the spacer portion is located between the defined sequence portion and the hybridizing portion.

11. The method of Claim 1, wherein the population of hybridizing portions in the second population of oligonucleotides is selected from the oligonucleotides comprising SEQ ID NOS:750-1498.

12. The method of Claim 6, wherein the population of hybridizing portions in the first population of oligonucleotides is selected from the oligonucleotides comprising SEQ ID NOS:1-749.

13. The method of Claim 8, further comprising amplifying at least one strand of the double stranded cDNA.

14. The method of Claim 13, further comprising sequencing the PCR amplified DNA.

15. The method of Claim 8, wherein the defined sequence portion of each oligonucleotide in the first population comprises a region of at least 8 consecutive nucleotides that are identical to a region of at least 8 consecutive nucleotides in the defined sequence portion of each oligonucleotide in the second population.

16. The method of Claim 8, wherein the defined sequence portion of at least one of the first or second population of oligonucleotides comprises an RNA portion and a DNA portion, wherein the RNA portion is 5' with respect to the DNA portion.

17. A method of transcriptome profiling comprising:

(a) synthesizing a population of single-stranded primer extension products from a target population of nucleic acid molecules within a population of RNA template molecules in a sample isolated from a mammalian subject using reverse transcriptase enzyme and a first population of oligonucleotide primers comprising a hybridizing portion and a first PCR primer binding site located 5' to the hybridizing portion;

(b) synthesizing double-stranded cDNA from the population of single-stranded primer extension products generated according to step (a) using a DNA polymerase and a second population of oligonucleotide primers comprising a hybridizing portion and a second PCR primer binding site located 5' to the hybridizing portion,

wherein the hybridizing portion is selected from all possible oligonucleotides having a length of 6 nucleotides that hybridize under defined conditions to the target population of nucleic acid molecules and do not hybridize under defined conditions to the non-target population of nucleic acid molecules in the population of single-stranded primer extension products, wherein the non-target population of nucleic acid molecules consists essentially of ribosomal RNA and mitochondrial ribosomal RNA of the same species as the mammalian subject; and

(c) PCR amplifying the double-stranded cDNA synthesized according to step (b) using a first PCR primer that binds to the first PCR primer binding site and a second PCR primer that binds to the second PCR primer binding site.

18. The method of Claim 17, further comprising cloning the PCR products into a vector to generate a library representative of the transcriptome of the mammalian subject at the time the sample was isolated.

19. The method of Claim 17, further comprising sequencing at least a portion of the PCR products.

20. The method of Claim 17, wherein the PCR amplification is carried out using at least 2 cycles of amplification with an annealing temperature between 40 to 50 degrees followed by additional amplification cycles with an annealing temperature of greater than 50 degrees.

21. The method of Claim 17, further comprising labeling at least a portion of the amplified PCR products.

22. The method of Claim 17, wherein the first PCR primer binding site of each oligonucleotide in the first population comprises a region of at least 8 consecutive nucleotides that are identical to a region of at least 8 consecutive nucleotides in the second PCR primer binding site of each oligonucleotide in the second population of oligonucleotides.

23. The method of Claim 17, wherein the PCR primer binding site of at least one of the first or second population of oligonucleotides comprises an RNA portion and a DNA portion, wherein the RNA portion is 5' with respect to the DNA portion.

24. A population of amplified nucleic acid molecules generated using the method of Claim 17.

25. A method of selectively amplifying a target population of nucleic acid molecules within a larger non-target population of nucleic acid molecules, the method comprising the steps of:

(a) synthesizing single-stranded cDNA from a sample comprising total RNA isolated from a mammalian subject using reverse transcriptase enzyme and a first population of oligonucleotide primers, and wherein each oligonucleotide within the first population of oligonucleotide primers comprises a hybridizing portion and a defined sequence portion located 5' to the hybridizing portion, wherein the hybridizing portion is a member of the population of oligonucleotides comprising SEQ ID NOS:1-749; and

(b) synthesizing double-stranded cDNA from the single-stranded cDNA synthesized according to step (a) using a DNA polymerase and a second population of oligonucleotide primers, wherein each oligonucleotide within the second population of oligonucleotide primers comprises a hybridizing portion and a defined sequence portion located 5' to the hybridizing portion, and wherein the hybridizing portion is a member of the population of oligonucleotides comprising SEQ ID NOS:750-1498.

26. The method of Claim 25, wherein the population of hybridizing portions of the first population of oligonucleotide primers comprises at least 10% of the oligonucleotides comprising SEQ ID NOS:1-749.

27. The method of Claim 25, wherein the population of hybridizing portions of the second population of oligonucleotide primers comprises at least 10% of the oligonucleotides comprising SEQ ID NOS:750-1498.

28. The method of Claim 25, further comprising sequencing at least a portion of the PCR products.

29. The method of Claim 25, further comprising labeling at least a portion of the PCR products.

30. A population of oligonucleotides comprising SEQ ID NOS:1-749 for use in first strand cDNA synthesis.

31. A population of oligonucleotides comprising SEQ ID NOS:750-1498 for use in second strand cDNA synthesis.

32. A reagent for selectively amplifying a target population of nucleic acid molecules, the reagent comprising at least 10% of the oligonucleotides comprising SEQ ID NOS:1-749.

33. A reagent for selectively amplifying a target population of nucleic acid molecules, the reagent comprising at least 10% of the oligonucleotides comprising SEQ ID NOS:750-1498.

34. A reagent for selectively amplifying a target population of nucleic acid molecules, the reagent comprising a population of oligonucleotides to prime the amplification of a target population of nucleic acid molecules wherein each oligonucleotide comprises a hybridizing portion and a defined sequence portion located 5' to the hybridizing portion, wherein the hybridizing portion is a member of the population of oligonucleotides comprising SEQ ID NOS:1-749.

35. A reagent for selectively amplifying a target population of nucleic acid molecules, the reagent comprising a population of oligonucleotides to prime the amplification of a target population of nucleic acid molecules wherein each oligonucleotide comprises a hybridizing portion and a defined sequence portion located 5' to the hybridizing portion, wherein the hybridizing portion is a member of the population of oligonucleotides comprising SEQ ID NOS:750-1498.

36. A kit for selectively amplifying a target population of nucleic acid molecules, the kit comprising a reagent comprising a first population of oligonucleotides for first strand cDNA synthesis wherein each oligonucleotide in the first population of oligonucleotides comprises a hybridizing portion and a defined sequence portion located

5' to the hybridizing portion, wherein the hybridizing portion is a member of the population of oligonucleotides comprising SEQ ID NOS:1-749.

37. The kit of Claim 36, wherein the population of hybridizing portions in the first population of oligonucleotides comprises at least 10% of the oligonucleotides comprising SEQ ID NOS:1-749.

38. The kit of Claim 36, further comprising a second population of oligonucleotides for second strand cDNA synthesis, wherein each oligonucleotide in the second population of oligonucleotides comprises a hybridizing portion and a defined sequence portion located 5' to the hybridizing portion, wherein the hybridizing portion is a member of the population of oligonucleotides comprising SEQ ID NOS:750-1498.

39. The kit of Claim 38, wherein the population of hybridizing portions in the second population of oligonucleotides comprises at least 10% of the oligonucleotides comprising SEQ ID NOS:750-1498.

40. The kit of Claim 38, wherein the population of hybridizing portions in the first population of oligonucleotides comprises the oligonucleotides consisting of SEQ ID NOS:1-749 and wherein the population of hybridizing portions in the second population of oligonucleotides comprises the oligonucleotides consisting of SEQ ID NOS:750-1498.

41. The kit of Claim 38, further comprising at least one of the following components: a reverse transcriptase, a DNA polymerase, a DNA ligase, a RNase H enzyme, a Tris buffer, a potassium salt, a magnesium salt, an ammonium salt, a reducing agent, deoxynucleoside triphosphates, or a ribonuclease inhibitor.

42. A kit for selectively amplifying a target population of nucleic acid molecules within a population of RNA template molecules in a sample obtained from a mammalian subject, the kit comprising:

(a) a first population of oligonucleotide primers comprising a hybridizing portion consisting of 6 nucleotides selected from all possible oligonucleotides having a length of 6 nucleotides that do not hybridize under defined conditions to the non-target population of nucleic acid molecules in the population of RNA template

molecules, and a defined sequence portion located 5' to the hybridizing portion, wherein the non-target population of nucleic acid molecules consists essentially of the most abundant nucleic acid molecules in the population of RNA template molecules;

(b) a second population of oligonucleotide primers comprising a hybridizing portion consisting of 6 nucleotides selected from the reverse complement of the nucleotide sequence of the hybridizing portion of the first population of oligonucleotide primers, and a defined sequence portion located 5' to the hybridizing portion;

(c) a first PCR primer that binds to the first defined sequence portion of the first population of oligonucleotides and a second PCR primer that binds to the second defined sequence portion of the second population of oligonucleotides.

43. The kit of Claim 42, wherein the non-target population of nucleic acid molecules consists essentially of ribosomal RNA and mitochondrial ribosomal RNA of the same species as the mammalian subject.

44. The kit of Claim 42, wherein the defined sequence portion of each oligonucleotide in the first and second population of oligonucleotides consists of a primer binding site for PCR amplification ranging in length from 10 nucleotides to 20 nucleotides.

45. The kit of Claim 42, wherein the defined sequence portion of each oligonucleotide in the first population comprises a region of at least 8 consecutive nucleotides that are identical to a region of at least 8 consecutive nucleotides in the defined sequence portion of each oligonucleotide in the second population.

46. The kit of Claim 42, wherein the defined sequence portion of at least one of the first or second population of oligonucleotides comprises an RNA portion and a DNA portion, wherein the RNA portion is 5' with respect to the DNA portion.

47. A method of selectively amplifying a target population of nucleic acid molecules to generate amplified DNA molecules, the method comprising the steps of:

(a) providing a first population of oligonucleotides wherein each oligonucleotide comprises a hybridizing portion and a first PCR primer binding site

located 5' to the hybridizing portion, wherein the hybridizing portion is a member of the population of oligonucleotides comprising SEQ ID NOS:1-749;

(b) annealing the first population of oligonucleotides to a sample comprising RNA isolated from a mammalian subject;

(c) synthesizing cDNA from the RNA using a reverse transcriptase enzyme;

(d) synthesizing double-stranded cDNA using a DNA polymerase and a second population of oligonucleotides, wherein each oligonucleotide comprises a hybridizing portion and a second PCR binding site located 5' to the hybridizing portion, wherein the hybridizing portion is a member of the population of oligonucleotides comprising SEQ ID NOS:750-1498;

(e) PCR amplifying the double stranded cDNA using thermostable DNA polymerase, a first PCR primer that binds to the first PCR primer binding site and a second PCR primer that binds to the second PCR primer binding site to generate amplified double stranded DNA; and

(f) sequencing the amplified double-stranded PCR products.

48. A population of selectively amplified nucleic acid molecules consisting of a representation of a target population of nucleic acid molecules within a population of RNA template molecules in a cell sample isolated from a mammalian subject, each amplified nucleic acid molecule comprising:

a 5' defined sequence portion flanking a member of the population of amplified nucleic acid sequences, and a 3' defined sequence, wherein the population of selectively amplified sequences includes an amplified nucleic acid sequence corresponding to a target RNA molecule expressed in the mammalian cell, and is characterized by having the following properties with reference to the particular mammalian species:

(a) having greater than 75% polyadenylated and non-polyadenylated transcripts; and having less than 10% ribosomal RNA.

49. The population of Claim 48 inserted into a cloning vector.

50. The population of Claim 48, wherein each nucleic acid molecule in the population is labeled.

51. The population of Claim 48 attached to a substrate.

52. The population of Claim 48, wherein the defined sequence portion of at least one of the first or second population of oligonucleotides comprises an RNA portion and a DNA portion, wherein the RNA portion is 5' with respect to the DNA portion.

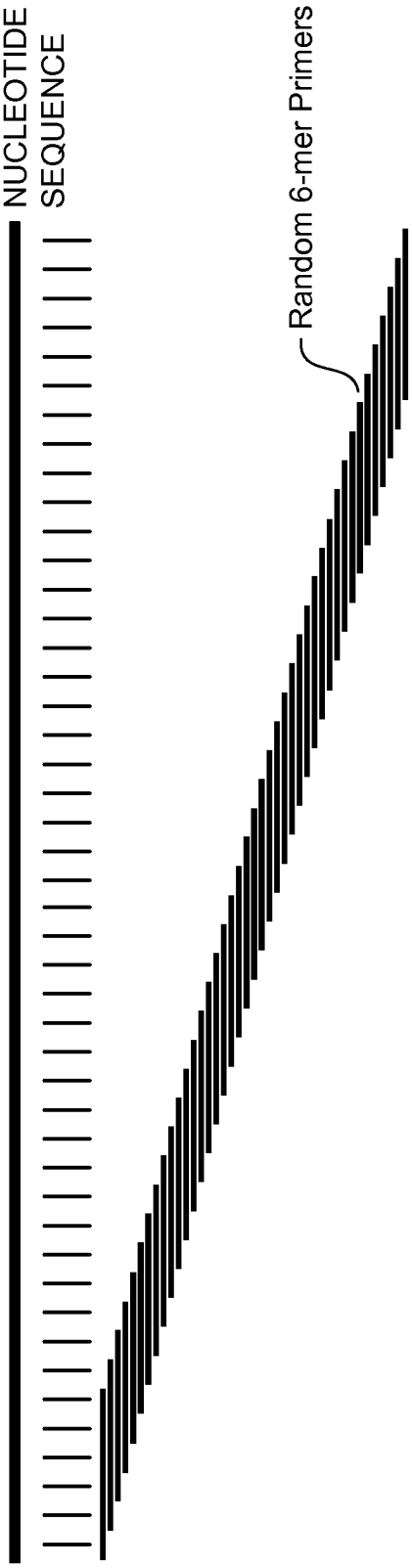
NUCLEOTIDE SEQUENCE
Random 6-mer Primers

Fig.1A.

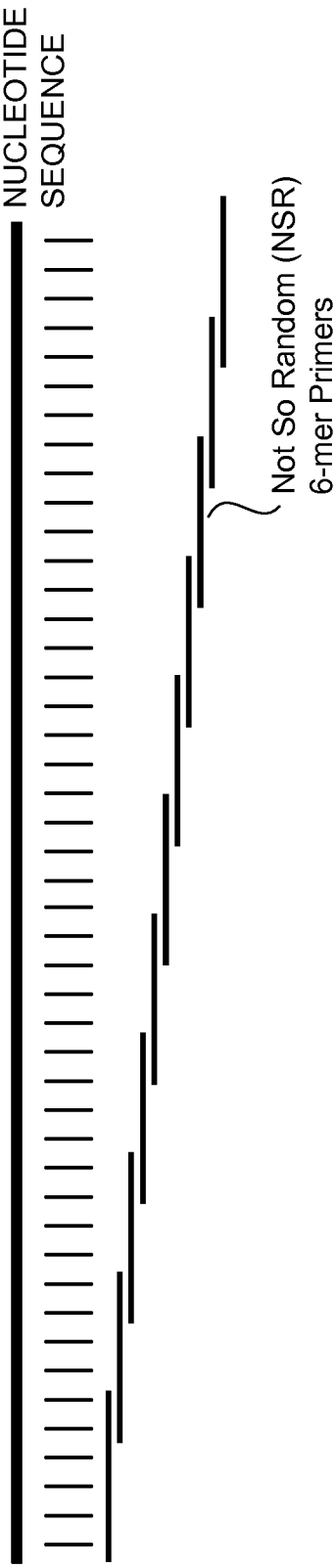
NUCLEOTIDE SEQUENCE
Not So Random (NSR) 6-mer Primers

Fig.1B.

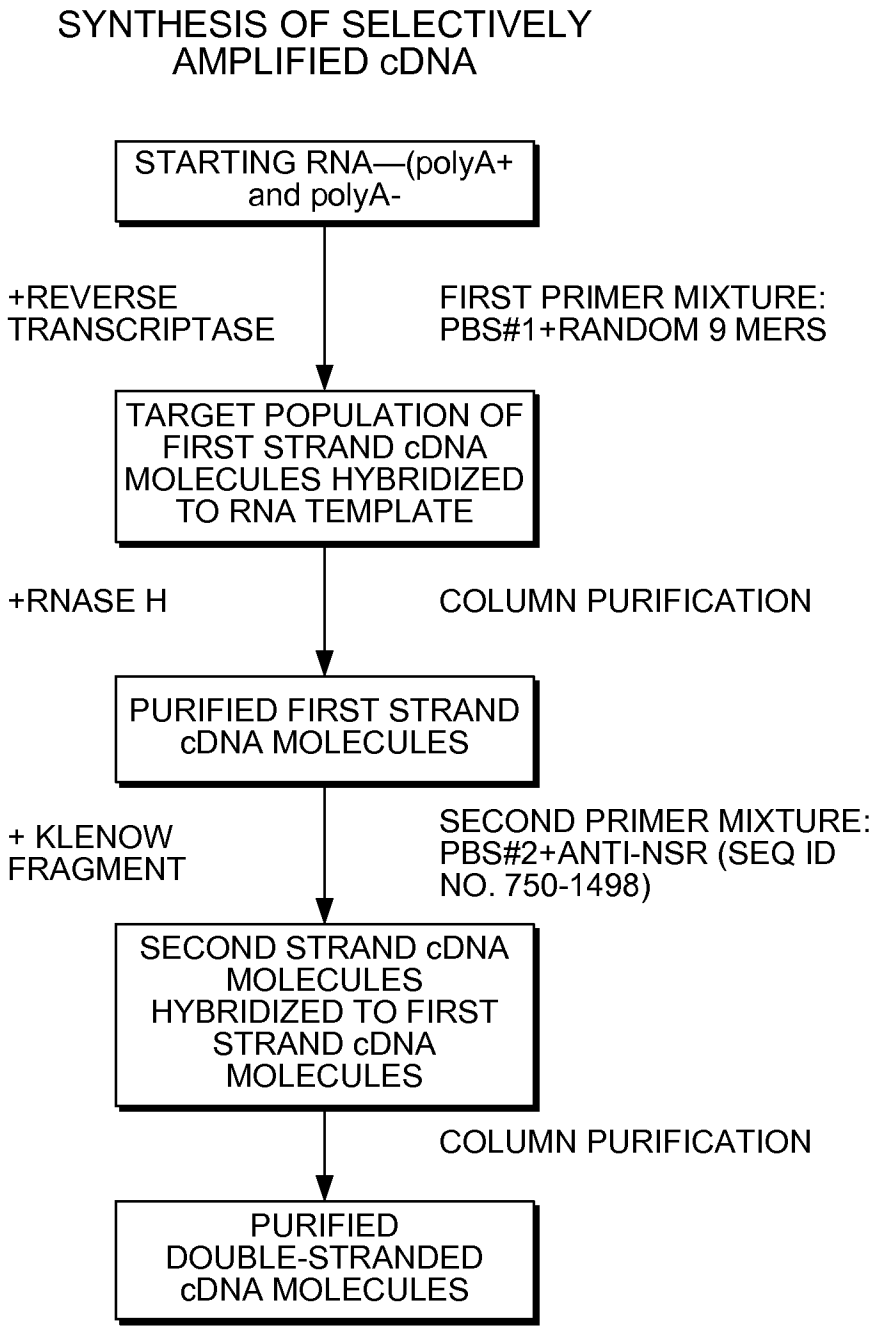
SYNTHESIS OF SELECTIVELY AMPLIFIED cDNA
STARTING RNA—(polyA+ and polyA-
+REVERSE TRANSCRIPTASE
FIRST PRIMER MIXTURE: PBS#1+RANDOM 9 MERS
TARGET POPULATION OF FIRST STRAND cDNA MOLECULES HYBRIDIZED TO RNA TEMPLATE
+RNASE H
COLUMN PURIFICATION
PURIFIED FIRST STRAND cDNA MOLECULES
+ KLENOW FRAGMENT
SECOND PRIMER MIXTURE: PBS#2+ANTI-NSR (SEQ ID NO. 750-1498)
SECOND STRAND cDNA MOLECULES HYBRIDIZED TO FIRST STRAND cDNA MOLECULES
COLUMN PURIFICATION
PURIFIED DOUBLE-STRANDED cDNA MOLECULES

Fig.1C.

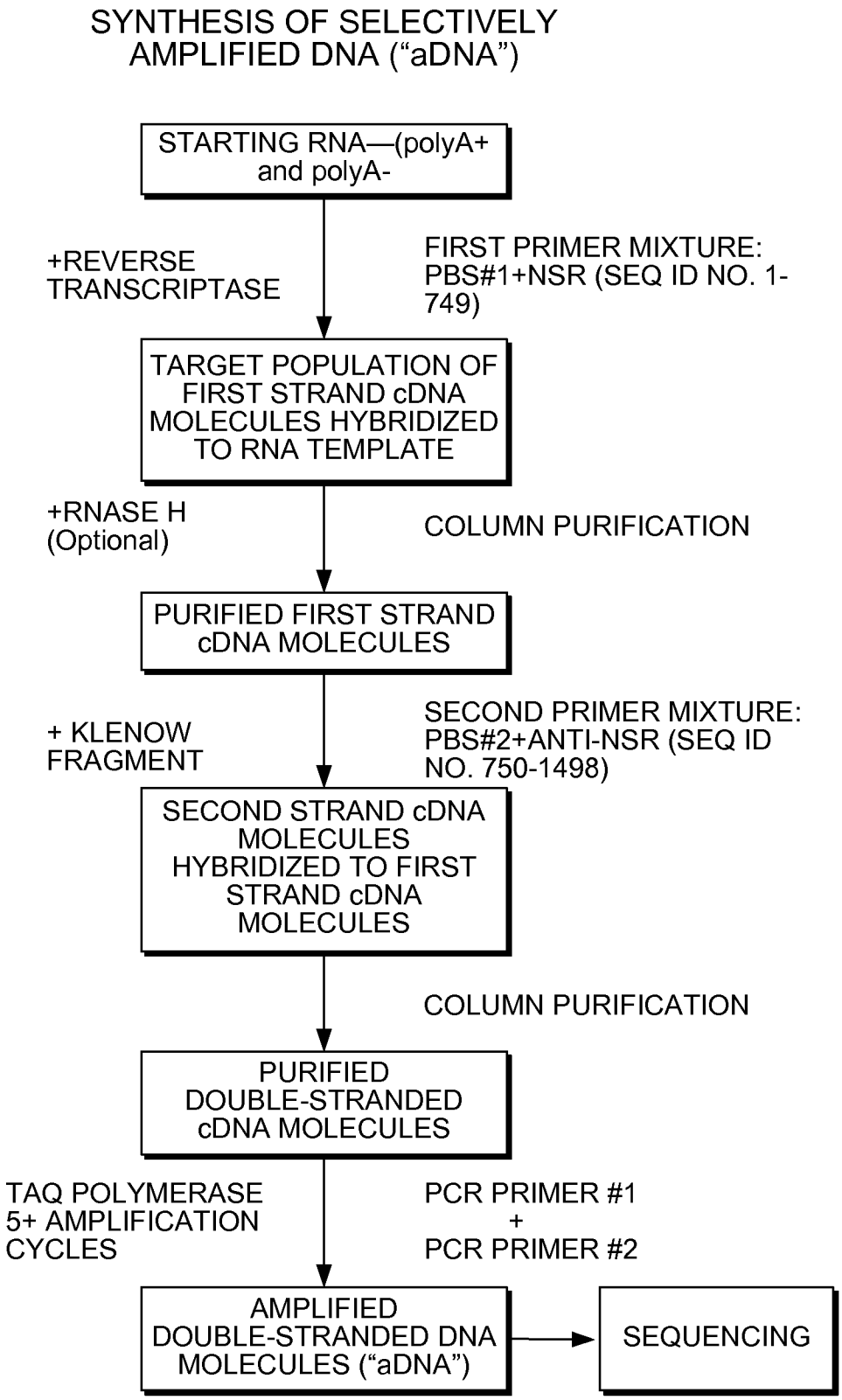
SYNTHESIS OF SELECTIVELY AMPLIFIED DNA ("aDNA")
STARTING RNA—(polyA+ and polyA-
+REVERSE TRANSCRIPTASE
FIRST PRIMER MIXTURE: PBS#1+NSR (SEQ ID NO. 1-749)
TARGET POPULATION OF FIRST STRAND cDNA MOLECULES HYBRIDIZED TO RNA TEMPLATE
+RNASE H (Optional)
COLUMN PURIFICATION
PURIFIED FIRST STRAND cDNA MOLECULES
+ KLENOW FRAGMENT
SECOND PRIMER MIXTURE: PBS#2+ANTI-NSR (SEQ ID NO. 750-1498)
SECOND STRAND cDNA MOLECULES HYBRIDIZED TO FIRST STRAND cDNA MOLECULES
COLUMN PURIFICATION
PURIFIED DOUBLE-STRANDED cDNA MOLECULES
TAQ POLYMERASE 5+ AMPLIFICATION CYCLES
PCR PRIMER #1 + PCR PRIMER #2
AMPLIFIED DOUBLE-STRANDED DNA MOLECULES ("aDNA")
SEQUENCING

Fig.1D.

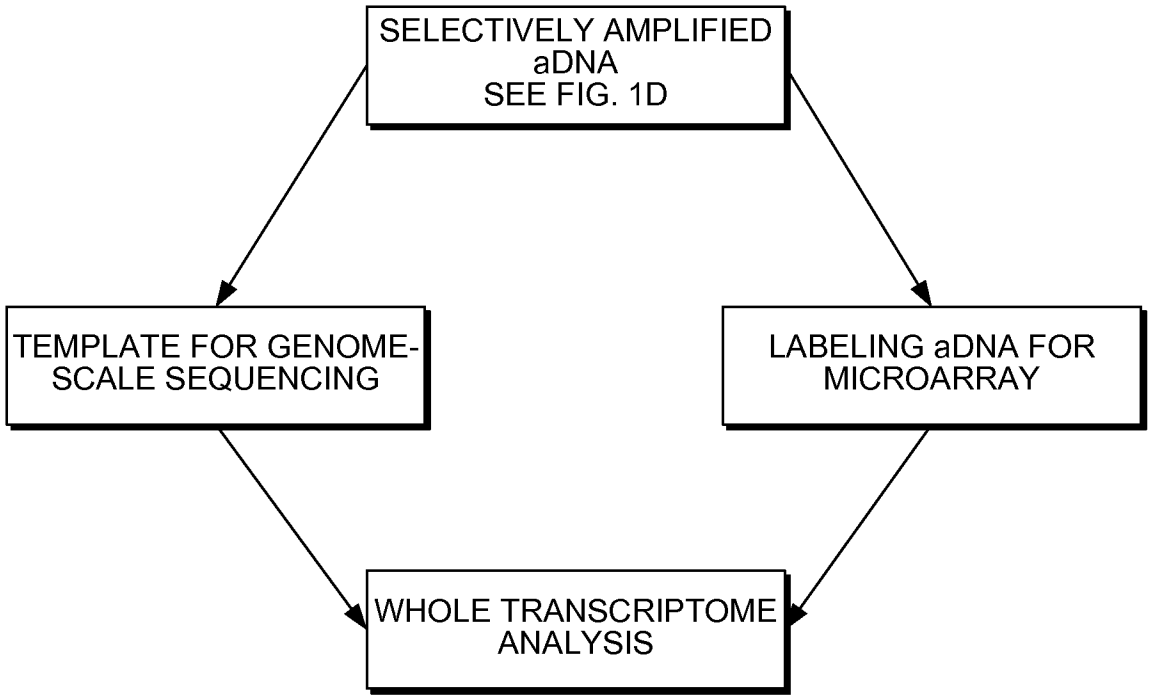
SELECTIVELY AMPLIFIED aDNA SEE FIG. 1D
TEMPLATE FOR GENOME-SCALE SEQUENCING
LABELING aDNA FOR MICROARRAY
WHOLE TRANSCRIPTOME ANALYSIS

Fig.2.

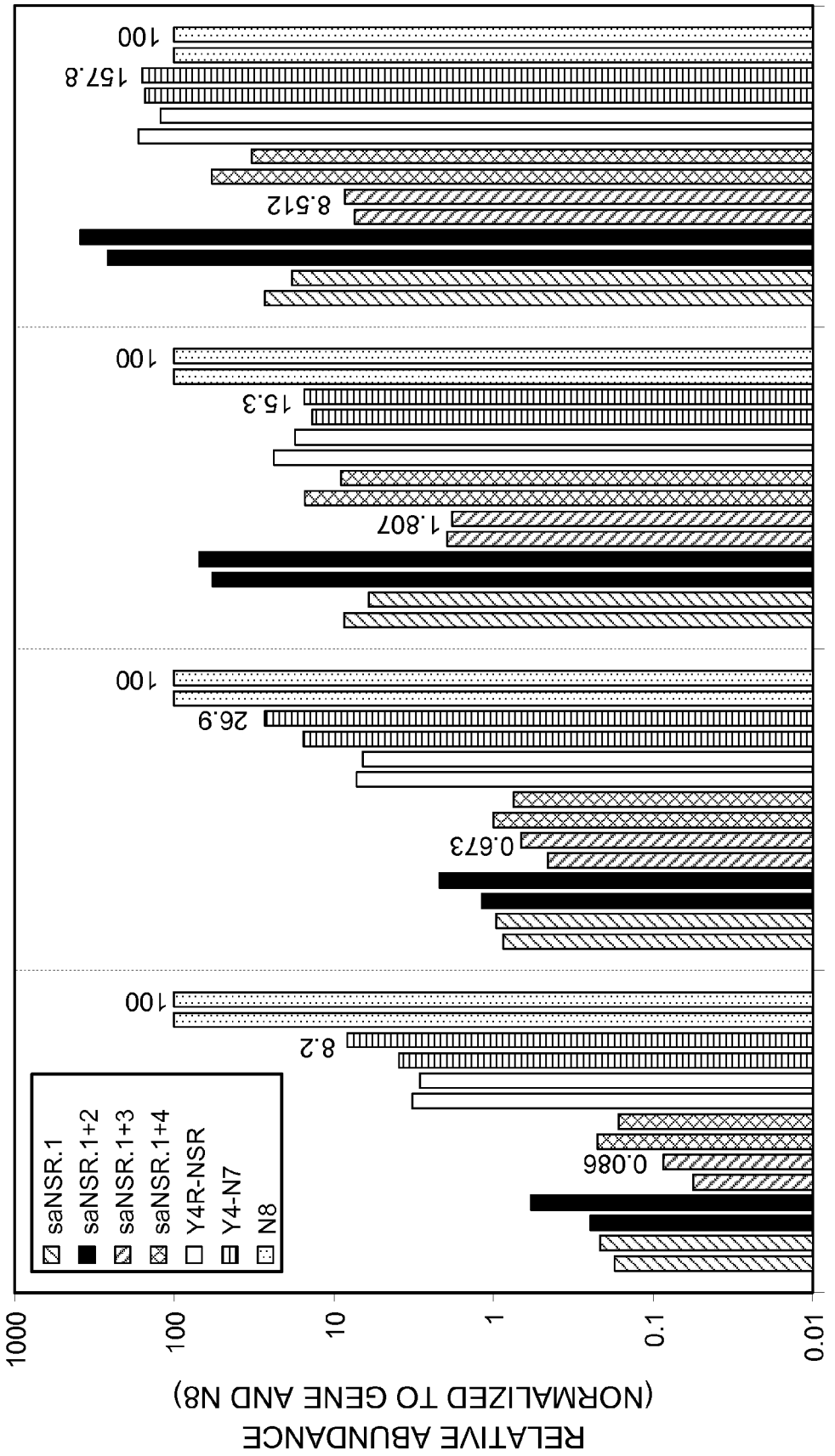
RELATIVE ABUNDANCE
(NORMALIZED TO GENE AND N8)
1000
100
10
1
0.1
0.01
saNSR.1
saNSR.1+2
saNSR.1+3
saNSR.1+4
Y4R-NSR
Y4-N7
N8
18S_all
28S_all
12S_all
16S_all
100
8.2
0.086
100
26.9
0.673
100
1.807
15.3
100
8.512
157.8
100

Fig.3A.

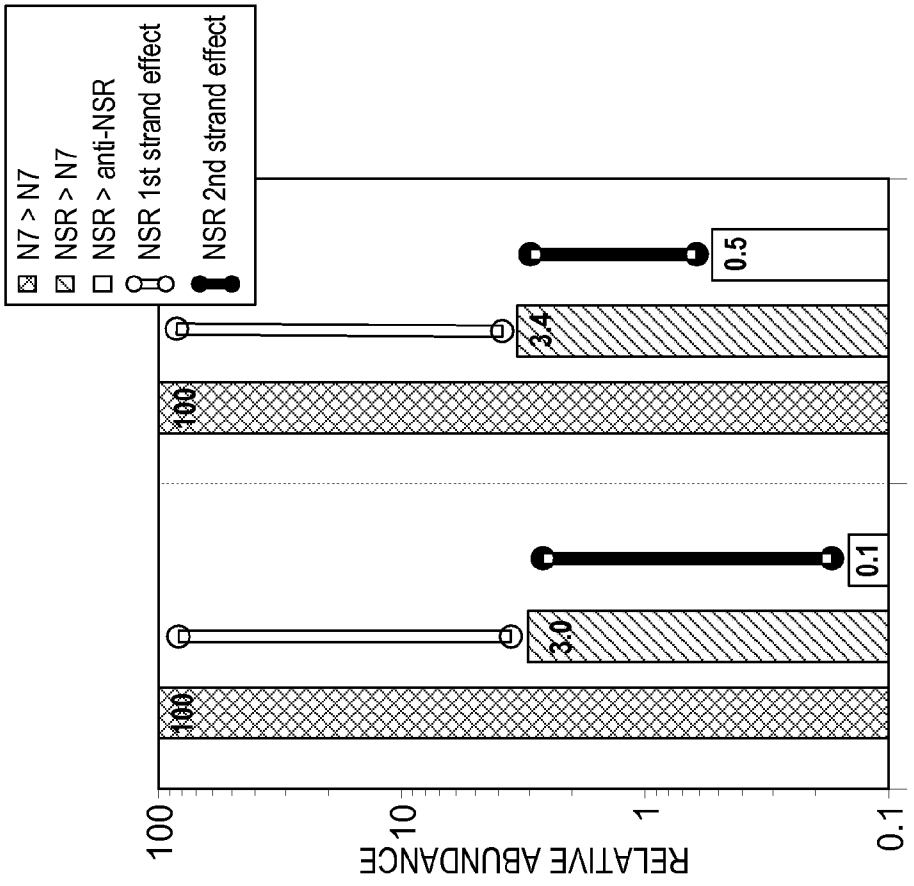
N7 > N7
NSR > N7
NSR > anti-NSR
NSR 1st strand effect
NSR 2nd strand effect
RELATIVE ABUNDANCE
100
10
1
0.1
100
3.0
0.1
100
3.4
0.5
18S
28S

Fig.3B.

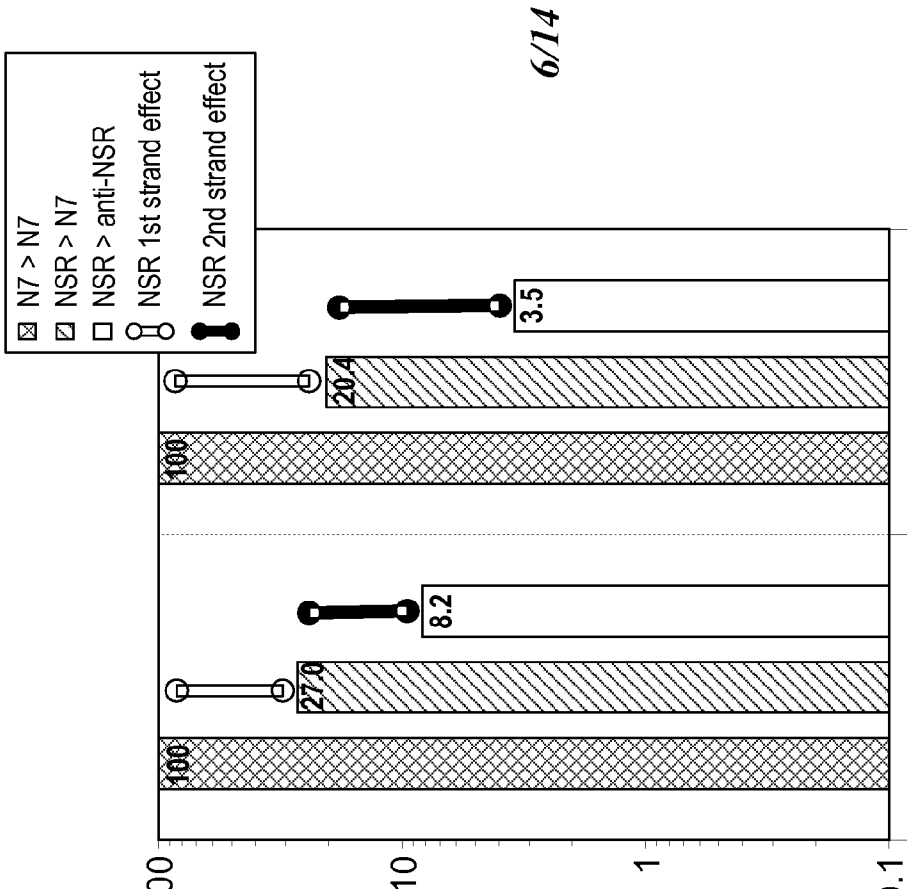
N7 > N7
NSR > N7
NSR > anti-NSR
NSR 1st strand effect
NSR 2nd strand effect
RELATIVE ABUNDANCE
100
10
1
0.1
100
27.0
8.2
100
20.4
3.5
12S
16S

Fig.3C.

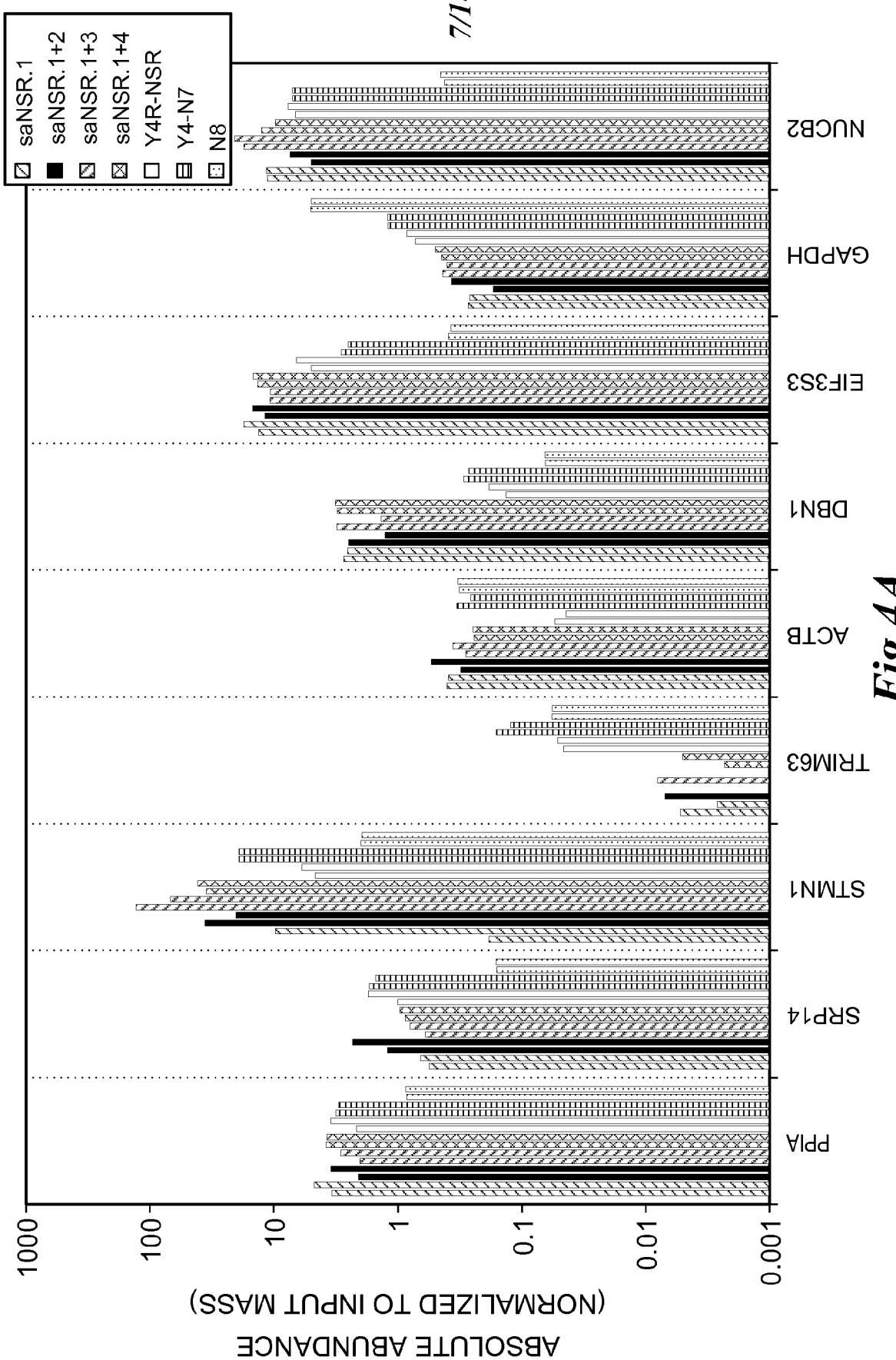
saNSR.1
saNSR.1+2
saNSR.1+3
saNSR.1+4
Y4R-NSR
Y4-N7
N8
ABSOLUTE ABUNDANCE
(NORMALIZED TO INPUT MASS)
1000
100
10
1
0.1
0.01
0.001
PPIA
SRP14
STMN1
TRIM63
ACTB
DBN1
EIF3S3
GAPDH
NUCB2

Fig. 4A.

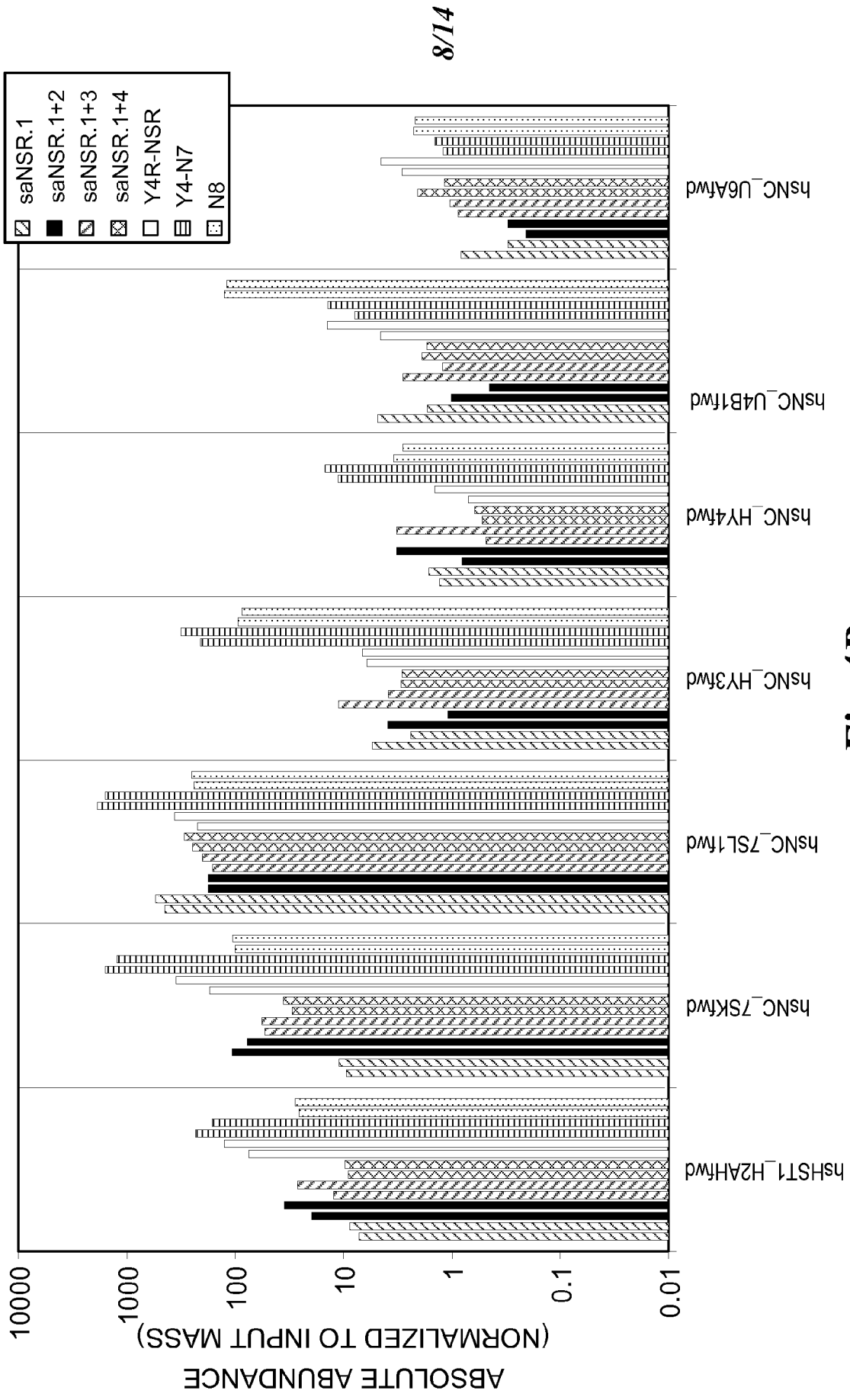
saNSR.1
saNSR.1+2
saNSR.1+3
saNSR.1+4
Y4R-NSR
Y4-N7
N8
ABSOLUTE ABUNDANCE
(NORMALIZED TO INPUT MASS)
10000
1000
100
10
1
0.1
0.01
hsHST1_H2AHfwd
hsNC_7SKfwd
hsNC_7SL1fwd
hsNC_HY3fwd
hsNC_HY4fwd
hsNC_U4B1fwd
hsNC_U6Afwd

Fig.4B.

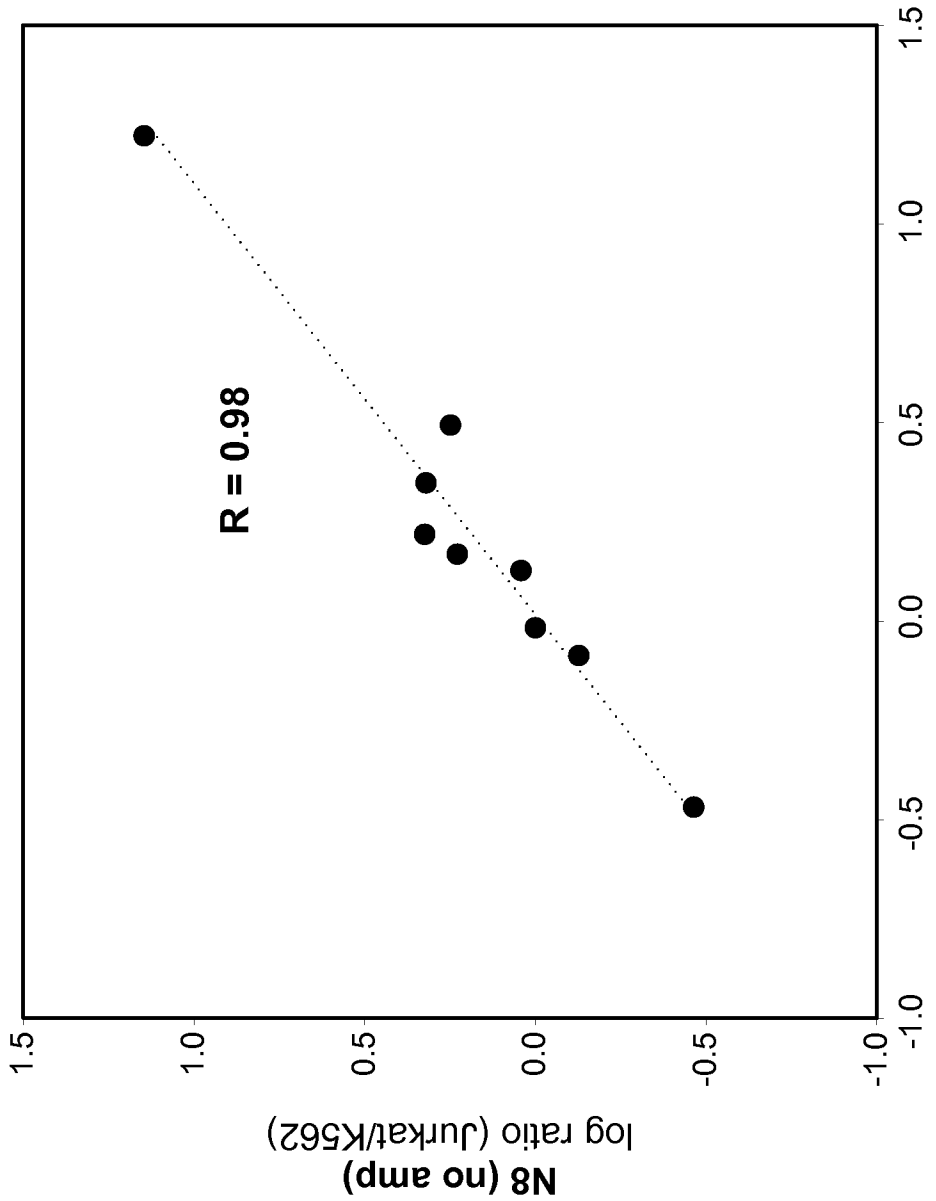
R = 0.98
N8 (no amp)
log ratio (Jurkat/K562)
saNSR1+3
log ratio (Jurkat/K562)
1.5
1.0
0.5
0.0
-0.5
-1.0

Fig. 5.

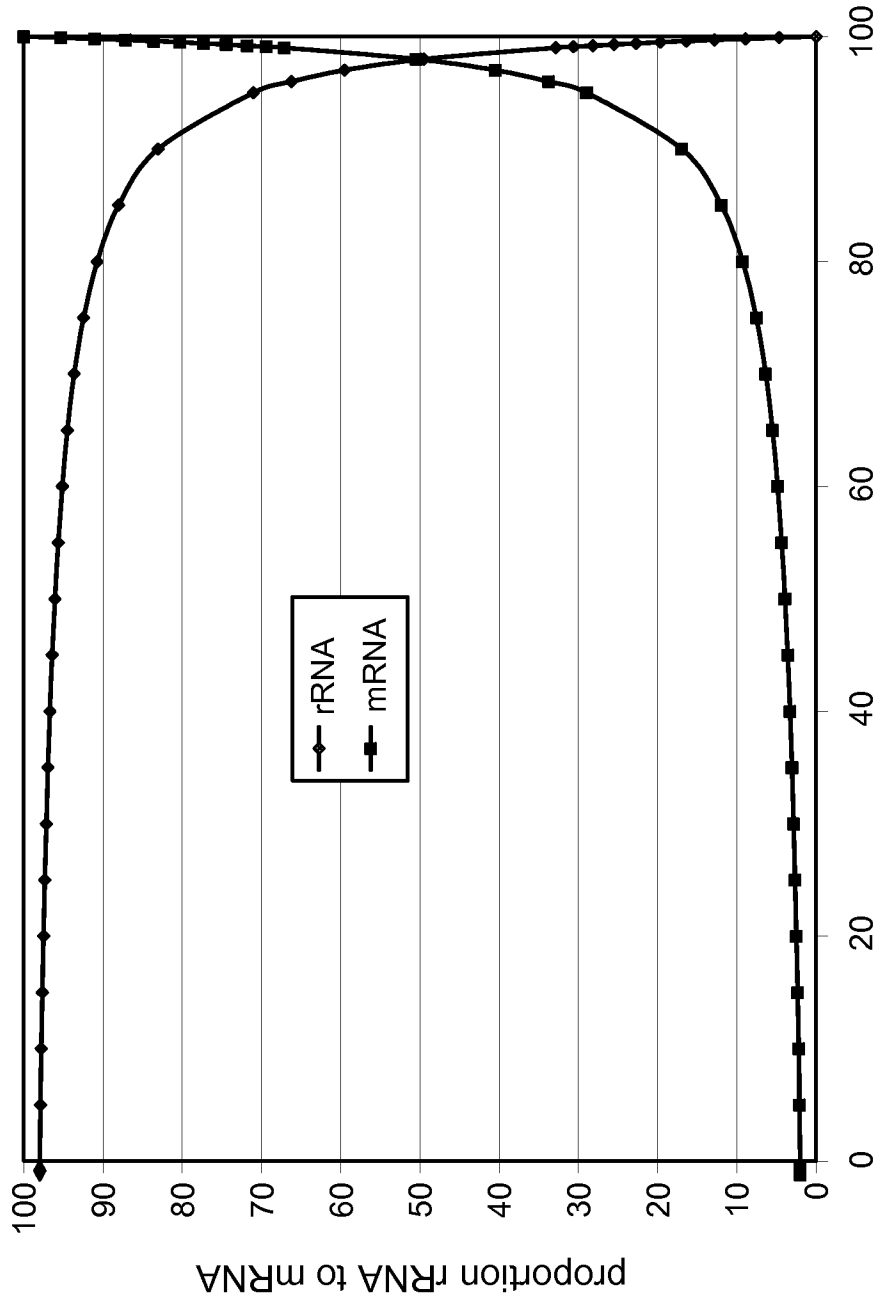
proportion rRNA to mRNA
100
90
80
70
60
50
40
30
20
10
0
rRNA
mRNA
0
20
40
60
80
100
% rRNA removal

Fig. 6A.

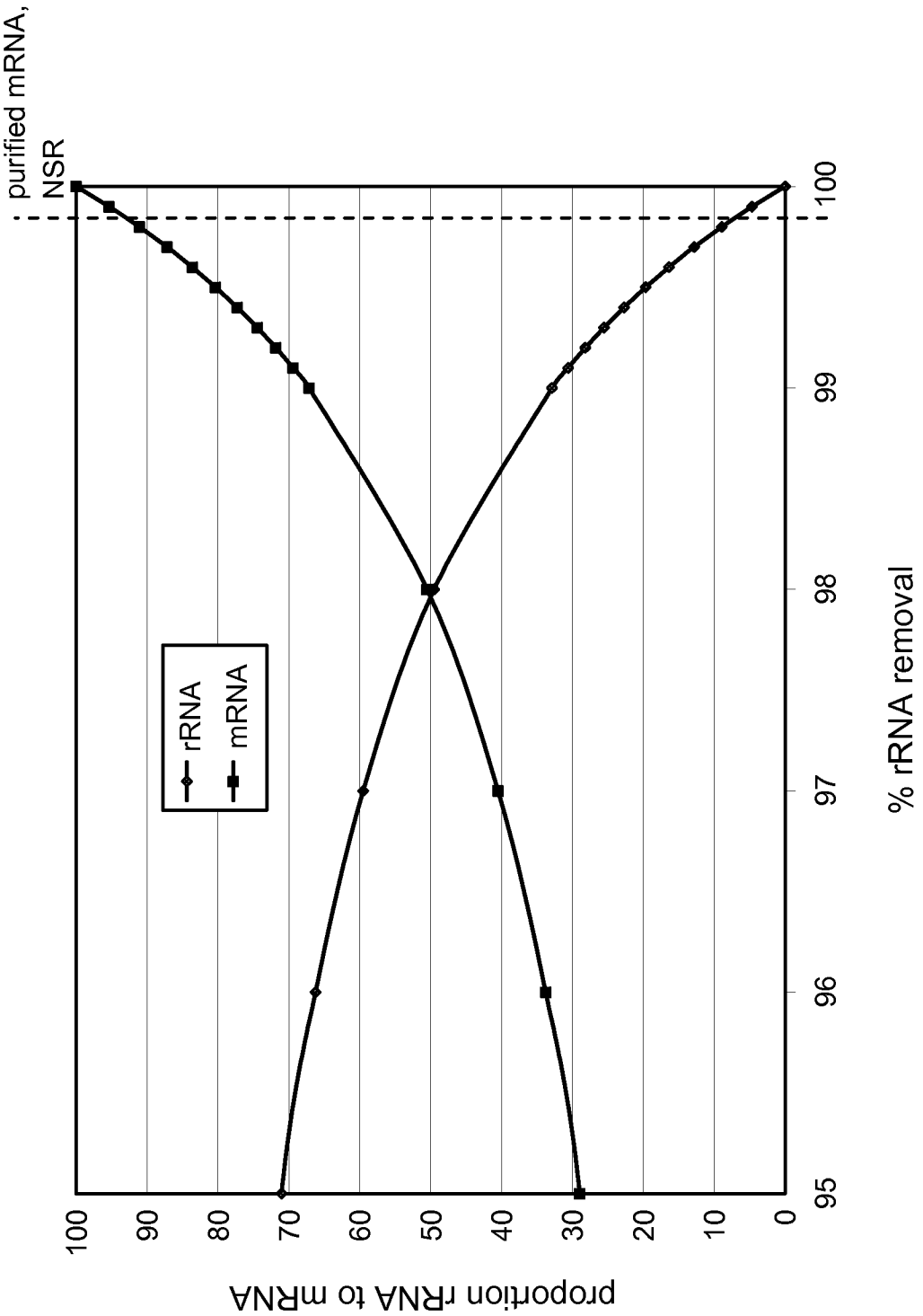
purified mRNA,
NSR
rRNA
mRNA
proportion rRNA to mRNA
100
90
80
70
60
50
40
30
20
10
0
95
96
97
98
99
100
% rRNA removal

Fig.6B.

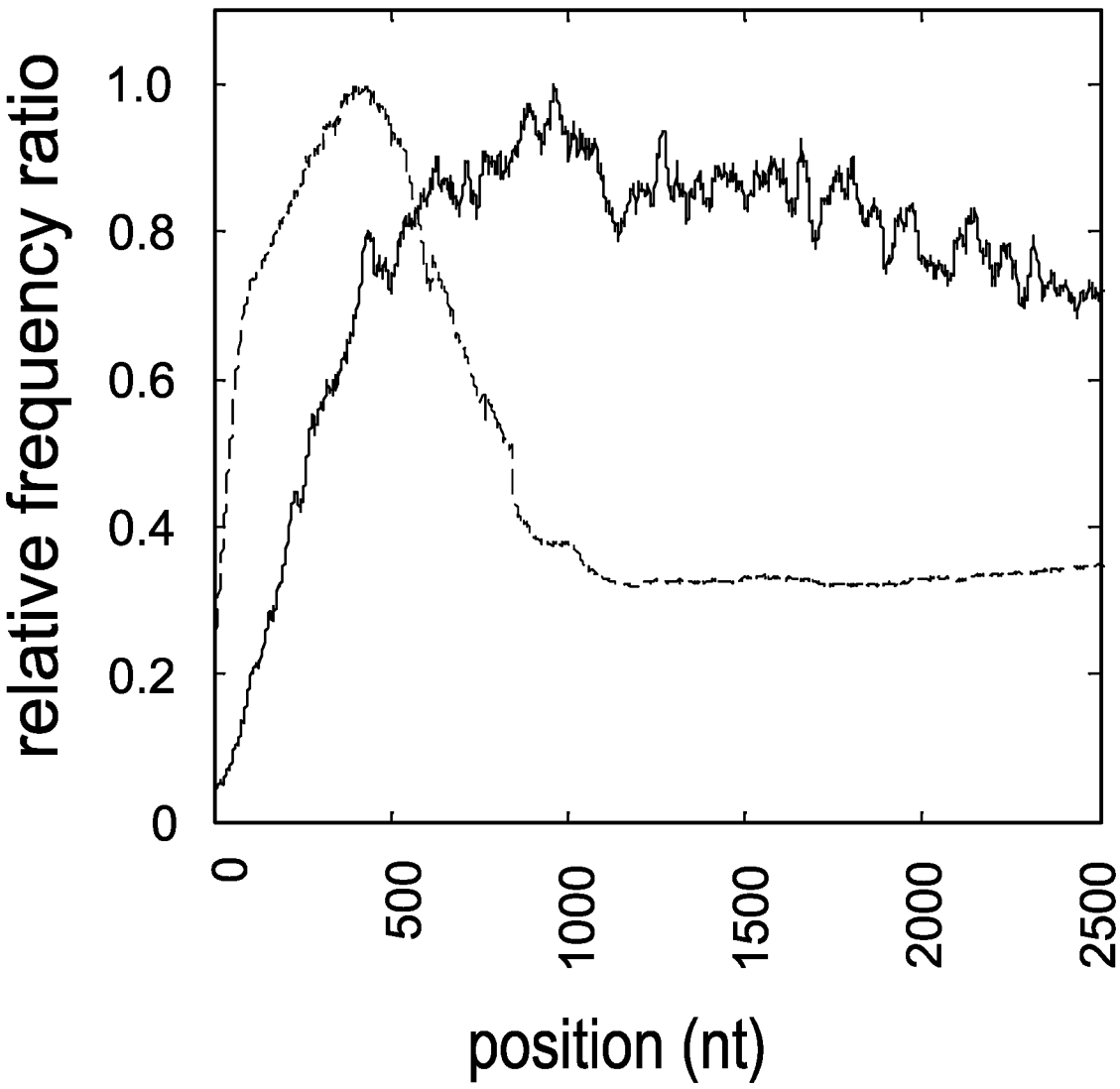
relative frequency ratio
1.0
0.8
0.6
0.4
0.2
0
0
500
1000
1500
2000
2500
position (nt)

Fig. 7A.

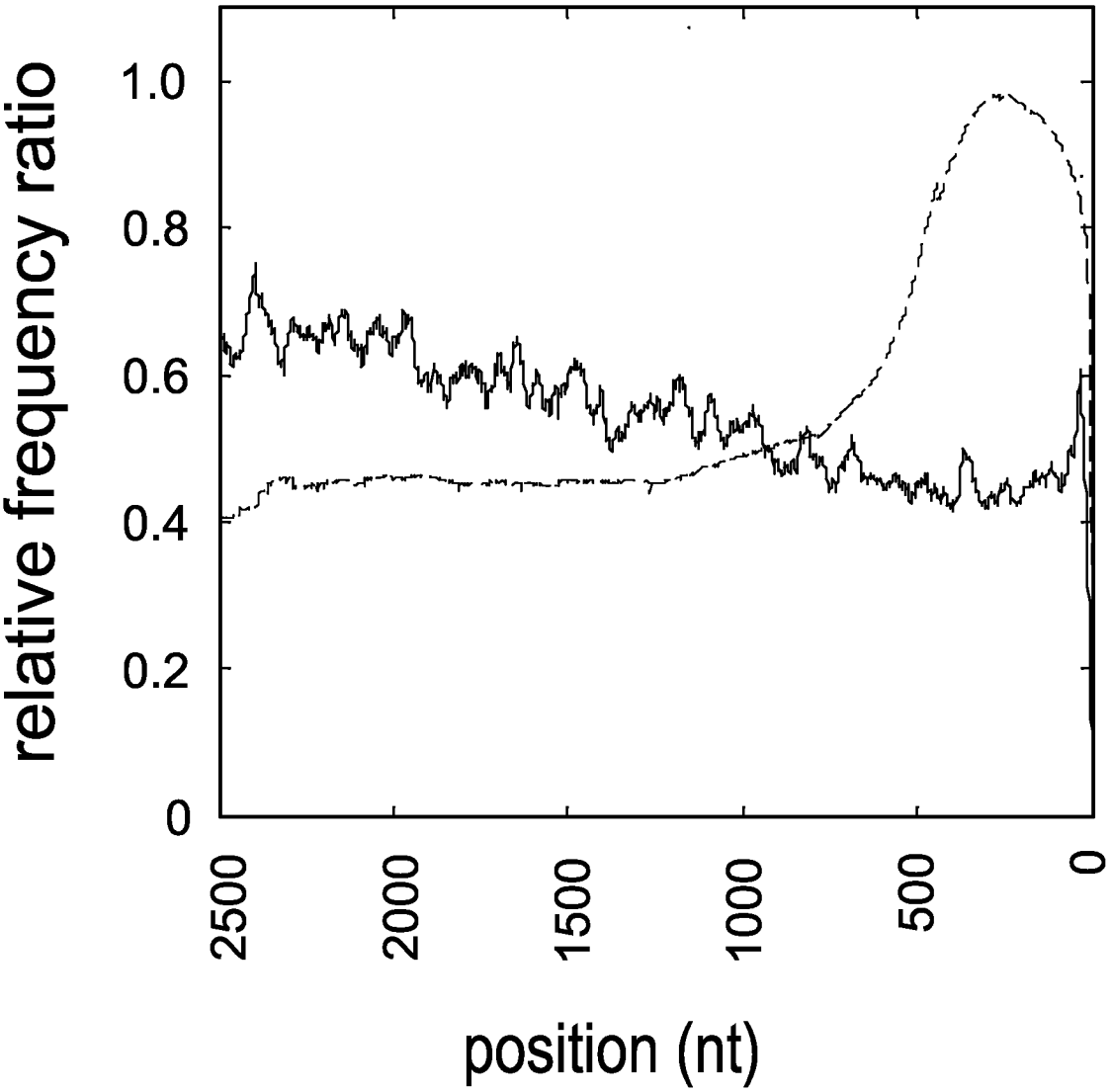
relative frequency ratio
1.0
0.8
0.6
0.4
0.2
0
2500
2000
1500
1000
500
0
position (nt)

Fig. 7B.

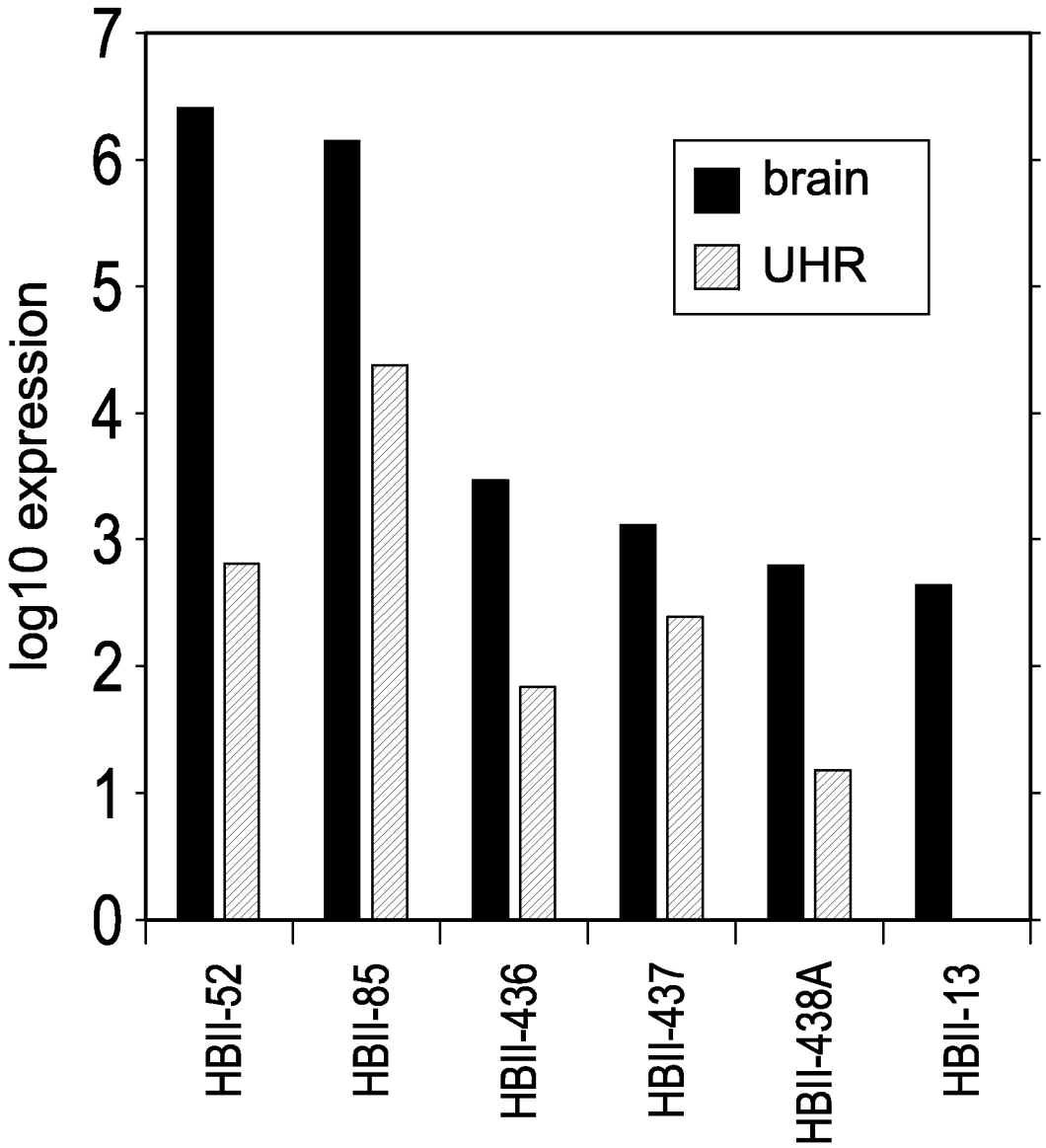
log10 expression
0
1
2
3
4
5
6
7
brain
UHR
HBII-52
HBII-85
HBII-436
HBII-437
HBII-438A
HBII-13

Fig.8.

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2008/081206

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68

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

B. FIELDS SEARCHED

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

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

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, BIOSIS, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2007/050990 A (ROSETTA INPHARMATICS LLC [US]; CASTLE JOHN [US]; RAYMOND CHRISTOPHER K) 3 May 2007 (2007-05-03) page 2, paragraph 3 - page 5, paragraph 2; claims 1,2,14,15,41,42 page 22, line 21 - line 33 ----- -/--	1-5, 7-11, 13-23, 31,33,35

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

[X] See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
4 February 2009	23/03/2009

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, Fax: (+31-70) 340-3016	Werner, Andreas

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2008/081206

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WONG K-K ET AL: "Use of tagged random hexamer amplification (TRHA) to clone and sequence minute quantities of DNA application to a 180 kb plasmid isolated from Sphingomonas F199" NUCLEIC ACIDS RESEARCH, OXFORD UNIVERSITY PRESS, SURREY, GB, vol. 24, no. 19, 1 October 1996 (1996-10-01), pages 3778-3783, XP008094799 ISSN: 0305-1048 the whole document	1-5, 7-11, 13-23, 31,33,35
A	ZHANG J ET AL: "Differential priming of RNA templates during cDNA synthesis markedly affects both accuracy and reproducibility of quantitative competitive reverse-transcriptase PCR" BIOCHEMICAL JOURNAL, THE BIOCHEMICAL SOCIETY, LONDON, vol. 337, no. PART 2, 15 January 1999 (1999-01-15), pages 231-241, XP002304124 ISSN: 0264-6021 the whole document	1-5, 7-11, 13-23, 31,33,35
A	EP 0 787 209 B (BIO MERIEUX [FR]) 20 March 2002 (2002-03-20) the whole document	15,16, 22,23

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box II.2

Claims Nos.: 1-10, 13-24 and 42-46 (in part)

Claims 1-10, 13-24 and 42-46 do not meet the requirements of Art. 6 PCT, because the result-to-be-achieved type of definition does not allow the scope of the claims to be ascertained: Said claims relate to hybridizing portions selected from all possible oligonucleotides of 6, 7, or 8 nucleotides that hybridize to the target but do not hybridize to the non-target population, without any indication about what these sequences might be. Any primer is normally designed to hybridize to a target, not to a non-target. Interpreting the claims in the light of the description, this definition merely amounts to a statement of the underlying problem of finding sequences that do not amplify the most abundant RNA. The hybridizing sequences potentially falling under the definition encompass up to 65536 different possible sequences (p.14 l.21-24), not all of which will solve the problem posed. Undue experimentation would be required to identify those sequences within a population of all possible sequences that discriminate between target and non-target populations. The skilled person would have no knowledge beforehand as to whether a given sequence of the prior art would fall within the claimed scope, except for those sequences positively identified by their SEQ ID NO. This non-compliance with the substantive provisions is to such an extent that a meaningful search over the whole subject-matter of the claims could not be carried out. The search and examination of said claims was consequently restricted to those hybridizing portions directly and unambiguously derivable from the application, namely hybridizing portions SEQ ID NO:1-749 for the first and SEQ ID NO:750-1498 for the second population of oligonucleotide primers.

The applicant's attention is drawn to the fact that claims relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure. If the application proceeds into the regional phase before the EPO, the applicant is reminded that a search may be carried out during examination before the EPO (see EPO Guideline C-VI, 8.2), should the problems which led to the Article 17(2)PCT declaration be overcome.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2008/081206

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 1-10, 13-24 and 42-46 (in part)
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers allsearchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search reportcovers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

see annex

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-5, 7-11, 13-16 (in part) and 17-23, 31, 33, 35 (in full)

Population of oligonucleotides comprising a hybridizing portion selected from SEQ ID NO:750-1498;
Reagent comprising a population of oligonucleotides to prime the amplification of a target, wherein each oligonucleotide comprises a hybridizing portion selected from said SEQ ID NO:750-1498 and a defined sequence portion in 5';
Kits thereof;
Methods thereof for selectively amplifying a target population, further comprising a first population of oligonucleotide primers comprising a hybridizing portion;
Methods thereof for transcriptome profiling.

---

2. claims: 6, 12, 36-37 (in part) and 6, 12, 30, 32, 34 (in full)

Population of oligonucleotides comprising a hybridizing portion selected from SEQ ID NO:1-749;
Reagent comprising a population of oligonucleotides to prime the amplification of a target, wherein each oligonucleotide comprises a hybridizing portion selected from said SEQ ID NO:1-749, and a defined sequence portion in 5';
Methods thereof for selectively amplifying a target population, further comprising a second population of oligonucleotide primers whose hybridizing portion is selected from all possible oligonucleotides having a length of 6, 7 or 8 nucleotides that hybridize to the target and do not hybridize to the non-target population.

---

3. claims: 1-16, 36-37 (in part) and 25-29, 38-47 (in full)

Kits for selectively amplifying a target population, comprising (a) a first population of oligonucleotide primers whose hybridizing portion is selected from all possible oligonucleotides having a length of 6 nucleotides that hybridize to the target and do not hybridize to the non-target population; (b) a second population of oligonucleotide primers comprising a hybridizing portion selected from the reverse complements of the hybridizing portions of the first population and a defined sequence portion in 5'; and (c) first and second PCR primers;
First population of oligonucleotides comprising a hybridizing portion selected from SEQ ID NO:1-749, and second population of oligonucleotides comprising a hybridizing portion selected from its reverse complement SEQ ID NO:750-1498;
Methods thereof for selectively amplifying a target population.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

4. claims: 24, 48-52 (in full)

Population of nucleic acid molecules comprising 5' and 3' defined sequences around an amplified nucleic acid sequence expressed in a mammalian cell; wherein less than 10% of said amplified nucleic acid sequences are ribosomal RNA.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2008/081206

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2007050990 A	03-05-2007	CA 2626977 A1	03-05-2007
		EP 1941058 A2	09-07-2008
		US 2008187969 A1	07-08-2008
EP 0787209 B	20-03-2002	AT 214741 T	15-04-2002
		CA 2200627 A1	06-02-1997
		DE 69619959 D1	25-04-2002
		DE 69619959 T2	07-11-2002
		EP 0787209 A1	06-08-1997
		FR 2737223 A1	31-01-1997
		WO 9704126 A1	06-02-1997
		US 5824517 A	20-10-1998