

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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
27 February 2003 (27.02.2003)

PCT

(10) International Publication Number
WO 03/016483 A2

- (51) International Patent Classification⁷: **C12N**
- (21) International Application Number: PCT/US02/26063
- (22) International Filing Date: 16 August 2002 (16.08.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/312,893 16 August 2001 (16.08.2001) US
- (63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/312,893 (CIP)
Filed on 16 August 2001 (16.08.2001)
- (71) Applicant (for all designated States except US): **CURAGEN CORPORATION** [US/US]; 555 Long Warf Drive, 11th floor, New Haven, CT 06511 (US).
- (72) Inventor; and
- (75) Inventor/Applicant (for US only): **HJALT, Tord** [SE/US]; 514 Main Street, Apartment 30, East Haven, CT 06512 (US).
- (74) Agent: **ELRIFI, Ivor, R.**; Mintz, Levin, Cohn, Gerris, Glovsky and Popeo PC, One Financial Center, Boston, MA 02111 (US).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:**
— without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: METHOD OF LABELLING cRNAs FOR PROBING OLIGO-BASED MICROARRAYS

(57) Abstract: A novel method of labelling cRNAs for probing oligo-based microarrays is disclosed. The polynucleotide template of interest is reverse transcribed into double-stranded cDNA using random primers that include an RNA polymerase promoter. This cDNA is then used as a template for synthesis of labeled RNA via *in vitro* transcription in the presence of labeled precursors.



WO 03/016483 A2

METHOD OF LABELING cRNAs FOR PROBING OLIGO-BASED MICROARRAYS

FIELD OF THE INVENTION

The invention relates in part to novel methods for producing cDNA and cRNA from a biological sample. Also included are methods for comparing the presence or amount of mRNA in a biological sample. The invention also includes random oligonucleotide primers which comprise an RNA polymerase promoter.

BACKGROUND OF THE INVENTION

Gene expression is important for understanding a wide range of biological phenomena, including development, differentiation, senescence, oncogenesis, and many other medically important processes. Recently, changes in gene expression have also been used to assess the activity of new drug candidates and to identify new targets for drug development. The latter objective is accomplished by correlating the expression of a gene or genes known to be affected by a particular drug with the expression profile of other genes of unknown function when exposed to that same drug. Genes of unknown function that exhibit the same pattern of regulation, or signature, in response to the drug are likely to represent novel targets for pharmaceutical development.

Generally, the level of expression of the protein product of a gene and its messenger RNA (mRNA) transcript are correlated, so that measuring one provides reliable information about the other. Since in most instances it is technically easier to measure levels of RNA than protein, variations in mRNA levels are commonly employed to assess gene expression in different cells and tissues or in the same cells and tissues at different stages of disease or development or exposed to different stimuli. One particularly useful method of assaying gene

expression at the level of transcription employs DNA microarrays (Ramsay, *Nature Biotechnol.* 16: 40-44, 1998; Marshall and Hodgson, *Nature Biotechnol.* 16: 27-31, 1998; Lashkari et al., *Proc. Natl. Acad. Sci. (USA)* 94: 130-157, 1997; DeRisi et al., *Science* 278: 680-6, 1997).

Mammalian cells contain as many as 1×10^5 to 3×10^5 different mRNA molecules, each of which varies in abundance (or frequency) within a given cell. The most abundant mRNAs are typically present at many thousands of copies per cell, while others may be present in as few as one copy or less per cell. Techniques for analyzing gene expression at the level of transcription typically require tens to hundreds of micrograms of mRNA, or as much mRNA as might be found in 10^7 - 10^9 mammalian cells. Oftentimes, it is impractical to obtain this number of cells from a tissue of interest. For example, a blood sample typically contains 10^6 nucleated cells/ml. Hence, to obtain 10^9 cells for analysis would necessitate taking a 1000 ml blood sample, which is clearly impractical in most instances.

Various methods have been described in the literature for amplifying the amount of a nucleic acid, such as DNA and RNA, present in a sample. Among these, the most widely practiced is the polymerase chain reaction (PCR), described in U.S. Pat. No. 4,683,195 and U.S. Pat. No. 4,683,202, herein incorporated by reference. Briefly, PCR consists of amplifying denatured, complementary strands of target nucleic acid by annealing each strand to a short oligonucleotide primer, wherein the primers are chosen so as to flank the sequence of interest. The primers are then extended by a polymerase enzyme to yield extension products that are themselves complementary to the primers and hence serve as templates for synthesis of additional copies of the target sequence. Each successive cycle of denaturation, primer annealing, and primer extension essentially doubles the amount of target synthesized in the previous cycle, resulting in exponential accumulation of the target.

When PCR is used to amplify mRNA into double-stranded (ds) DNA, the polyadenylated (poly(A)+) fraction is first selected, then a complementary DNA (cDNA) copy of the mRNA is made using reverse transcriptase and an oligo(dT) primer. The products of this reaction can then be amplified directly using random priming (Ausubel et al., eds., 1994, *Current Protocols in Molecular Biology*, vol. 2, Current Protocols Publishing, New York).

cDNA is a deoxyribonucleic acid that contains the information coding for the synthesis of proteins, but lacks the intervening introns present in genomic DNA. The synthesis and/or use of cDNA and/or libraries thereof plays a critical role in a variety of

different applications in biotechnology and related fields. Applications in which cDNAs and/or libraries thereof are employed include gene discovery, differential gene expression analysis, and the like. A variety of protocols have been developed to prepare cDNA and libraries thereof, where such methods are continually being modified.

PCR methodologies in general suffer from several limitations that are well-known in the art. See U.S. Pat. No. 5,716,785 and Innis et al., eds. (1990, PCR Protocols: A Guide to Methods and Applications, Academic Press Inc., San Diego, Calif.) for review and discussion of limitations. One such limitation results from the poor fidelity of commonly used, thermostable polymerase enzymes, such as *Taq*. This results in nucleotide base misincorporations that are propagated from one cycle to the next. It is estimated that such misincorporations may occur as often as once per one thousand bases of incorporation. A second limitation is that different cDNAs are amplified with different efficiencies, resulting in under representation of some cDNA sequences and overrepresentation of others in the amplified product. Even a small difference in efficiency may result in a several-thousand fold differential in the representation of these cDNAs in the product after only 30 cycles of amplification.

An alternative method of mRNA synthesis and amplification is known as reverse transcription and is described in U.S. Pat. No. 5,716,785, herein incorporated by reference. Unlike PCR, reverse transcription does not result in geometric amplification, but rather in linear amplification. In reverse transcription, an oligo(dT) primer that is extended at the 3'-end with a bacteriophage T7 RNA polymerase promoter is used to prime the poly-A⁺ mRNA population for cDNA synthesis. After synthesis of the first-strand cDNA, the second-strand cDNA is made using the method of Gubler and Hoffman (Gene 25:263-69, 1983). Addition of RNA polymerase results in reverse transcription and linear amplification of mRNA that is anti-sense to the poly-A⁺ RNA. This method has several limitations. It is not as sensitive as PCR and requires a much larger sample of mRNA to generate the same amount of material and thus displays low efficiency.

Moreover, conventional reverse transcription cannot produce full length cDNAs from mRNAs because the conventional reverse transcription method cannot complete reverse transcription to the end cap sites of mRNAs.

That is, current techniques for reverse transcription have a technical limitation that the reaction ends prematurely because of a stable secondary structure of mRNA and thus the probability of complete transcription over the whole transcription unit including its 5' end is

extremely low. In many instances, the 5' information is of great interest. This technical limitation affects the quality of libraries. That is, most cloned cDNAs synthesized from the poly A at the 3' end using oligo (dT) as a primer have only the 3' end and do not cover the full mRNA length because of premature termination of synthesis. Several attempts have been made to overcome this problem. While some of these techniques are effective for increasing efficiency of the synthesis of the first chain to some extent, they are not yet sufficient to efficiently obtain full length cDNAs. In particular, they show particularly low efficiency for the reverse transcription of long mRNAs of several kbp or more.

Parallel quantification of large numbers of mRNA transcripts using microarray technology promises to provide detailed insight into cellular processes involved in the regulation of gene expression. This allows new understanding of signaling networks that operate in the cell and of the molecular basis and classification of disease. The ability of DNA microarrays to monitor gene expression simultaneously in a large-scale format is helping to usher in a post-genomic age, where simple constructs about the role of nature versus nurture are being replaced by a functional understanding of gene expression in living organisms. At present, high-density DNA microarrays allow researchers to quickly and accurately quantify gene-expression changes in a massively parallel manner. Microarray technology is a rapidly advancing area, which is gaining popularity in many biological disciplines from drug target identification to predictive toxicology. Over the past few years, there has been a dramatic increase in the number of methods and techniques available for carrying out this form of gene expression analysis. The techniques and associated peripherals, such as slide types, deposition methods, robotics, and scanning equipment, are undergoing constant improvement, helping to drive the technology forward in terms of robustness and ease of use.

Microarrays can be divided into two groups based on what is immobilized: A) cDNAs or B) oligodeoxyribonucleotides (oligos). For oligo-based microarrays, one can use sense oligos *i.e.* oligos having the same polarity as the mRNA/cDNA one wishes to study). This makes oligo design and matching oligos to cDNA records easy. The probe mixture then has to be made with anti-sense polarity, so that it can hybridize to the sense oligos. The anti-sense probe mixture is generally made as follows: A total RNA (or mRNA) preparation from a given tissue is subjected to reverse transcription using an oligo-dT primer with a dangling T7 promoter in the 5' end, and reverse transcriptase. Second strand cDNA synthesis is performed in the classical way using DNA polymerase I and DNA ligase. Next, the double-stranded

cDNA mixture is transcribed *in vitro* using T7 RNA polymerase, including labeling nucleotides, *i.e.* biotinylated nucleotides, producing a biotin-labeled, anti-sense, cRNA probe mixture. The cRNA probe mix is hybridized to the oligo microarray, containing sense orientation oligos representing a collection of cDNA sequences. In the example of a biotin label, a three-step sandwiched detection procedure can be applied, such as one involving streptavidin that binds biotin, a primary antibody to streptavidin, and a fluorophore-labeled secondary antibody, *i.e.* Cy3, to the primary antibody, with high-stringency washes between each step. The fluorescence of specific Cy3 signals are detected by a confocal scanner and digitized. These primary data are then normalized using background controls and internal standards. The same general procedure may also be used with cRNAs directly labeled with Cy3 or other commonly used fluorophore labels. In this case, longer oligos (50-60-mers) are used. With the sandwich method described above, sensitivity is greater, and one can use 30-mers.

The current methods of synthesizing cRNA probes hinges on the use of an oligo-dT-T7 primer for reverse transcription. This primer is intended to hybridize to the poly-A tail of most mRNA. However, the problem with this approach is two-fold: 1) some mRNAs don't have a poly-A tail; 2) some mRNA have a long 3' UTR and thus the reverse transcriptase may have difficulties copying the entire mRNA, and may produce truncated cDNAs that mostly represent non-coding sequences.

Thus, there exists a need in the art for improved methods of synthesizing and amplifying nucleic acids, especially mRNA, which methods can achieve a high degree of synthesis and amplification from a limited amount of mRNA and which simultaneously avoid the infidelity with respect to sequence and representation often introduced by other synthesis and amplification methods.

SUMMARY OF THE INVENTION

In one aspect, the present invention provides a method of production of double-stranded molecules from a single-stranded ribonucleic acid molecule, which involves three steps. The first step involves contacting the ribonucleic acid molecule with one or more first random oligonucleotide primers containing an RNA polymerase promoter and reverse transcriptase under conditions sufficient for template-driven enzymatic deoxyribonucleic acid synthesis to occur, whereby a first deoxyribonucleic acid molecule is produced. The second

step involves contacting this first deoxyribonucleic acid molecule with a second random primer which does not comprise an RNA polymerase promoter, and DNA polymerase under conditions sufficient for template driven enzymatic deoxyribonucleic acid synthesis to occur, whereby a second deoxyribonucleic acid molecule is produced. The third step involves isolating and purifying the resulting double-stranded deoxyribonucleic acid molecule, such that a double-stranded deoxyribonucleic acid molecule is produced.

In one embodiment, the single-stranded ribonucleic acid molecule used to produce double-stranded DNA molecules is an mRNA. In another embodiment, the molecule produced from the RNA of interest is a cDNA. In various embodiments, the RNA polymerase used is T3, T7 or SP6 polymerases. In other embodiments, the DNA polymerase used is E. coli DNA polymerase I, the Klenow fragment of E. coli DNA polymerase I, or T4 DNA polymerase. In another embodiment, the double-stranded DNA produced by reverse transcription of the mRNA of interest is introduced into a vector. In one embodiment, the primers are each 4-50 nucleotides in length.

Another aspect of the present invention provides a method for synthesizing at least one cRNA from a sample containing a plurality of different mRNAs. First strand cDNA is synthesized by contacting at least one mRNA in the sample with one or more random oligonucleotide primers including an RNA polymerase promoter that is sufficiently complementary to a sequence in the mRNA so as to prime first strand cDNA synthesis, and reverse transcriptase under conditions sufficient for reverse transcriptase activity to occur. Next, double-stranded cDNA is synthesized by contacting the first strand cDNA with a second random primer that does not contain an RNA polymerase promoter but is sufficiently complementary to a sequence in the first strand cDNA so to prime second strand cDNA synthesis, and with a DNA polymerase under conditions sufficient for RNA polymerase activity to occur. The resulting double-stranded cDNA is isolated and purified, and then subjected to *in vitro* transcription under conditions sufficient for template driven RNA transcription to occur, such that cRNA is produced.

In one embodiment, the single-stranded ribonucleic acid molecule used to produce double-stranded DNA molecules is an mRNA. In another embodiment, the molecule produced from the RNA of interest is a cDNA. In various embodiments, the RNA polymerase used is T3, T7 or SP6 polymerases. In other embodiments, the DNA polymerase used is E. coli DNA polymerase I, the Klenow fragment of E. coli DNA polymerase I, or T4 DNA polymerase. In another embodiment, the double-stranded DNA produced by reverse

transcription of the mRNA of interest is introduced into a vector. In one embodiment, the primers are each 4-50 nucleotides in length. In various embodiments, the samples contain total RNA or mRNA from cells. In separate embodiments, the cells are prokaryotic, yeast, eukaryotic, mammalian and/or human cells. In one embodiment, the sample contains total RNA from 1×10^6 cells or less. In another embodiment, the sample contains at least 10,000 different mRNAs.

Another aspect of the present invention discloses a method for comparing the presence or amount of at least one mRNA of interest in a first sample and in a second sample, where these first and second samples each contain a plurality of different mRNAs from one or more cells. First strand cDNA is synthesized by contacting at least one mRNA in the sample with one or more random oligonucleotide primers including an RNA polymerase promoter that is sufficiently complementary to a sequence in the mRNA so as to prime first strand cDNA synthesis, and reverse transcriptase under conditions sufficient for reverse transcriptase activity to occur. Next, double-stranded cDNA is synthesized by contacting the first strand cDNA with a second random primer that does not contain an RNA polymerase promoter but is sufficiently complementary to a sequence in the first strand cDNA so to prime second strand cDNA synthesis, and with a DNA polymerase under conditions sufficient for RNA polymerase activity to occur. The resulting double-stranded cDNA is isolated and purified, and then subjected to *in vitro* transcription under conditions sufficient for template driven RNA transcription to occur, such that cRNA is produced. The cRNA from the first sample is labeled with a first label; the cRNA produced using the second sample is labeled with a second label distinguishable from the first label. The mRNA of interest in the first sample is detected or measured by contacting the first cRNA labeled with the first label with a polynucleotide probe capable of hybridizing to the first cRNA of the mRNA of interest under conditions conducive to hybridization; and detecting any hybridization that occurs between said probe and said first cRNA. The mRNA of interest in the second sample is detected or measured by contacting the second cRNA, labeled with second label, with polynucleotide probe capable of hybridizing to second cRNA of the mRNA of interest under conditions conducive to hybridization; and detecting any hybridization that occurs between probe and second cRNA. The levels of the mRNA of interest detected or measured in the first sample are compared to levels of the mRNA of interest detected or measured in the second sample.

In one embodiment, the single-stranded ribonucleic acid molecule used to produce double-stranded DNA molecules is an mRNA. In another embodiment, the molecule produced from the RNA of interest is a cDNA. In various embodiments, the RNA polymerase used is T3, T7 or SP6 polymerases. In other embodiments, the DNA polymerase used is E. coli DNA polymerase I, the Klenow fragment of E. coli DNA polymerase I, or T4 DNA polymerase. In another embodiment, the double-stranded DNA produced by reverse transcription of the mRNA of interest is introduced into a vector. In one embodiment, the primers are each 4-50 nucleotides in length. In various embodiments, the samples contain total RNA or mRNA from cells. In separate embodiments, the cells are prokaryotic, yeast, eukaryotic, mammalian and/or human cells. In one embodiment, the sample contains total RNA from 1×10^6 cells or less. In another embodiment, the sample contains at least 10,000 different mRNAs.

In another aspect, the present invention provides a method for producing a 5' enriched cDNA library from a sample of mRNA molecules. The mRNA is contacted with one or more first random oligonucleotide primers including an RNA polymerase promoter, and reverse transcriptase under conditions sufficient for template driven enzymatic deoxyribonucleic acid synthesis to occur, whereby a first population of cDNA molecules is produced. This first population of cDNA molecules is contacted with a second random primer not comprising an RNA polymerase promoter, and a DNA polymerase under conditions sufficient for template driven enzymatic deoxyribonucleic acid synthesis to occur, whereby a second population of cDNA molecules is produced. This second population of double-stranded cDNA molecules is isolated in order to produce a 5' enriched library. In one embodiment, the mRNA contains a poly-adenylated tail. In another embodiment, the mRNA is contacted with a primer mixture that includes one or more first random oligonucleotide primers containing an RNA polymerase promoter and oligo (dT) primers containing an RNA polymerase promoter. In a further embodiment, the oligo (dT) primer is stably associated with the surface of a solid support. In yet another embodiment, the first and second labels are fluorescent, radioactive, enzymatic, hapten, biotin or digoxigenin labels. In a further embodiment, the fluorescent label is fluorescein isothiocyanate, lissamine, Cy3, Cy5, or rhodamine 110. In another embodiment, a first aliquot of the first cRNA is labeled with a first fluorophore having a first emission spectrum, and a second aliquot of the second cRNA is labeled with a second fluorophore with a second emission spectrum differing from that of the first emission

spectrum. In another embodiment, the first fluorophore is Cy3 and the second fluorophore is Cy5.

Still another aspect of the present invention discloses a kit including, in one or more containers a mixture of first random oligonucleotide primers, where each first primer contains a 3' end sequence of 10-50 nucleotides that contains an A, a G, a T, or a C nucleotide randomly present in each position, and a 5' end sequence of 10-50 nucleotides that includes an RNA polymerase promoter, a mixture of oligo (dT) primers, where each oligo (dT) primer contains a 3' end sequence of 10-50 nucleotides that includes a T nucleotide present in each position; and also includes a 5' end sequence of 10-50 nucleotides containing an RNA polymerase promoter and a mixture of second random oligonucleotide primers, each second primer containing a sequence of 10-50 nucleotides that comprises an A, a G, a T, or a C nucleotide randomly present in each position.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides a method capable of reverse transcription of mRNA over its full length, particularly the 5' region of the cDNA, and hence capable of providing a full length cDNA even if a long chain mRNA is used as a template. The present invention also provides a method for *in vitro* transcription of double-stranded cDNA molecule to produce a labeled cRNA molecule for use in various methodologies including microarray technology.

The present invention relates to random oligonucleotide primers that contain an RNA polymerase promoter for synthesis of nucleic acids. The invention further relates to methods of use of the oligonucleotide primers of the invention for the synthesis of nucleic acid sequences in a population of cells, which methods preserve fidelity with respect to sequence and transcript representation.

More specifically, the present invention incorporates a change in the methods used for priming of the mRNA. Instead of using only oligo-dT primers in conjunction with T7 promoters, random oligonucleotide primers of varying lengths (*i.e.* hexamer or nanomer) are employed that have RNA polymerase promoters in their 5' ends. Using these random primers containing the RNA polymerase promoters, alone or in combination with an oligo-dT primer, will increase probe coverage towards the 5' end of the template mRNA, as compared to methods using only an oligo-dT primer.

Use of the random primers of the invention, comprising an RNA polymerase promoter, and subsequent second strand cDNA synthesis generates a library of cDNAs of different lengths and different start and stop positions. This results in the generation of a tiled network, completely covering the mRNA population; in particular the cDNAs generated provide coverage across the full length of most of the individual mRNA sequences. The ensuing *in vitro* transcription of these cDNAs generates a complete, full-coverage set of either directly labeled or biotin-tagged cRNA probes. The present invention does not necessitate any changes in existing chip oligo design, hybridization, or detection protocols. Thus it is a stream-lined improvement to existing procedures.

For clarity of disclosure, and not by way of limitation, the detailed description of the invention is divided into the subsections set forth below.

OLIGONUCLEOTIDES

The oligonucleotide primers (*i.e.* random primers or oligo (dT) primers) for use in the methods of the invention can be of any suitable size, and are preferably 4-50 nucleotides in length. The terms "oligonucleotide" and "oligo" as used herein denotes single stranded nucleotide multimers of from about 10 to about 100 nucleotides in length. The term "polynucleotide" as used herein refers to a single- or double-stranded polymer composed of nucleotide monomers of greater than about 100 nucleotides in length up to about 1000 nucleotides in length.

An oligo (dT) primer for use in the subject methods is sufficiently long to provide for efficient hybridization to the polyA tail of mRNA. Generally, the length of the oligo (dT) primer ranges from about 10 to 50 nt in length, preferably from about 10 to 30 nt in length, and more preferably from about 20 to 25 nt length. The oligo (dT) primer contains a T nucleotide present in each position of the oligo (dT) primer sequence. In many embodiments, the oligo (dT) primer is stably attached to the surface of a solid support. The solid support may be any convenient solid support known to those skilled in the art. A variety of different solid-phases are suitable for use in the subject methods, such phases being known in the art and commercially available. Specific solid-phases of interest include polymeric supports including, but not limited to, polystyrene, pegs, sheets, beads, magnetic beads, and the like. By "stably attached" is meant that the oligo (dT) primer remains associated with the surface of the solid support under at least conditions of enzymatic template driven nucleic acid

synthesis. The oligo (dT) primer may be covalently or non-covalently bonded to the solid phase. The oligo (dT) primer is contacted with the mRNA under conditions sufficient for template driven enzymatic deoxyribonucleic acid synthesis to occur, where the mRNA molecule serves as the template molecule.

A random oligonucleotide primer for use in the subject methods is sufficiently long to provide for efficient hybridization to a complementary nucleic acid. By "random oligonucleotide primer" or "random primer" is meant a primer of random base sequence relative to the single-stranded RNA of the first hybrid, *i.e.* a primer that is not known with certainty to hybridize to the mRNA component of the first hybrid or duplex prior to use. In other words, the random primer has a sequence that is not known for certainty to have a complementary region in the single-stranded portion of the RNA of the hybrid prior to contact with the random primer. The random primer is sufficiently long to provide for efficient hybridization to a complementary region of the mRNA component, where the random primer typically ranges from about 4 to 50 nt, preferably from about 10 to 30 nt and more preferably from about 20 to 30 nt in length. Primers contain random arrangements of A, T, C, G, U or I nucleotides, such that each random primer contains an A, a G, a T, a C or an I nucleotide randomly present in each position of the sequence. In preferred embodiments, the nucleotides are deoxyribonucleotide triphosphates (dNTPs) *i.e.* dATP, dTTP, dCTP, dGTP, or dITP or ribonucleotide triphosphates (rNTPs) *i.e.*, rATP, rUTP, rCTP, rGTP, or rITP. Specific primers of interest include those that have from about 5 to 15, preferably from about 6 to 9 nt random nts fused in one embodiment to a restriction enzyme cut site or more preferably fused to an RNA polymerase promoter from about 5 to 20 nt, preferably from about 8 to 16 nt, *e.g.* 3'(N)₉GCGGCCGCGCGGCCGC5', where N is any nucleotide. Hexamer or nanomer oligonucleotides are preferred embodiments for random primers. In another preferred embodiment, the primers of the present invention are present in a kit.

The oligonucleotide primers can be DNA, RNA, chimeric mixtures or derivatives or modified nucleic acid versions thereof, so 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. The term "nucleic acid" as used herein means a polymer composed of nucleotides, *e.g.* deoxyribonucleotides or ribonucleotides. The terms "ribonucleic acid" and "RNA" as used herein mean a polymer composed of ribonucleotides. RNA can be single-stranded or in a duplex. In a preferred embodiment,

single-stranded RNA is message RNA (mRNA). The terms "deoxyribonucleic acid" and "DNA" as used herein mean a polymer composed of deoxyribonucleotides. Deoxyribonucleic acid can be single-stranded or double-stranded. In a preferred embodiment, DNA is complementary DNA (cDNA).

In a preferred embodiment, the random oligonucleotide primer contains an RNA polymerase promoter. Preferably, the RNA polymerase promoter can be at the 5' end of the random oligonucleotide primer. The RNA polymerase promoter can be any RNA polymerase promoter sequence known in the art. In preferred embodiments, the RNA polymerase promoter is a T7 RNA polymerase promoter sequence, a T3 RNA polymerase promoter sequence, or a SP6 RNA polymerase promoter sequence; T7 RNA polymerase, T3 RNA polymerase and SP6 RNA polymerase can be used to initiate RNA synthesis from these promoters, respectively.

In one embodiment, a random oligonucleotide primer contains a 3' end sequence of 10-50 nucleotides, said 3' end sequence having an A, a G, a T, a C, or a I nucleotide randomly present in each position of said 3' end sequence; and contains a 5' end sequence of 10-50 nucleotides, said 5' end sequence having an RNA polymerase promoter (*i.e.* any RNA polymerase promoter sequence known in the art). More preferably, the RNA polymerase promoter is a T7 RNA polymerase promoter sequence, a T3 RNA polymerase promoter sequence, or a SP6 RNA polymerase promoter sequence.

In another embodiment, an oligo (dT) primer contains a 3' end sequence of 10-50 nucleotides, said 3' end sequence having a T nucleotide present in each position of said 3' end sequence; and contains a 5' end sequence of 10-50 nucleotides, said 5' end sequence having an RNA polymerase promoter (*i.e.* any RNA polymerase promoter sequence known in the art). More preferably, the RNA polymerase promoter is a T7 RNA polymerase promoter sequence, a T3 RNA polymerase promoter sequence, or a SP6 RNA polymerase promoter sequence.

In another embodiment, a random oligonucleotide primer, contains a sequence of 10-50 nucleotides, said sequence having an A, a G, a T, a C, or a I nucleotide randomly present in each position of said sequence.

In another embodiment, the oligonucleotide primer may contain at least one modified base moiety, including, but not limited to, 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) 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 (v), wybutosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine.

In another embodiment, the oligonucleotide primer contains at least one modified sugar moiety selected from the group including, but not limited to, arabinose, 2-fluoroarabinose, xylulose, and hexose.

In yet another embodiment, the oligonucleotide primer contains 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.

The oligonucleotide primers 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 synthesized by standard methods known in the art, *e.g.* by use of an automated DNA synthesizer (such as are commercially available from Biosearch, Applied Biosystems) and standard phosphoramidite chemistry. Phosphorothioate oligonucleotides may be synthesized by the method of Stein et al. (1988, Nucl. Acids Res. 16:3209-3221), methylphosphonate oligonucleotides can be prepared by use of controlled pore glass polymer supports (Sarin et al., 1988, Proc. Natl. Acad. Sci. U.S.A. 85:7448-7451), etc.

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 oligonucleotide that has been separated on an acrylamide gel, or by measuring the optical density at 260 nm in a spectrophotometer.

METHODS OF LABELING OF NUCLEIC ACID PRODUCTS

Nucleic acid synthesis or amplification products such as synthesized RNA (*i.e.* cRNA) may be labeled with any art-known detectable marker, including radioactive labels such as ³²P, ³⁵S, ³H, and the like; fluorophores; chemiluminescers; or enzymatic markers. In a preferred embodiment, the label is fluorescent.

Exemplary suitable fluorophore moieties that can be used as labels are selected from the group including but not limited to: 4-acetamido-4'-isothiocyanatostilbene-2,2'-disulfonic acid acridine and derivatives: acridine, acridine isothiocyanate, 5-(2'-aminoethyl)aminonaphthalene-1-sulfonic acid (EDANS), 4-amino-N-[3-vinylsulfonyl]phenyl]naphthalimide-3,5 disulfonate (Lucifer Yellow VS), -(4-anilino-1-naphthyl)maleimide, anthranilamide, Brilliant Yellow; coumarin and derivatives: coumarin, 7-amino-4-methylcoumarin (AMC, Coumarin 120), 7-amino-4-trifluoromethylcoumarin (Coumarin 151), Cy3, Cy5, cyanosine, 4',6-diaminidino-2-phenylindole (DAPI), 5',5"-dibromopyrogallol-sulfonephthalein (Bromopyrogallol Red), 7-diethylamino-3-(4'-isothiocyanatophenyl)-4-methylcoumarin, diethylenetriamine pentaacetate, 4,4'-diisothiocyanatodihydro-stilbene-2,2'-disulfonic acid, 4,4'-diisothiocyanatostilbene-2,2'-disulfonic acid, 5-[dimethylamino]naphthalene-1-sulfonyl chloride (DNS, dansyl chloride), 4-(4'-dimethylaminophenylazo)benzoic acid (DABCYL), 4-dimethylaminophenylazophenyl-4'-isothiocyanate (DABITC); eosin and derivatives: eosin, eosin isothiocyanate; erythrosin and derivatives: erythrosin B, erythrosin isothiocyanate, ethidium; fluorescein and derivatives: 5-carboxyfluorescein (FAM), 5-(4,6-dichlorotriazin-2-yl)aminofluorescein (DTAF), 2',7'-dimethoxy-4',5'-dichloro-6-carboxyfluorescein (JOE), fluorescein, fluorescein isothiocyanate, QFITC (XRITC), fluorescamine, IR144, IR1446, Malachite Green isothiocyanate, 4-methylumbelliferone, ortho cresolphthalein, nitrotyrosine, pararosanine, Phenol Red, B-phycoerythrin, o-phthaldialdehyde; pyrene and derivatives: pyrene, pyrene butyrate, succinimidyl 1-pyrene butyrate, Reactive Red 4 (Cibacron Brilliant Red 3B-A); rhodamine and derivatives: 6-carboxy-X-rhodamine (ROX), 6-carboxyrhodamine (R6G), lissamine rhodamine B sulfonyl chloride, rhodamine (Rhod), rhodamine B, rhodamine 110, rhodamine 123, rhodamine X isothiocyanate, sulforhodamine B, sulforhodamine 101, sulfonyl chloride derivative of sulforhodamine 101 (Texas Red), N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA), tetramethyl rhodamine, tetramethyl rhodamine isothiocyanate (TRITC), riboflavin, rosolic acid, terbium chelate derivatives.

In a preferred embodiment, synthesized cRNA is labeled with a label, *i.e.*, any label known in the art and as described herein. In other preferred embodiments, the synthesized cRNA is labeled with a label wherein the label is a fluorescent, radioactive, enzymatic, hapten, biotin, or digoxigenin label, more preferably the label is fluorescent. The preferred method of labeling in this invention is with a fluorophore, such as fluorescein isothiocyanate, lissamine, Cy3, Cy5, and rhodamine 110, with Cy3 and Cy5 particularly preferred.

In addition to fluorophores, chemiluminescers and enzymes, among others, may also be used as labels. In another embodiment, the oligonucleotide may be labeled with an enzymatic marker that produces a detectable signal when a particular chemical reaction is conducted, such as alkaline phosphatase or horseradish peroxidase. Such enzymatic markers are preferably heat stable, so as to survive the denaturing steps of the amplification process.

Oligonucleotides may also be indirectly labeled by incorporating a nucleotide linked covalently to a hapten or to a molecule such as biotin, to which a labeled avidin molecule or streptavidin may be bound, or digoxigenin, to which a labeled anti-digoxigenin antibody may be bound.

Oligonucleotides of the invention may be labeled with labeling moieties during chemical synthesis or the label may be attached after synthesis by methods known in the art.

LABELING OF RNA

In one embodiment, the synthesized RNA of the invention (*i.e.* cRNA) is labeled during synthesis to facilitate its detection in subsequent steps. The RNA may be labeled with any art-known detectable marker, including but not limited to radioactive labels such as ^{32}P , ^{35}S , ^3H , and the like; fluorophores; chemiluminescers; or enzymatic markers.

Labeling of RNA is preferably accomplished by including one or more labeled NTPs in the *in vitro* transcription reaction mixture (*See*, Example 1). NTPs may be directly labeled with a radioisotope, such as ^{32}P , ^{35}S , ^3H . Radiolabeled NTPs are available from several sources, including New England Nuclear (Boston, Mass.) and Amersham. NTPs may be directly labeled with a fluorescent label such as Cy3 or Cy5. In one embodiment, biotinylated or allylamine-derivatized NTPs are incorporated during the *in vitro* transcription reaction and the resultant cRNAs thereafter labeled indirectly, for example, by the addition of fluorophore-conjugated avidin (in the case of biotin) or the NHS ester of a fluorophore (in the case of

allylamine). In another embodiment, fluorescently labeled NTPs may be incorporated during the *in vitro* transcription reaction, which fluorescently labels the resultant cRNAs directly.

Exemplary fluorophore moieties that can be used as labels are set forth herein. The preferred label in the methods of this invention is a fluorophore, such as fluorescein isothiocyanate, lissamine, Cy3, Cy5, and rhodamine 110, with Cy3 and Cy5 particularly preferred.

Not only fluorophores, but also chemiluminescers and enzymes, among others, may be used as labels. In yet another embodiment, the RNA may be labeled with an enzymatic marker that produces a detectable signal when a particular chemical reaction is conducted, such as alkaline phosphatase or horseradish peroxidase. Such enzymatic markers are preferably heat stable, so as to survive the denaturing steps of the amplification process.

RNA may also be indirectly labeled by incorporating a nucleotide linked covalently to a hapten or to a molecule such as biotin, to which a labeled avidin molecule may be bound, or digoxigenin, to which a labeled anti-digoxigenin antibody may be bound. RNA may be labeled with labeling moieties during chemical synthesis or the label may be attached after synthesis by methods known in the art.

Labeling of RNA is preferably accomplished by preparing cRNA that is fluorescently labeled with NHS-esters. Most preferably, labeling of RNA is accomplished in a two-step procedure in which allylamine-derivatized UTP is incorporated during *in vitro* transcription. Following the *in vitro* transcription reaction, unincorporated nucleotides are removed and the allylamine-containing RNAs are conjugated to the N-hydroxysuccinimide (NHS) esters of Cy3 or Cy5.

This two-step method of preparing fluorescent-labeled cRNA is relatively inexpensive and suitable for use in two color hybridizations to DNA microarrays. In the first step, aminoallyl (AA)-labeled nucleic acids are prepared by incorporation of AA-nucleotides. AA-UTP (Sigma A-5660) may be used for labeling cRNA. AA-cRNA is prepared using the Ambion MegaScript T7 RNA polymerase *in vitro* transcription kit, with AA-UTP substituted at 50-100% of the total UTP concentration. It is essential to remove all traces of amine-containing buffers such as Tris prior to derivatizing the AA-nucleic acids. AA-nucleic acids prepared in enzymatic reactions are preferably cleaned up on appropriate QIAGEN columns: RNeasy Mini kit (for RNA) or QIAquick PCR Purification kit (for DNA) (QIAGEN Inc.--USA, Valencia, Calif.). For the QIAGEN columns, samples are applied twice. For washes, 80% EtOH is preferably substituted for the buffer provided with the QIAGEN kit. Samples

are eluted twice with 50 μ l volumes of 70° C. H₂O. Alternatively (but less preferably), samples may be cleaned up by repeated cycles of dilution and concentration on Microcon-30 filters.

In the second step, AA-nucleic acids are derivatized with NHS-esters, preferably Cy 3 or Cy 5. Preferably, 2-6 μ g of AA-labeled nucleic acid are aliquoted into a microfuge tube, adjusting the total volume to 12 μ l with H₂O. The NHS-ester is dissolved at a concentration of ~15 mM in anhydrous DMSO (~200 nmoles in 13 μ l). 27 μ l of 0.1 M sodium carbonate buffer, pH 9, are added. 12 μ l of the dye mix (containing ~60 nmoles dye-NHS ester) are then immediately added to the AA-labeled nucleic acid (~6-20 pmoles of a 1 kb molecule). The samples are then incubated in the dark at 23° C for 1 hour. The coupling reaction is stopped by adding 5 μ l of a 4M solution of hydroxylamine. Incubation is continued at 23° C for an additional 0.25 hr. Dye-coupled nucleic acid is separated from unincorporated dye on an RNeasy Mini kit or QIAquick PCR Purification kit (QIAGEN Inc.--USA, Valencia, Calif.). Samples are washed with 80% EtOH instead of buffer, as described above, and eluted twice with 50 μ l volumes of 70° C H₂O.

The spectrum of the labeled nucleic acid is preferably measured from 220 nm-700 nm. The percent recovery of nucleic acid and molar incorporation of dye is calculated from extinction coefficients and absorbance values at $1_{\max}$. Recovery of nucleic acid is typically ~80%. The mole percent of dye incorporated per nucleotide ranges from 1.5-5% of total nucleotides.

In a preferred embodiment, *in vitro* transcribed RNA is labeled with biotin (*See*, Example 1) or other labels, as described above. Most preferably the label is fluorescent and the label is incorporated by procedures commonly used in the art, as well as by the procedure described in Example 1 for a biotin label.

Often, it is desirable to compare gene expression in two different populations of cells, perhaps derived from different tissues or perhaps the same tissue exposed to different stimuli (*i.e.* comparison of a diseased tissue sample with a normal tissue sample or the comparison of a drug-treated tissue with a untreated tissue). Such comparisons are facilitated by labeling the RNAs from one population with a first fluorophore and the RNAs from the other population, with a second fluorophore, where the two fluorophores have distinct emission spectra. Again, Cy3 and Cy5 are particularly preferred fluorophores for use in comparing gene expression between two different populations of cells.

METHODS OF PREPARATION OF RNA

The methods of the invention are applicable to nucleic acid sequences derived from both eukaryotic and prokaryotic cells (*i.e.* yeast cells), although they are preferably used with eukaryotic cells, more preferably, with mammalian cells and most preferably, with human cells. Among cells that may serve as sources of DNA or RNA are nucleated blood cells, established cell lines, tumor cells, and tissue biopsy specimens, among others that will be readily apparent to those of skill in the art.

Although the synthesis methods of the invention can synthesize DNA, it is preferred to utilize the methods to synthesize RNA (*i.e.* cRNA) from a population of cells. Total cellular RNA, cytoplasmic RNA, mRNA, or poly(A)+ RNA may be used. Methods for preparing total and poly(A)+ RNA are well known and are described generally in Sambrook et al. (1989, *Molecular Cloning--A Laboratory Manual* (2nd Ed.), Vols. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.) and Ausubel et al., eds. (1994, *Current Protocols in Molecular Biology*, vol. 2, Current Protocols Publishing, New York).

RNA may be isolated from eukaryotic cells by procedures that involve lysis of the cells and denaturation of the proteins contained therein. Cells of interest include wild-type cells, drug-exposed wild-type cells, diseased cells, modified cells, and drug-exposed modified cells.

Additional steps may be employed to remove DNA. Cell lysis may be accomplished with a nonionic detergent, followed by microcentrifugation to remove the nuclei and hence the bulk of the cellular DNA. In one embodiment, RNA is extracted from cells of the various types of interest using guanidinium thiocyanate lysis followed by CsCl centrifugation to separate the RNA from DNA (Chirgwin et al., 1979, *Biochemistry* 18:5294-5299). Poly(A)+ RNA is selected by selection with oligo-dT cellulose (see Sambrook et al., 1989, *Molecular Cloning--A Laboratory Manual* (2nd Ed.), Vols. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.). Alternatively, separation of RNA from DNA can be accomplished by organic extraction, for example, with hot phenol or phenol/chloroform/isoamyl alcohol.

If desired, RNase inhibitors may be added to the lysis buffer. Likewise, for certain cell types, it may be desirable to add a protein denaturation/digestion step to the protocol.

For many applications, it is desirable to preferentially enrich mRNA with respect to other cellular RNAs, such as transfer RNA (tRNA) and ribosomal RNA (rRNA). Most

mRNAs contain a poly(A) tail at their 3' end. This allows them to be enriched by affinity chromatography, for example, using oligo(dT) or poly(U) coupled to a solid support, such as cellulose or SephadexTM (see Ausubel et al., eds., 1994, Current Protocols in Molecular Biology, vol. 2, Current Protocols Publishing, New York). Once bound, poly(A)+ mRNA is eluted from the affinity column using 2 mM EDTA/0.1% SDS.

The sample of RNA can contain a plurality of different mRNA molecules, each different mRNA molecule having a different nucleotide sequence. In a specific embodiment, the mRNA molecules in the RNA sample contain at least 100 different nucleotide sequences. More preferably, the mRNA molecules of the RNA sample contain at least 500, 1,000, 5,000, 10,000, 20,000, 30,000, 40,000, 50,000, 60,000, 70,000, 80,000 90,000 or 100,000 different nucleotide sequences. In another specific embodiment, the RNA sample is a mammalian RNA sample (*i.e.* human RNA sample), the mRNA molecules of the mammalian RNA sample containing about 20,000 to 30,000 different nucleotide sequences.

In a specific embodiment, total RNA or mRNA from cells are used in the methods of the invention. The source of the RNA can be cells of a plant or animal, human, mammal, primate, non-human animal, dog, cat, mouse, rat, bird, yeast, eukaryote, prokaryote, etc. In specific embodiments, the method of the invention is used with a sample containing total mRNA or total RNA from 1×10^6 cells or less.

METHODS OF DOUBLE-STRANDED cDNA SYNTHESIS

In the broadest sense, the subject invention is directed to methods of preparing double-stranded nucleic acid molecules from single-stranded nucleic acid molecules, and specifically to methods of preparing double-stranded deoxyribonucleic acid molecules from single-stranded ribonucleic acid molecules. Generally, the single-stranded ribonucleic acid molecule is an mRNA molecule having a polyA tail while the double-stranded deoxyribonucleic acid molecule is a cDNA molecule. However, the mRNA molecule need not contain a polyA tail. In particular, the subject invention is directed to methods of producing double-stranded cDNA molecules that include the sequence information from the 5' end of the template mRNA from which they are prepared. The length of the cDNA molecules prepared by the subject methods typically ranges from about 0.5 to 3.0 kb, usually from about 1.0 to 2.0 kb and more usually from about 1.0 to 1.5 kb.

The first step in the subject methods is to prepare a first deoxyribonucleic acid strand (*i.e.* cDNA). This first DNA strand is prepared from the initial single-stranded RNA molecule, *e.g.* the mRNA. This step is generally accomplished by contacting the initial RNA with a primer and a reverse transcriptase under conditions sufficient for template driven enzymatic DNA synthesis to occur. The terms "sufficient for" and "conducive to" as used herein refers to a quantity or condition that can fulfill a need or requirement but without being abundant, *i.e.* an circumstance, situation or environment which allows the method, procedure or protocol of the invention to occur. In a preferred embodiment, this first step involves contacting a single stranded RNA molecule, *i.e.* mRNA, with one or more first random primers containing an RNA polymerase promoter and reverse transcriptase under conditions sufficient for template driven enzymatic DNA synthesis to occur, where the mRNA molecule serves as the template molecule. In preferred embodiments, the RNA polymerase promoter is a T7 RNA polymerase promoter sequence, a T3 RNA polymerase promoter sequence or a SP6 RNA polymerase promoter sequence and T7 RNA polymerase, T3 RNA polymerase and SP6 RNA polymerase are used to initiate RNA or DNA synthesis from these promoters, respectively. In a further embodiment, where the initial single-stranded RNA molecule contains a polyadenylated tail, one or more random primers containing an RNA polymerase promoter contacts the mRNA in combination with an oligo (dT) primer containing an RNA polymerase promoter under conditions for template driven enzymatic DNA synthesis to occur. In another embodiment the oligo (dT) primer is stably associated with the surface of a solid support.

In a preferred embodiment, contact of the initial single stranded RNA, *i.e.* mRNA, with one or more first random primers containing an RNA polymerase promoter alone or in combination with an oligo (dT) primer containing an RNA polymerase promoter under conditions sufficient for template driven enzymatic DNA synthesis to occur produces a first deoxyribonucleic acid molecule, *i.e.* first strand cDNA. In another embodiment, the primers used in the method of the invention are each 4-50 nucleotides in length. Suitable oligonucleotide primers are described herein and can be modified as necessary by the skilled artisan.

The initial mRNA that serves as template in the first step may be present in a variety of different samples, where the sample will typically be derived from a physiological source. The physiological source may be derived from a variety of eukaryotic sources, with physiological sources of interest including sources derived from single celled organisms such

as yeast and multicellular organisms, including plants and animals, particularly mammals, where the physiological sources from multicellular organisms may be derived from particular organs or tissues of the multicellular organism, or from isolated cells derived therefrom. In obtaining the sample of RNAs to be analyzed from the physiological source from which it is derived, the physiological source may be subjected to a number of different processing steps, where such processing steps might include tissue homogenization, cell isolation and cytoplasmic extraction, nucleic acid extraction and the like, where such processing steps are known to those of skill in the art. Methods of isolating RNA from cells, tissues, organs or whole organisms are known to those of skill in the art and are described herein and in Maniatis et al., *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Press)(1989).

In addition to the mRNA and the primer, reverse transcriptase, RNA polymerase and other reagents necessary for primer extension are present. In preferred embodiments, the RNA polymerase is any known RNA polymerase but most preferably the RNA polymerase is T7 RNA polymerase, T3 RNA polymerase and SP6 RNA polymerase. In other embodiments, a reverse transcription or first strand DNA synthesis buffer, dithiothreitol (DTT), a mixture of deoxyribonucleotide triphosphates (dNTPs) *i.e.* dATP, dTTP, dCTP, dGTP, or dTTP, and nuclease-free H₂O are present. An illustrative procedure for first strand cDNA synthesis as provided in Example 1. The order in which the reagents are combined may be modified as desired, and the entire illustrative procedure described therein can be modified by the skilled artisan.

Second strand cDNA synthesis is carried out by any method known in the art, as illustrated in Example 1. In a preferred embodiment, the second step of the subject methods is to contact the first deoxyribonucleic acid molecule with a second random primer which does not contain an RNA polymerase promoter, and DNA polymerase under conditions sufficient for template driven enzymatic deoxyribonucleic acid synthesis to occur, whereby a second deoxyribonucleic acid molecule is produced resulting in a double-stranded deoxyribonucleic acid molecule (cDNA). In another embodiment, the double-stranded deoxyribonucleic acid molecule can be isolated and purified according to any method known in the art and illustrated in Example 1. In specific embodiments, the DNA polymerase used for second strand cDNA synthesis is any DNA polymerase known in the art and more preferably *E. coli* DNA polymerase I, Klenow fragment of *E. coli* DNA polymerase I, or T4 DNA polymerase. In another embodiment, the primers used in the method of the invention are each 4-50 nucleotides in length. Suitable oligonucleotide primers are described herein and can be

modified by the skilled artisan. In other embodiments, a second strand DNA synthesis buffer, DNA ligase, a mixture of deoxyribonucleotide triphosphates (dNTPs) *i.e.* dATP, dTTP, dCTP, dGTP, or dTTP, RNase H (to degrade any remaining RNA molecules), and nuclease-free H₂O are present (*See*, Example 1).

Contact of the first strand DNA with the second random primer under conditions sufficient for template driven enzymatic DNA synthesis (as described above), results in the production of double-stranded cDNA molecules with different lengths and different start and stop positions. Of particular interest are double-stranded cDNA molecules in which the cDNA molecule includes the genetic information of the 5' terminus of the RNA, *e.g.* mRNA. The length of the double-stranded cDNA molecule typically ranges from about 500 to 3000 nt, preferably from about 1000 to 2000 nt and more preferably from about 1500 to 2000 nt.

In another embodiment, the present invention is directed to a method for producing a 5' enriched cDNA library from a sample of mRNA molecules, said method containing: (a) contacting said mRNA with (i) one or more first random oligonucleotide primers containing an RNA polymerase promoter, and (ii) reverse transcriptase under conditions sufficient for template driven enzymatic deoxyribonucleic acid synthesis to occur, whereby a first population of cDNA molecules is produced; (b) contacting said first population of cDNA molecules (i) with a second random primer not containing an RNA polymerase promoter, and (ii) DNA polymerase under conditions sufficient for template driven enzymatic deoxyribonucleic acid synthesis to occur, whereby a second population of cDNA molecules is produced; and (c) isolating said second population of double-stranded cDNA molecules; such that said 5' enriched library is produced. By "5' enriched" is meant that a significant proportion of the cDNAs in the library contain the nucleotide sequence information of the 5' end of the mRNAs from which the cDNAs are derived. The 5' enriched cDNA libraries of the subject invention contain at least 5, usually at least 50 and more usually at least 100 distinct cDNAs (*i.e.* cDNAs that differ in sequence from each other), where the number of distinct cDNAs in the library may be as high as 10,000 or higher. A significant portion of the cDNA constituents of the library include the 5' sequence information of their corresponding mRNA from which they were derived, where the percentage of distinct cDNAs in the library that include 5' sequence information from their corresponding mRNAs is at least about 10%, usually at least about 20% and more usually at least about 25%, where the percentage may be as high as 30% or higher, including in certain embodiments, 40%, 50%, 60% 70% 80%, 90% or higher. The initial mRNA that serves as template in the first step may be present in a

variety of different samples, where the sample will typically be derived from a physiological source. In preferred embodiments, the physiological source is prokaryotic tissue or cell (*i.e.* yeast cells) and eukaryotic tissue or cell. In more preferred embodiments, the sample source is a mammalian tissue or cell and most preferably the sample source is a human tissue or cell. In another embodiment, the sample contains total RNA or total mRNA from the tissue or cells described herein. In a preferred embodiment the sample contains total RNA from 1×10^6 cells or less and the sample contains at least 10,000 different mRNAs. In preferred embodiments, the RNA polymerase promoter is a T7 RNA polymerase promoter sequence, a T3 RNA polymerase promoter sequence or a SP6 RNA polymerase promoter sequence and T7 RNA polymerase, T3 RNA polymerase and SP6 RNA polymerase are used to initiate RNA or DNA synthesis from these promoters, respectively. In specific embodiments, the DNA polymerase used for second strand cDNA synthesis is any DNA polymerase known in the art and more preferably *E. coli* DNA polymerase I, Klenow fragment of *E. coli* DNA polymerase I, or T4 DNA polymerase. In another embodiment, the primers used in the method of the invention are each 4-50 nucleotides in length. Suitable oligonucleotide primers are described herein and can be modified by the skilled artisan. In a further embodiment, where the initial single-stranded RNA molecule contains a polyadenylated tail, one or more random primers containing an RNA polymerase promoter contacts the mRNA in combination with an oligo (dT) primer containing an RNA polymerase promoter under conditions for template driven enzymatic DNA synthesis to occur. In another embodiment the oligo (dT) primer is stably associated with the surface of a solid support. In another embodiment, the second population of double-stranded cDNA molecules can be introduced into a vector as described below.

In another embodiment, the resultant isolated double-stranded cDNA may then be ligated into an appropriate vector for propagation and subsequent use, as desired, using methods well known to those of skill in the art. Appropriate vectors include viral, phagemid and plasmid vectors. Generally, this step involves contact of the vector with the double-stranded cDNA under conditions sufficient for ligation of the cDNA with the vector to occur. Typically, the vector and the cDNA will have been treated to produce complementary or "sticky-ends" with at least one, and preferably two different restriction endonucleases. This pretreatment step may further include ligation of linker or adapter sequences onto the ends of the vector and/or ds cDNA in order to introduce desired restriction sites, etc.

METHODS OF cRNA SYNTHESIS

In a preferred embodiment, the present invention provides a method for synthesizing at least one cRNA from a sample containing a plurality of different mRNAs, said method containing: (a) synthesizing first strand cDNA by contacting at least one mRNA in said sample with (i) one or more random oligonucleotide primers containing an RNA polymerase promoter that is sufficiently complementary to a sequence in the mRNA so as to prime first strand cDNA synthesis, and (ii) reverse transcriptase under conditions sufficient for reverse transcriptase activity to occur; (b) synthesizing double-stranded cDNA by contacting the first strand cDNA with (i) a second random primer not containing an RNA polymerase promoter wherein said second random primer is sufficiently complementary to a sequence in the first strand cDNA so to prime second strand cDNA synthesis, and (ii) a DNA polymerase under conditions sufficient for DNA polymerase activity to occur; (c) isolating and purifying said double-stranded cDNA; and (d) subjecting said purified double-stranded cDNA to *in vitro* transcription under conditions sufficient for template driven DNA transcription to occur, such that cRNA is produced.

In vitro transcription can be accomplished by any method well known in the art, *i.e.*, Example 1; Ausubel et al., eds., 1994, Current Protocols in Molecular Biology, vol. 2, Current Protocols Publishing, New York. Prior to subjecting double-stranded cDNA to *in vitro* transcription, it may be necessary to introduce the double-stranded cDNA molecule into a vector using any methods well known to those of skill in the art (Ausubel et al., eds., 1994, Current Protocols in Molecular Biology, vol. 2, Current Protocols Publishing, New York) and as described herein.

In preferred embodiments, the physiological source is prokaryotic tissue or cell (*i.e.* yeast cell) and eukaryotic tissue or cell. In more preferred embodiments, the sample source is a mammalian tissue or cell and most preferably the sample source is a human tissue or cell. In another embodiment, the sample contains total RNA or total mRNA from the tissue or cells described herein. In a preferred embodiment the sample contains total RNA from up to 1×10^6 cells and the sample contains at least 10,000 different mRNAs. In preferred embodiments, the RNA polymerase promoter is a T7 RNA polymerase promoter sequence, a T3 RNA polymerase promoter sequence or a SP6 RNA polymerase promoter sequence and T7 RNA polymerase, T3 RNA polymerase and SP6 RNA polymerase are used to initiate RNA or DNA synthesis from these promoters, respectively. In specific embodiments, the

DNA polymerase used for second strand cDNA synthesis is any DNA polymerase known in the art and more preferably E. coli DNA polymerase I, Klenow fragment of E. coli DNA polymerase I, or T4 DNA polymerase. In another embodiment, the primers used in the method of the invention are each 4-50 nucleotides in length. Suitable oligonucleotide primers are described herein and can be modified by the skilled artisan. In a further embodiment, where the initial single-stranded RNA molecule contains a polyadenylated tail, one or more random primers containing an RNA polymerase promoter contacts the mRNA in combination with an oligo (dT) primer containing an RNA polymerase promoter under conditions for template driven enzymatic DNA synthesis to occur. In another embodiment, the oligo (dT) primer is stably associated with the surface of a solid support.

In a preferred embodiment, synthesized cRNA is labeled with a label, *i.e.*, any label known in the art and as described herein. In other preferred embodiments, the synthesized cRNA is labeled with a label, wherein the label is a fluorescent, radioactive, enzymatic, hapten, biotin, or digoxigenin label. More preferably, the label is fluorescent. The preferred label in the methods of this invention is a fluorophore, such as fluorescein isothiocyanate, lissamine, Cy3, Cy5, and rhodamine 110, with Cy3 and Cy5 particularly preferred.

In additional embodiments, the present invention provides a method for determining the presence or absence of a target mRNA from a sample and determining the effect of drug treatment on a target mRNA in a sample following the *in vitro* transcription step and production of cRNA.

In another embodiment, the present invention is directed to a method wherein the mRNA is extracted from at least one cell of interest, and further containing (e) contacting the cRNA produced in step (d) described above with an array containing one or more species of polynucleotide positioned at preselected sites on the array, under conditions conducive to hybridization; and (f) detecting any hybridization that occurs between said one or more species of polynucleotide and said cRNA. Methods of preparing and using microarrays are described herein.

METHODS FOR DETERMINING BIOLOGICAL RESPONSE PROFILES

In one embodiment, the invention is directed to a method for comparing the presence or amount of at least one mRNA of interest in a first sample and in a second sample, said first sample and said second sample each containing a plurality of different mRNAs from one or

more cells, said method containing: (a) synthesizing first strand cDNA by contacting at least one mRNA in said sample with (i) one or more first random oligonucleotide primers containing an RNA polymerase promoter that is sufficiently complementary to a sequence in the mRNA so as to prime first strand cDNA synthesis, and (ii) reverse transcriptase under conditions sufficient for reverse transcriptase activity to occur; (b) synthesizing double-stranded cDNA by contacting the first strand cDNA with (i) a second random primer not containing an RNA polymerase promoter that is sufficiently complementary to a sequence in the first strand cDNA so to prime second strand cDNA synthesis, and (ii) a DNA polymerase under conditions sufficient for DNA polymerase activity to occur; (c) isolating and purifying said double-stranded cDNA; (d) subjecting said purified double-stranded cDNA to *in vitro* transcription under conditions sufficient for template driven DNA transcription to occur, such that cRNA is produced; (e) labeling the cRNA produced in step (d) with a first label; (f) repeating steps (a)-(d) with said second sample; (g) labeling the cRNA produced in step (f) with a second label distinguishable from said first label; (h) detecting or measuring the mRNA of interest in the first sample by contacting the first cRNA labeled with said first label with a polynucleotide probe capable of hybridizing to said first cRNA of the mRNA of interest under conditions conducive to hybridization; and detecting any hybridization that occurs between said probe and said first cRNA; (i) detecting or measuring the mRNA of interest in the second sample by contacting the second cRNA labeled with said second label with said polynucleotide probe capable of hybridizing to said second cRNA of the mRNA of interest under conditions conducive to hybridization; and detecting any hybridization that occurs between said probe and said second cRNA; and (j) comparing the levels of the mRNA of interest detected or measured in said first sample with levels of the mRNA of interest detected or measured in said second sample.

The initial mRNA that serves as template in the first step may be present in a variety of different samples, where the sample will typically be derived from a physiological source. In preferred embodiments, the physiological source is prokaryotic tissue or cell (*i.e.* yeast cell) and eukaryotic tissue or cell. In more preferred embodiments, the sample source is a mammalian tissue or cell and most preferably the sample source is a human tissue or cell. In another embodiment, the sample contains total RNA or total mRNA from the tissue or cells described herein. In a preferred embodiment the sample contains total RNA from 1×10^6 cells or less and the sample contains at least 10,000 different mRNAs. In preferred embodiments, the RNA polymerase promoter is a T7 RNA polymerase promoter sequence, a

T3 RNA polymerase promoter sequence or a SP6 RNA polymerase promoter sequence and T7 RNA polymerase, T3 RNA polymerase and SP6 RNA polymerase are used to initiate RNA or DNA synthesis from these promoters, respectively. In specific embodiments, the DNA polymerase used for second strand cDNA synthesis is any DNA polymerase known in the art and more preferably E. coli DNA polymerase I, Klenow fragment of E. coli DNA polymerase I, or T4 DNA polymerase. In another embodiment, the primers used in the method of the invention are each 4-50 nucleotides in length. Suitable oligonucleotide primers are described herein and can be modified by the skilled artisan. In a further embodiment, where the initial single-stranded RNA molecule contains a polyadenylated tail, one or more random primers containing an RNA polymerase promoter contacts the mRNA in combination with an oligo (dT) primer containing an RNA polymerase promoter under conditions for template driven enzymatic DNA synthesis to occur. In another embodiment the oligo (dT) primer is stably associated with the surface of a solid support

In a preferred embodiment, synthesized first and second cRNAs is labeled with a label, *i.e.*, any label known in the art and described herein. In other preferred embodiments, the synthesized first and second cRNAs is labeled with a label wherein the label is a fluorescent, radioactive, enzymatic, hapten, biotin, or digoxigenin label. Preferably the label is fluorescent. The preferred label in the methods of this invention is a fluorophore, such as fluorescein isothiocyanate, lissamine, Cy3, Cy5, and rhodamine 110, with Cy3 and Cy5 particularly preferred.

In a preferred embodiment, a first aliquot of the first cRNA is labeled with a first fluorophore having a first emission spectrum, and a second aliquot of the second cRNA is labeled with a second fluorophore with a second emission spectrum differing from that of the first emission spectrum; more preferably the first fluorophore is Cy3 and the second fluorophore is Cy5.

In another embodiment, the present invention is directed to a method wherein in steps (h) and (i) as described above, the steps of contacting the first cRNA labeled with said first label with said polynucleotide probe, and contacting the second cRNA labeled with said second label with said polynucleotide probe, are carried out concurrently or sequentially. The term "concurrently" refers to two events occurring at the same time and the term "sequentially" and refers to succeeding or following in a particular order.

In preferred embodiments, the first sample contains mRNAs from cells that are diseased and wherein said second sample contains mRNAs from normal cells; or first sample

contains mRNAs from drug-treated cells and wherein said second sample contains mRNAs from untreated cells, wherein said levels of said mRNAs from these cells can be compared.

In another embodiment, the present invention is directed to a method wherein in steps (h) and (i) are carried out by a method containing contacting said first and second cRNAs with an array containing one or more species of polynucleotide probe positioned at preselected sites on the array, under conditions sufficient for hybridization to occur; and detecting any hybridization that occurs between said polynucleotide probes and said cRNAs. Methods of preparing and using microarrays are described herein.

The invention utilizes the ability to measure the responses of a biological system to a large variety of perturbations. Exemplary methods for measuring biological responses are described herein. One of skill in the art will appreciate that this invention is not limited to the following specific methods for measuring the responses of a biological system. In particular, the presence of cRNA(s) of interest (and thus mRNA(s) of interest in the sample) can be detected or measured by procedures including but not limited to Northern blotting or using bead-bound oligonucleotides as probes, or the use of polynucleotide microarrays. In additional embodiments, the synthesized labeled cRNA probes of the invention can be utilized for various other methodologies, including but not limited to dot blot analysis, differential hybridization, RT-PCR, and anti-sense RNA studies.

Dot blot analysis: For any type of dot blot experiment that employs cDNA, genomic DNA, mRNA, or oligos, one skilled in the art can use the invention to synthesize full-length coverage cRNA probes. For genomic DNA, this method can be used to ask evolutionary-type questions such as: "are there exons present in this organism that are similar to exons from this other organism, but less so for this one?"

Differential hybridization: When studying expression differences between cell types or tissues, it is sometimes useful to employ differential hybridization. A cDNA library displayed as bacteriophage plaques may be hybridized with the cRNA probes, synthesized and labeled by the methods of the present invention, from the cell type or tissue one wants to compare.

RT-PCR and anti-sense RNA studies: A reverse transcription reaction of an mRNA pool from a certain tissue can be primed using the random oligonucleotide primer of the present invention containing an RNA polymerase promoter. In an illustrative and non-limiting example, using a random-T7 primers, the next step, PCR, one would use a gene-specific upper primer directed at the 5 prime end of the desired mRNA, and a T7 promoter-

complementary primer. If used during stringent conditions (a long (27-30-mer) upper primer, high annealing temperature, or preferably a touch-down PCR protocol) one can isolate the full-length cDNA of a gene of choice. In addition, this cDNA is ready to be *in vitro* transcribed in the anti-sense direction using T7 RNA polymerase. The anti-sense RNA can be used in anti-sense RNA studies for a) use in secondary structure determination of the anti-sense RNA; b) use directly labeled with radioactivity or biotin as an anti-sense RNA probe for *in situ* hybridization; c) use in RNA Interference experiments in duplex with the corresponding sense mRNA, produced by conventional means; d) other anti-sense RNA experiments. Furthermore, a single reverse transcription reaction can be used for multiple different genes only by changing the upper primer. This is particularly useful if the tissue or cell type under study is scarce.

In one specific embodiment of the invention, one or more labels is introduced into the RNA during the *in vitro* transcription step, as described herein, to facilitate gene expression profiling. Gene expression can be profiled in any of several ways. The preferred method is to probe a DNA microarray with the labeled RNA transcripts generated as described above. A DNA microarray, or chip, is a microscopic array of DNA fragments or synthetic oligonucleotides, disposed in a defined pattern on a solid support, wherein they are amenable to analysis by standard hybridization methods (Schena, *BioEssays* 18: 427, 1996).

The DNA in a microarray may be derived from genomic or cDNA libraries, from fully sequenced clones, or from partially sequenced cDNAs known as expressed sequence tags (ESTs). Methods for obtaining such DNA molecules are generally known in the art (See, *e.g.*, Ausubel et al., eds., 1994, *Current Protocols in Molecular Biology*, vol. 2, Current Protocols Publishing, New York). Alternatively, oligonucleotides may be synthesized by conventional methods, such as phosphoramidite-based synthesis.

Gene expression profiling can be done for purposes of screening, diagnosis, staging of a disease, monitoring response to therapy, as well as for identifying genetic targets of drugs and of pathogens.

TRANSCRIPT ASSAY USING DNA ARRAYS

The methods of this invention are particularly useful for the analysis of gene expression profiles. For expression profiling, DNA microarrays are typically probed using mRNA, extracted and synthesized to cRNA from the cells whose gene expression profile it is

desired to analyze, using the methods of the invention (*i.e.* producing double-stranded DNA, synthesizing at least one cRNA molecule, labeling at least one synthesized cRNA molecule, producing 5' enriched cDNA library). To facilitate comparison between any two samples of interest, the mRNAs are typically labeled separately with fluorescent dyes that emit at different wavelengths, as described above.

Some embodiments of this invention are based on measuring the transcriptional rate of genes. Transcriptional rates can be measured by techniques of hybridization to arrays of nucleic acid or nucleic acid mimic probes. However measured, the result is either the absolute, relative amounts of transcripts or response data, including values representing RNA abundance ratios, which usually reflect DNA expression ratios (in the absence of differences in RNA degradation rates).

In various alternative embodiments of the present invention, aspects of the biological state other than the transcriptional state, such as the translational state, the activity state, or mixed aspects can be measured. Preferably, measurement of the transcriptional state is made by hybridization to transcript arrays, which are described herein. Certain other methods of transcriptional state measurement are also described herein.

In a preferred embodiment the present invention makes use of "transcript arrays" (also referred to herein as "microarrays"). Transcript arrays can be employed for analyzing the transcriptional state in a biological sample and especially for measuring the transcriptional states of a biological sample exposed to graded levels of a drug of interest or to graded perturbations to a biological pathway of interest.

In one embodiment, transcript arrays are produced by hybridizing detectably labeled polynucleotides representing the mRNA transcripts present in a cell (*e.g.*, fluorescently labeled cRNA that is synthesized by the methods of the present invention) to a microarray. A microarray is a surface with an ordered array of binding (*e.g.*, hybridization) sites for products of many of the genes in the genome of a cell or organism, preferably most or almost all of the genes. Microarrays can be made in a number of ways. However produced, microarrays share certain preferred characteristics. Specifically, the arrays are reproducible, allowing multiple copies of a given array to be produced and easily compared with each other. Preferably the microarrays are small, usually smaller than 5 cm², and they are made from materials that are stable under binding (*e.g.*, nucleic acid hybridization) conditions. A given binding site or unique set of binding sites in the microarray will specifically bind the product of a single gene in the cell. Although there may be more than one physical binding

site (hereinafter "site") per specific mRNA, for the sake of clarity the discussion below will assume that there is a single site.

In one embodiment, the microarray is an array of polynucleotide probes, the array containing a support with at least one surface and at least 100 different polynucleotide probes, each different polynucleotide probe containing a different nucleotide sequence and being attached to the surface of the support in a different location on the surface. Preferably, the nucleotide sequence of each of the different polynucleotide probes is in the range of 40 to 80 nucleotides in length. More preferably, the nucleotide sequence of each of the different polynucleotide probes is in the range of 50 to 70 nucleotides in length. Even more preferably, the nucleotide sequence of each of the different polynucleotide probes is in the range of 50 to 60 nucleotides in length.

In specific embodiments, the array contains polynucleotide probes of at least 2,000, 4,000, 10,000, 15,000, 20,000, 50,000, 80,000, or 100,000 different nucleotide sequences.

In another embodiment, the nucleotide sequence of each polynucleotide probe in the array is specific for a particular target polynucleotide sequence. In yet another embodiment, the target polynucleotide sequences contain expressed polynucleotide sequences of a cell or organism. In a specific embodiment, the cell or organism is a mammalian cell or organism. In another specific embodiment, the cell or organism is a human cell or organism.

In specific embodiments, the nucleotide sequences of the different polynucleotide probes of the array are specific for at least 50%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% of the genes in the genome of the cell or organism. Most preferably, the nucleotide sequences of the different polynucleotide probes of the array are specific for all of the genes in the genome of the cell or organism.

In specific embodiments, the polynucleotide probes of the array hybridize specifically and distinguishably to at least 10,000, to at least 20,000, to at least 50,000, different polynucleotide sequences, to at least 80,000, or to at least 100,000 different polynucleotide sequences.

In other specific embodiments, the polynucleotide probes of the array hybridize specifically and distinguishably to at least 90%, at least 95%, or at least 99% of the genes or gene transcripts of the genome of a cell or organism. Most preferably, the polynucleotide probes of the array hybridize specifically and distinguishably to the genes or gene transcripts of the entire genome of a cell or organism.

In specific embodiments, the array has at least 100, at least 250, at least 1,000, or at least 2,500 probes per 1 cm², preferably all or at least 25% or 50% of which are different from each other. In another embodiment, the array is a positionally addressable array (in that the sequence of the polynucleotide probe at each position is known).

In another embodiment, the nucleotide sequence of each polynucleotide probe in the array is a DNA sequence. In another embodiment, the DNA sequence is a single-stranded DNA sequence. The DNA sequence may be, *e.g.*, a cDNA sequence, or a synthetic sequence. In a preferred embodiment, the nucleotide sequence of each polynucleotide probe in the array is a oligodeoxyribonucleotides sequence.

It will be appreciated by those skilled in the art that when cRNA complementary to the RNA of a cell is synthesized and hybridized to a microarray under suitable hybridization conditions, the level of hybridization to the site in the array corresponding to any particular gene will reflect the prevalence in the cell of mRNA transcribed from that gene. For example, when detectably labeled (*e.g.*, with a fluorophore) cRNA complementary to the total cellular mRNA is hybridized to a microarray, the site on the array corresponding to a gene (*i.e.*, capable of specifically binding the product of the gene) that is not transcribed in the cell will have little or no signal (*e.g.*, fluorescent signal), and a gene for which the encoded mRNA is prevalent will have a relatively strong signal.

In preferred embodiments, cRNAs from two different cells are hybridized to the binding sites of the microarray. In the case of monitoring drug responses one biological sample is exposed to a drug and another biological sample of the same type is not exposed to the drug. In the case of pathway responses one cell is exposed to a pathway perturbation and another cell of the same type is not exposed to the pathway perturbation. The cRNA derived from each of the two cell types are differently labeled so that they can be distinguished. In one embodiment, for example, cRNA from a cell treated with a drug (or exposed to a pathway perturbation) is synthesized using a fluorescein-labeled NTP, and cRNA from a second cell, not drug-exposed, is synthesized using a rhodamine-labeled NTP. When the two cRNAs are mixed and hybridized to the microarray, the relative intensity of signal from each cRNA set is determined for each site on the array, and any relative difference in abundance of a particular mRNA detected.

In the example described above, the cRNA from the drug-treated (or pathway perturbed) cell will fluoresce green when the fluorophore is stimulated and the cRNA from the untreated cell will fluoresce red. As a result, when the drug treatment has no effect, either

directly or indirectly, on the relative abundance of a particular mRNA in a cell, the mRNA will be equally prevalent in both cells and, upon reverse transcription, red-labeled and green-labeled cRNA will be equally prevalent. When hybridized to the microarray, the binding site(s) for that species of RNA will emit wavelengths characteristic of both fluorophores (and appear brown in combination). In contrast, when the drug-exposed cell is treated with a drug that, directly or indirectly, increases the prevalence of the mRNA in the cell, the ratio of green to red fluorescence will increase. When the drug decreases the mRNA prevalence, the ratio will decrease.

The use of a two-color fluorescence labeling and detection scheme to define alterations in gene expression has been described, *e.g.*, in Schena et al., 1995, Science 270:467-470, which is incorporated herein by reference in its entirety. An advantage of using cRNA labeled with two different fluorophores is that a direct and internally controlled comparison of the mRNA levels corresponding to each arrayed gene in two cell states can be made, and variations due to minor differences in experimental conditions (*e.g.*, hybridization conditions) will not affect subsequent analyses. However, it will be recognized that it is also possible to use cRNA from a single cell, and compare, for example, the absolute amount of a particular mRNA in, *e.g.*, a drug-treated or pathway-perturbed cell and an untreated cell.

In a preferred embodiment, cRNA from two different cells are hybridized to the binding sites of a microarray. In this embodiment, one biological sample is from a diseased tissue or cell and the other biological sample of the same type is from a normal non-diseased tissue or cell. As described herein, these samples can be labeled in one embodiment and subjected to the two-color fluorescence labeling and detection scheme to define alterations in gene expression. Thus, synthesis of a labeled cRNA molecule using the random oligonucleotides of the invention containing an RNA polymerase promoter can be used in methods of diagnosis, wherein a synthesized cRNA sequence is complementary to a sequence (*e.g.*, genomic) of an infectious disease agent, *e.g.* of human disease including but not limited to viruses, bacteria, parasites, and fungi, thereby diagnosing the presence of the infectious agent in a sample of nucleic acid from a patient. The target nucleic acid can be genomic or cDNA or mRNA or synthetic, human or animal, or of a microorganism, etc. In another embodiment that can be used in the diagnosis or prognosis of a disease or disorder, the target sequence is a wild type human genomic or RNA or cDNA sequence, mutation of which is implicated in the presence of a human disease or disorder, or alternatively, can be the mutated sequence. In such an embodiment, the mutation can be an insertion, substitution, and/or

deletion of one or more nucleotides, or a translocation.

PREPARATION OF MICROARRAYS

Microarrays are known in the art and consist of a surface to which probes that correspond in sequence to gene products (*e.g.*, cDNAs, mRNAs, cRNAs, polypeptides, and fragments thereof), can be specifically hybridized or bound at a known position. In one embodiment, the microarray is an array (*i.e.*, a matrix) in which each position represents a discrete binding site for a product encoded by a gene (*e.g.*, a protein or RNA), and in which binding sites are present for products of most or almost all of the genes in the organism's genome. In a preferred embodiment, the "binding site" (hereinafter, "site") is a nucleic acid or nucleic acid analogue to which a particular cognate cRNA can specifically hybridize. The nucleic acid or analogue of the binding site can be, *e.g.*, a synthetic oligomer, a full-length cRNA, a less-than full length cRNA, or a gene fragment.

In one embodiment, the microarray contains binding sites for products of all or almost all genes in the target organism's genome. This microarray will have binding sites corresponding to at least about 50% of the genes in the genome, often at least about 75%, more often at least about 85%, even more often more than about 90%, and most often at least about 99%.

Such comprehensiveness, however, is not necessarily required. In another embodiment, the microarray contains binding sites for products of human genes. This microarray will have binding sites corresponding to at least about 5-10% of the genes in the genome, preferably at least about 10-15%, and more preferably at least about 40%.

Preferably, the microarray has binding sites for genes relevant to the action of a drug of interest, disease of interest, or in a biological pathway of interest. A "gene" is identified as an open reading frame (ORF) of preferably at least 50, 75, or 99 amino acids from which a mRNA is transcribed in the organism (*e.g.*, if a single cell) or in some cell in a multicellular organism. The number of genes in a genome can be estimated from the number of mRNAs expressed by the organism, or by extrapolation from a well-characterized portion of the genome. When the genome of the organism of interest has been sequenced, the number of ORFs can be determined and mRNA coding regions identified by analysis of the DNA sequence. For example, the *Saccharomyces cerevisiae* genome has been completely sequenced and is reported to have approximately 6275 open reading frames (ORFs) longer

than 99 amino acids. Analysis of these ORFs indicates that there are 5885 ORFs that are likely to specify protein products (Goffeau et al., 1996, Science 274:546-567, which is incorporated by reference in its entirety). In contrast, the human genome is estimated to contain approximately 10^5 genes.

PREPARATION OF NUCLEIC ACIDS FOR MICROARRAYS

As noted above, the "binding site" to which a particular cognate cRNA specifically hybridizes is usually a nucleic acid or nucleic acid analogue attached at that binding site. In one embodiment, the binding sites of the microarray are DNA polynucleotides corresponding to at least a portion of each gene in an organism's genome. These DNAs can be obtained by, *e.g.*, polymerase chain reaction (PCR) amplification of gene segments from genomic DNA, cDNA (*e.g.*, by RT-PCR), or cloned sequences. PCR primers are chosen, based on the known sequence of the genes or cDNA, that result in amplification of unique fragments (*i.e.*, fragments that do not share more than bases of contiguous identical sequence with any other fragment on the microarray). Computer programs are useful in the design of primers with the required specificity and optimal amplification properties. See, *e.g.*, Oligo version 5.0 (National Biosciences). In the case of binding sites corresponding to very long genes, it will sometimes be desirable to amplify segments near the 3' end of the gene so that when oligo-dT primed cDNA probes are hybridized to the microarray, less-than-full-length probes will bind efficiently. Typically each gene fragment on the microarray will be between about 50 bp and about 2000 bp, more typically between about 100 bp and about 1000 bp, and usually between about 300 bp and about 800 bp in length.

PCR methods are well known and are described, for example, in Innis et al., eds., 1990, PCR Protocols: A Guide to Methods and Applications, Academic Press Inc., San Diego, Calif., which is incorporated herein by reference in its entirety. It will be apparent that computer controlled robotic systems are useful for isolating and amplifying nucleic acids.

An alternative means for generating the nucleic acid for the microarray is by synthesis of synthetic polynucleotides or oligonucleotides, *e.g.*, using N-phosphonate or phosphoramidite chemistries (*e.g.*, Froehler et al., 1986, Nucleic Acid Res 14:5399-5407). Synthetic sequences are between about 15 and about 100 bases in length, preferably between about 20 and about 50 bases.

In some embodiments, synthetic nucleic acids include non-natural bases, *e.g.*, inosine. Where the particular base in a given sequence is unknown or is polymorphic, a universal base, such as inosine or 5-nitroindole, may be substituted. Additionally, it is possible to vary the charge on the phosphate backbone of the oligonucleotide, for example, by thiolation or methylation, or even to use a peptide rather than a phosphate backbone. The making of such modifications is within the skill of one trained in the art.

As noted above, nucleic acid analogues may be used as binding sites for hybridization. An example of a suitable nucleic acid analogue is peptide nucleic acid (see, *e.g.*, Egholm et al., 1993, *Nature* 365:566-568 and U.S. Pat. No. 5,539,083).

In an alternative embodiment, the binding (hybridization) sites are made from plasmid or phage clones of genes, cDNAs (*e.g.*, expressed sequence tags), or inserts therefrom (Nguyen et al., 1995, *Genomics* 29:207-209). In yet another embodiment, the polynucleotide of the binding sites is RNA.

ATTACHING NUCLEIC ACIDS TO THE SOLID SURFACE

The nucleic acid or analog is attached to a solid support, which may be made from glass, silicon, plastic (*e.g.*, polypropylene, nylon, polyester), polyacrylamide, nitrocellulose, cellulose acetate or other materials. In general, non-porous supports, and glass in particular, are preferred. The solid support may also be treated in such a way as to enhance binding of oligonucleotides thereto, or to reduce non-specific binding of unwanted substances thereto. Preferably, the glass support is treated with polylysine or silane to facilitate attachment of oligonucleotides to the slide.

Methods of immobilizing DNA on the solid support may include direct touch, micropipetting (Yershov et al., *Proc. Natl. Acad. Sci. USA* (1996) 93(10):4913-4918), or the use of controlled electric fields to direct a given oligonucleotide to a specific spot in the array (U.S. Pat. No. 5,605,662.). DNA is typically immobilized at a density of 100 to 10,000 oligonucleotides per cm² and preferably at a density of about 1000 oligonucleotides per cm².

A preferred method for attaching the nucleic acids to a surface is by printing on glass plates, as is described generally by Schena et al., 1995, *Science* 270:467-470. This method is especially useful for preparing microarrays of cDNA. *See also* DeRisi et al., 1996, *Nature Genetics* 14:457-460; Shalon et al., 1996, *Genome Res.* 6:639-645; and Schena et al., *Proc. Natl. Acad. Sci. USA*, 1996, 93(20):10614-19.)

In a preferred alternative to immobilizing pre-fabricated oligonucleotides onto a solid support, it is possible to synthesize oligonucleotides directly on the support (Maskos et al., Nucl. Acids Res. 21: 2269-70, 1993; Fodor et al., Science 251: 767-73, 1991; Lipshutz et al., 1999, Nat. Genet. 21(1 Suppl):20-4). Among methods of synthesizing oligonucleotides directly on a solid support, particularly preferred methods include photolithography (see Fodor et al., Science 251: 767-73, 1991; McGall et al., Proc. Natl. Acad. Sci. (USA) 93: 13555-60, 1996) and piezoelectric printing (Lipshutz et al., 1999, Nat. Genet. 21(1 Suppl):20-4), with the piezoelectric method being the most preferred.

In one embodiment, a high-density oligonucleotide array is employed. Techniques are known for producing arrays containing thousands of oligonucleotides complementary to defined sequences, at defined locations on a surface using photolithographic techniques for synthesis in situ (see, Fodor et al., 1991, Science 251:767-773; Pease et al., 1994, Proc. Natl. Acad. Sci. USA 91:5022-5026; Lockhart et al., 1996, Nature Biotechnol. 14:1675-80; U.S. Pat. No. 5,578,832; U.S. Pat. No. 5,556,752; and U.S. Pat. No. 5,510,270; each of which is incorporated by reference in its entirety) or other methods for rapid synthesis and deposition of defined oligonucleotides (Lipshutz et al., 1999, Nat. Genet. 21(1 Suppl):20-4.)

When these methods are used, oligonucleotides (e.g., 20-mers) of known sequence are synthesized directly on a surface such as a derivatized glass slide. Usually, the array produced contains multiple probes against each target transcript. Oligonucleotide probes can be chosen to detect alternatively spliced mRNAs or to serve as various type of control.

Another preferred method of making microarrays is by use of an inkjet printing process to synthesize oligonucleotides directly on a solid phase, as described, e.g., in U.S. Pat. No. 6,419,883, which is incorporated by reference herein in its entirety.

Other methods for making microarrays, e.g., by masking (Maskos and Southern, 1992, Nuc. Acids Res. 20:1679-1684), may also be used. In principal, any type of array, for example, dot blots on a nylon hybridization membrane (see Sambrook et al., 1989, Molecular Cloning--A Laboratory Manual (2nd Ed.), Vols. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.), could be used, although, as will be recognized by those of skill in the art, very small arrays will be preferred because hybridization volumes will be smaller.

HYBRIDIZATION TO MICROARRAYS

Nucleic acid hybridization and wash conditions are optimally chosen so that the probe "specifically binds" or "specifically hybridizes" to a specific array site, *i.e.*, the probe hybridizes, duplexes or binds to a sequence array site with a complementary nucleic acid sequence but does not hybridize to a site with a non-complementary nucleic acid sequence. As used herein, one polynucleotide sequence is considered complementary to another when, if the shorter of the polynucleotides is less than or equal to 25 bases, there are no mismatches using standard base-pairing rules or, if the shorter of the polynucleotides is longer than 25 bases, there is no more than a 5% mismatch. Preferably, the polynucleotides are perfectly complementary (*i.e.* no mismatches). It can easily be demonstrated that specific hybridization conditions result in specific hybridization by carrying out a hybridization assay including negative controls (see, *e.g.*, Shalon et al., 1996, *Genome Research* 6:639-645, and Chee et al., 1996, *Science* 274:610-614).

Optimal hybridization conditions will depend on the length (*e.g.*, oligomer versus polynucleotide greater than 200 bases) and type (*e.g.*, RNA, DNA, PNA) of labeled probe and immobilized polynucleotide or oligonucleotide. General parameters for specific (*i.e.*, stringent) hybridization conditions for nucleic acids are described in Sambrook et al. (1989, *Molecular Cloning--A Laboratory Manual* (2nd Ed.), Vols. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.) and in Ausubel et al. (1987, *Current Protocols in Molecular Biology*, Greene Publishing, Media, Pa., and Wiley-Interscience, New York). When the cDNA microarrays of Schena et al. (1996, *Proc. Natl. Acad. Sci. USA*, 93:10614-19) are used, typical hybridization conditions are hybridization in 533 SSC plus 0.2% SDS at 65° C. for 4 hours followed by washes at 25° C in low stringency wash buffer (1xSSC plus 0.2% SDS) followed by 10 minutes at 25° C in high stringency wash buffer (0.1xSSC plus 0.2% SDS) (Schena et al., 1996, *Proc. Natl. Acad. Sci. USA*, 93:10614-19). Useful hybridization conditions are also provided in, *e.g.*, Tijssen, 1993, *Hybridization With Nucleic Acid Probes*, Elsevier Science Publishers B.V., Amsterdam and New York, and Kricka, 1992, *Nonisotopic DNA Probe Techniques*, Academic Press, San Diego, Calif.

Although simultaneous hybridization of differentially labeled mRNA samples is preferred, it is also possible to use a single label and to perform hybridizations sequentially rather than simultaneously.

SINGLE DETECTION AND DATA ANALYSIS

When fluorescently labeled probes are used, the fluorescence emissions at each site of a transcript array can preferably be detected by scanning confocal laser microscopy. In one embodiment, a separate scan, using the appropriate excitation line, is carried out for each of the two fluorophores used. Alternatively, a laser can be used that allows simultaneous specimen illumination at wavelengths specific to the two fluorophores and emissions from the two fluorophores can be analyzed simultaneously (see Shalon et al., 1996, *Genome Research* 6:639-645, which is incorporated by reference in its entirety for all purposes). In a preferred embodiment, the arrays are scanned with a laser fluorescent scanner with a computer controlled X-Y stage and a microscope objective. Sequential excitation of the two fluorophores is achieved with a multi-line, mixed gas laser and the emitted light is split by wavelength and detected with two photomultiplier tubes. Fluorescence laser scanning devices are described in Shalon et al., 1996, *Genome Res.* 6:639-645 and in other references cited herein. Alternatively, the fiber-optic bundle described by Ferguson et al., 1996, *Nature Biotechnol.* 14:1681-1684, may be used to monitor mRNA abundance levels at a large number of sites simultaneously.

Signals are recorded and, in a preferred embodiment, analyzed by computer, *e.g.*, using a 12 bit analog to digital board. In one embodiment the scanned image is despeckled using a graphics program (*e.g.*, Hijaak Graphics Suite) and then analyzed using an image gridding program that creates a spreadsheet of the average hybridization at each wavelength at each site. If necessary, an experimentally determined correction for "cross talk" (or overlap) between the channels for the two fluors may be made. For any particular hybridization site on the transcript array, a ratio of the emission of the two fluorophores can be calculated. The ratio is independent of the absolute expression level of the cognate gene, but is useful for genes whose expression is significantly modulated by drug administration, gene deletion, or any other tested event.

According to the method of the invention, the relative abundance of an mRNA in two biological samples is scored as a perturbation and its magnitude determined (*i.e.*, the abundance is different in the two sources of mRNA tested), or as not perturbed (*i.e.*, the relative abundance is the same). In various embodiments, a difference between the two sources of RNA of at least a factor of about 25% (RNA from one source is 25% more abundant in one source than the other source), more usually about 50%, even more often by a factor of about 2 (twice as abundant), 3 (three times as abundant) or 5 (five times as abundant) is scored as a perturbation.

Preferably, in addition to identifying a perturbation as positive or negative, it is advantageous to determine the magnitude of the perturbation. This can be carried out, as noted above, by calculating the ratio of the emission of the two fluorophores used for differential labeling, or by analogous methods that will be readily apparent to those of skill in the art.

In one embodiment, two samples, each labeled with a different fluor, are hybridized simultaneously to permit differential expression measurements. If neither sample hybridizes to a given spot in the array, no fluorescence will be seen. If only one hybridizes to a given spot, the color of the resulting fluorescence will correspond to that of the fluor used to label the hybridizing sample (for example, green if the sample was labeled with Cy3, or red, if the sample was labeled with Cy5). If both samples hybridize to the same spot, an intermediate color is produced (for example, yellow if the samples were labeled with fluorescein and rhodamine). Then, applying methods of pattern recognition and data analysis known in the art, it is possible to quantify differences in gene expression between the samples. Methods of pattern recognition and data analysis are described in *e.g.*, U.S. Pat. No. 6,203,569 and U.S. Pat. No. 6,271,002; which are incorporated herein by reference in their entireties.

KITS FOR THE SYNTHESIS OF NUCLEOTIDE SEQUENCES

An additional aspect provided by the subject invention are kits for carrying out the subject methods. In specific embodiments, the kits contain in one or more containers: (a) a mixture of first random oligonucleotide primers, each first primer containing a 3' end sequence of 10-50 nucleotides, said 3' end sequence having an A, a G, a T, or a C nucleotide randomly present in each position of said 3' end sequence; and each first primer having a 5' end sequence of 10-50 nucleotides, said 5' end sequence containing an RNA polymerase promoter; (b) a mixture of oligo (dT) primers, each oligo (dT) primer containing a 3' end sequence of 10-50 nucleotides, said 3' end sequence having a T nucleotide present in each position of said 3' end sequence; and each oligo (dT) primer having a 5' end sequence of 10-50 nucleotides, said 5' end sequence containing an RNA polymerase promoter; and (c) a mixture of second random oligonucleotide primers, each second primer having a sequence of 10-50 nucleotides, said sequence containing an A, a G, a T, or a C nucleotide randomly present in each position of said sequence. Oligonucleotide primers can be in any form, *e.g.*, lyophilized, or in solution (*e.g.*, a distilled water or buffered solution), etc. Oligonucleotides

for use in the same reaction can be combined in a single container or can be in separate containers.

In another embodiment, the kit contains in a separate container an RNA polymerase specific to said RNA polymerase promoter. In another embodiment, the kit contains an RNA polymerase for *in vitro* transcription of double-stranded cDNA. More preferably, the RNA polymerase is T7 RNA polymerase, T3 RNA polymerase or SP6 RNA polymerase.

In another embodiment, the kit contains in a separate container a DNA polymerase. More preferably, the DNA polymerase is E. coli DNA polymerase I, Klenow fragment of E. coli DNA polymerase I, or T4 DNA polymerase.

In another embodiment, the kit contains a buffer for reverse transcription or first strand DNA synthesis, a buffer for DNA synthesis, a buffer for *in vitro* transcription of double-stranded cDNA, DNA ligase, an RNase (*i.e.* RNase H).

In another embodiment, the kit contains deoxyribonucleotide triphosphate molecules (dNTPs) and ribonucleotide triphosphate molecules (rNTPs). More preferably the dNTPs are dATP, dCTP, dTTP and dGTP and the rNTPs are rATP, rCTP, rUTP and rGTP. In a preferred embodiment, the dNTPs and rNTPs are labeled, *i.e.*, any label known in the art and as described herein. In other preferred embodiments, the label is a fluorescent, radioactive, enzymatic, hapten, biotin, or digoxigenin label, more preferably the label is fluorescent. The preferred label is a fluorophore, such as fluorescein isothiocyanate, lissamine, Cy3, Cy5, and rhodamine 110, with Cy3 and Cy5 particularly preferred.

The kit optionally further contains a set of directions for carrying out reverse transcription or first strand cDNA synthesis and second strand cDNA synthesis; as well as a set of directions for carrying out the *in vitro* transcription of double-stranded cDNA into cRNA. The instructions may be associated with a package insert and/or the packaging of the kit or the components thereof.

The kit optionally further contains a control nucleic acid, and/or a microarray, and/or means for stimulating and detecting fluorescent light emissions from fluorescently labeled RNA, and/or expression profile projection and analysis software capable of being loaded into the memory of a computer system.

ANALYTIC KIT IMPLEMENTATION

In a preferred embodiment, the methods of this invention can be implemented by use of kits containing random oligonucleotide primers of the invention containing an RNA polymerase promoter and microarrays. The microarrays contained in such kits contain a solid phase, *e.g.*, a surface, to which probes are hybridized or bound at a known location of the solid phase. Preferably, these probes consist of nucleic acids of known, different sequence, with each nucleic acid being capable of hybridizing to an RNA species or to a cDNA species derived therefrom. In particular, the probes contained in the kits of this invention are nucleic acids capable of hybridizing specifically to nucleic acid sequences derived from RNA species which are known to increase or decrease in response to perturbations to the particular protein whose activity is determined by the kit. The probes contained in the kits of this invention preferably substantially exclude nucleic acids which hybridize to RNA species that are not increased in response to perturbations to the particular protein whose activity is determined by the kit.

In another preferred embodiment, the kits of the invention further contains expression profile projection and analysis software capable of being loaded into the memory of a computer system. An example of such a system is described in reference in U.S. Pat. No. 6,271,002, which is incorporated herein by reference in its entirety. Preferably, the expression profile analysis software contained in the kits of this invention, is essentially identical to the expression profile analysis software 512 described in reference in U.S. Pat. No. 6,271,002.

Alternative kits for implementing the analytic methods of this invention will be apparent to those of skill in the art and are intended to be comprehended within the accompanying claims. In particular, the accompanying claims are intended to include the alternative program structures for implementing the methods of this invention that will be readily apparent to one of skill in the art.

The invention is further defined by reference to the following examples. It is understood that the foregoing detailed description and the following examples are illustrative only and are not to be taken as limitations upon the scope of the invention. It will be apparent to those skilled in the art that many modifications, both to the materials and methods, may be practiced without departing from the purpose and interest of the invention. Further, all patents, patent applications and publications cited herein are incorporated herein by reference.

EXAMPLES

Example 1.**Labeling Protocol:****1. First Strand cDNA Synthesis**

This 40 μ l reaction converts a single-stranded RNA molecule (*i.e.* mRNA sequences) to first strand cDNA from up to 5 μ g of mRNA (or up to 30 μ g total RNA). For best results use only high quality RNA and nuclease-free reagents and materials. RNA can be isolated according to any method well known in the art and as described herein. Primers are random oligonucleotide primers containing an RNA polymerase promoter as described herein. Random primers can be used alone or in combination with an oligo (dT) primer containing an RNA polymerase promoter up to 2 μ g total primer. Note that the procedure calls for a Reverse Transcriptase with no RNase H activity.

- 1) Add the following in a 1.5 ml micro-centrifuge tube:

Primers (2 μ g)	4 μ l
Water, nuclease-free	x μ l
mRNA (up to 5 μ g)	y μ l
Final volume	20 μ l

- 2) Heat at 68 °C for 10 minutes. Chill on ice-water slush. Spin down briefly, and keep on ice.
- 3) In a separate tube, mix the following components (*i.e.* SuperScript II Reverse Transcriptase and buffer from GIBCO BRL/Roche). Do not use nuclease inhibitors, as this will interfere with the second strand synthesis. 5X First-Strand buffer comprises 250 mM Tris-HCl (pH 8.3), 375 mM KCl, 15 mM MgCl₂. The volume of water depends on the amount of SuperScript II added in Step 6.

5X First-Strand buffer	8 μ l
0.1 M DTT	4 μ l
10 mM dNTP mix	2 μ l
Water (nuclease-free)	x μ l

- 4) Mix the contents of the tube by gently vortexing, spin briefly and add to the Primer/RNA mixture from step 2.
- 5) Place the tube at 42 °C for 2 minutes.
- 6) Add SuperScript II RT (GIBCO BRL/Roche); the amount of SuperScript II RT varies with the amount of initial RNA (use 200 units of SuperScript II RT for each μ g of mRNA).

SuperScript II RT (200 U/ μ l)	x μ l
------------------------------------	-----------

Final volume:

40 μ l

- 7) Incubate at 42 to 50 °C for 1 hour. Higher temperature dissolves some secondary RNA structures.
- 8) Chill on ice to stop reaction and immediately proceed to second strand synthesis.

2. Second Strand cDNA Synthesis

The yield of the second strand reaction is typically > 80% relative to the amount of first strand cDNA synthesized.

- 1) On ice, add the following reagents to the first strand reaction, in this order.

For the composition of 10X Second Strand buffer see New England Biolabs catalog (same as DNA ligase buffer).

Water (nuclease-free)	86 μ l
10X Second Strand buffer	15 μ l
10 mM dNTP	3 μ l
<i>E. coli</i> DNA ligase (10 U/ μ l)	1 μ l
<i>E. coli</i> DNA polymerase I (10 U/ μ l)	4 μ l
<i>E. coli</i> RNase H (2 U/ μ l)	1 μ l

Final volume

150 μ l

- 2) Mix gently, spin down briefly, and incubate at 16 °C for 2-3 hours.
- 3) Without allowing the temperature to rise, add 4 μ l of T4 DNA polymerase (3 U/ μ l) and incubate at 16 °C for 5 minutes.
- 4) Stop the reaction with 10 μ l of 0.5 M EDTA and proceed with cDNA purification immediately.

The RNA is destroyed by adding 30 μ l of 1 M NaOH / 2 mM EDTA. Incubate at 65 °C for 10 minutes. Neutralize with 30 μ l of 1N HCl, and 30 μ l of 1 M Tris, pH 7.

3. cDNA Purification

- 1) Add an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1; pH 8). Vortex and centrifuge at room temperature for 2 minutes at full speed. Carefully recover the upper phase and transfer to a clean 1.5 ml microcentrifuge tube.

- 2) Extract with an equal volume of chloroform to remove traces of phenol. Transfer upper phase to a new 1.5 ml microcentrifuge tube.
- 3) Add 1/10 volume of sodium acetate (3 M, pH 5.2), 1 μ l of glycogen (10 μ g/ μ l), and 2.5 volume of cold 100% ethanol. Vortex and leave at -20 °C for at least 1 hour, preferably over night.
- 4) Centrifuge at 13K for 60 minutes at 4 °C.
- 5) Discard the supernatant and add 200 μ l of cold 70% ethanol to the pellet. Carefully decant and discard the supernatant.
- 6) Air-dry the double-strand cDNA pellet at room temperature for 10 minutes.
- 7) Resuspend the pellet in 50 μ l of MilliQ filtered water.
- 8) Purify the double-strand cDNA using a Micro-Spin column.
- 9) Lyophilize the flow-through to 7.5 μ l in a Speedvac, or precipitate according to Step 10.
- 10) Adjust the sample to 50 μ l final volume with nuclease-free water. Add 5 μ l of sodium acetate (3 M, pH 5.2), 1 μ l of the supplied Glycogen solution (10 μ g/ μ l), and 125 μ l of ice-chilled 100% ethanol. Vortex and precipitate at -20 °C for at least 1 hour, preferably over night.

Centrifuge at 13K for 60 minutes at 4 °C. Discard the supernatant and gently add 200 μ l of ice-chilled 70% ethanol (diluted in nuclease-free water) to the pellet.

Gently discard the supernatant and completely air-dry the double-strand cDNA pellet at room temperature for 10 minutes in a clean and dust-free environment.

Resuspend pellet in 7.5 μ l of nuclease-free water by pipeting up and down several times.

- 11) Leave the sample on ice and proceed with the transcription procedure, or keep samples at -80 °C for long-term storage.

4. *In vitro* Transcription

This protocol produces biotin-labeled cRNA probes from the double strand cDNA template synthesized above. One can use a high yield transcription kit such as T7 MEGAscript[™] from Ambion Inc. (cat. #1334). Please see the instructions of the manufacturer. Biotin-labeling is meant to be illustrative and not to limit the invention. The procedure can be modified by one of ordinary skill in the art to introduce any label as described herein. Furthermore, it may be necessary to introduce the double-stranded cDNA into a vector suitable for *in vitro*

transcription. This can be accomplished by any suitable method and any suitable vector well known to the skilled artisan or as described herein.

- 1) Add the following MEGAscript™ reagents to the ds-DNA sample from Step 11. Proceed at room temperature to avoid precipitation of DNA with the spermidine in Transcription Buffer. Keep T7 Enzyme Mix on ice before adding to translation mixture:

		<u>Final Concentration</u>
Water (nuclease-free)	___ μ l	
10X Transcription Buffer	3.0 μ l	
rATP (75 mM)	3.0 μ l	7.5 mM
rGTP (75 mM)	3.0 μ l	7.5 mM
rUTP (75 mM)	3.0 μ l	7.5 mM
rCTP (15 mM)	3.0 μ l	1.5 mM
Biotin-rCTP (10 mM)	4.5 μ l	1.5 mM
<u>T7 Enzyme Mix</u>	<u>3.0 μl</u>	
Total volume:	<u>30 μl</u>	

- 2) Incubate for 4 hours at 37 °C.
- 3) Adjust the volume of the transcription mixture to 40 μ l final volume using water.
- 4) Purify by using one of the Micro-Spin columns (provided).
- 5) Adjust flow-through to 50 μ l final volume with nuclease-free water. Keep probe at -80°C, unless used immediately for hybridization.
- 6) The final yield with Ambion's MEGAscript™ kit is typically 25-35 μ g of cRNA transcript per 1 μ g of mRNA.

OTHER EMBODIMENTS

Although particular embodiments have been disclosed herein in detail, this has been done by way of example for purposes of illustration only, and is not intended to be limiting with respect to the scope of the appended claims, which follow. In particular, it is contemplated by the inventors that various substitutions, alterations, and modifications may be made to the invention without departing from the spirit and scope of the invention as defined by the claims. Other aspects, advantages, and modifications are considered to be within the scope of the following claims.

What is claimed is:

1. A method for comparing the presence or amount of at least one mRNA of interest in a first sample and in a second sample, said first sample and said second sample each containing a plurality of different mRNAs from one or more cells, said method comprising:

- (a) synthesizing first strand cDNA by contacting at least one mRNA in said sample with (i) one or more first random oligonucleotide primers comprising an RNA polymerase promoter that is sufficiently complementary to a sequence in the mRNA so as to prime first strand cDNA synthesis, and (ii) reverse transcriptase under conditions sufficient for reverse transcriptase activity to occur;
- (b) synthesizing double-stranded cDNA by contacting the first strand cDNA with (i) a second random primer not comprising an RNA polymerase promoter that is sufficiently complementary to a sequence in the first strand cDNA so to prime second strand cDNA synthesis, and (ii) a DNA polymerase under conditions sufficient for DNA polymerase activity to occur;
- (c) isolating and purifying said double-stranded cDNA;
- (d) subjecting said purified double-stranded cDNA to *in vitro* transcription under conditions sufficient for template driven DNA transcription to occur, such that cRNA is produced;
- (e) labeling the cRNA produced in step (d) with a first label;
- (f) repeating steps (a)-(d) with said second sample;
- (g) labeling the cRNA produced in step (f) with a second label distinguishable from said first label;
- (h) detecting or measuring the mRNA of interest in the first sample by contacting the first cRNA labeled with said first label with a polynucleotide probe capable of hybridizing to said first cRNA of the mRNA of interest under conditions conducive to hybridization; and detecting any hybridization that occurs between said probe and said first cRNA;
- (i) detecting or measuring the mRNA of interest in the second sample by contacting the second cRNA labeled with said second label with said polynucleotide probe capable of hybridizing to said second cRNA of the mRNA of interest under conditions conducive to hybridization; and detecting any hybridization that occurs between said probe and said second cRNA; and
- (j) comparing the levels of the mRNA of interest detected or measured in said first sample with levels of the mRNA of interest detected or measured in said second sample.

2. The method of claim 1 wherein said sample contains total RNA or total mRNA from cells.

3. The method of claim 2 wherein said cells are prokaryotic cells.

4. The method of claim 3 wherein said prokaryotic cells are yeast cells.
5. The method of claim 2 wherein said cells are eukaryotic cells.
6. The method of claim 5 wherein said eukaryotic cells are mammalian cells.
7. The method of claim 6 wherein said mammalian cells are human cells.
8. The method of claim 1 wherein said sample contains total RNA from 1×10^6 cells or less.
9. The method of claim 1 wherein said sample contains at least 10,000 different mRNAs.
10. The method of claim 1, wherein the RNA polymerase promoter is a T7 RNA polymerase promoter sequence and the RNA polymerase is T7 RNA polymerase.
11. The method of claim 1, wherein the RNA polymerase promoter is a T3 RNA polymerase promoter sequence and the RNA polymerase is T3 RNA polymerase.
12. The method of claim 1, wherein the RNA polymerase promoter is an SP6 RNA polymerase promoter sequence and the RNA polymerase is SP6 RNA polymerase.
13. The method of claim 1, wherein said DNA polymerase is E. coli DNA polymerase I, Klenow fragment of E. coli DNA polymerase I, or T4 DNA polymerase.
14. The method of claim 1, wherein the primers are each 4-50 nucleotides in length.
15. The method of claim 1, wherein said mRNA further comprises a polyadenylated tail and said mRNA is contacted with one or more first random oligonucleotide primers comprising an RNA polymerase promoter in combination with an oligo (dT) primer comprising an RNA polymerase promoter under conditions sufficient for reverse transcriptase activity to occur.
16. The method of claim 15, wherein said oligo (dT) primer is stably associated with the surface of a solid support.
17. The method of claim 1, wherein the first and second label is a fluorescent, radioactive, enzymatic, hapten, biotin, or digoxigenin label.
18. The method of claim 17, wherein the label is fluorescent.
19. The method of claim 18 wherein the fluorescent label is fluorescein isothiocyanate, lissamine, Cy3, Cy5, or rhodamine 110.
20. The method of claim 18, wherein a first aliquot of the first cRNA is labeled with a first fluorophore having a first emission spectrum, and a second aliquot of the second cRNA is labeled with a second fluorophore with a second emission spectrum differing from that of the first emission spectrum.
21. The method of claim 20, wherein the first fluorophore is Cy3 and the second fluorophore is Cy5.

22. The method of claim 1 wherein in steps (h) and (i), the steps of contacting the first cRNA labeled with said first label with said polynucleotide probe, and contacting the second cRNA labeled with said second label with said polynucleotide probe, are carried out concurrently.
23. The method of claim 1 wherein said first sample contains mRNAs from cells that are diseased and wherein said second sample contains mRNAs from normal cells.
24. The method of claim 1 wherein said first sample contains mRNAs from drug-treated cells and wherein said second sample contains mRNAs from untreated cells.
25. The method of claim 1 wherein said detecting or measuring steps (h) and (i) are carried out by a method comprising contacting said first and second cRNAs with an array containing one or more species of polynucleotide probe positioned at preselected sites on the array, under conditions sufficient for hybridization to occur; and detecting any hybridization that occurs between said polynucleotide probes and said cRNAs.
26. The method of claim 25 wherein the array comprises a support with at least one surface and at least 100 different polynucleotide probes, each different polynucleotide probe comprising a different nucleotide sequence and being attached to the surface of the support in a different, selected location on said surface.
27. The method of claim 25 wherein the nucleotide sequence of each of the different polynucleotide probes is in the range of 40-80 nucleotides in length.
28. The method of claim 25 wherein the polynucleotide probes comprise oligodeoxyribonucleotides sequences.
29. The method of claim 25 wherein the polynucleotide probes comprise cDNA sequences.
30. The method of claim 25 wherein the array comprises polynucleotide probes of at least 2,000 different sequences.
31. The method of claim 25 wherein the polynucleotide probes hybridize to at least 10,000 different polynucleotide sequences.
32. The method of claim 25 wherein the array has at least 1,000 polynucleotide probes per 1 cm².
33. The method of claim 27 wherein the nucleotide sequence of the polynucleotide probes are specific for at least 50% of the genes in the genome of the cells of interest.
34. A method for synthesizing at least one cRNA from a sample containing a plurality of different mRNAs, said method comprising:
- (a) synthesizing first strand cDNA by contacting at least one mRNA in said sample with (i) one or more random oligonucleotide primers comprising an RNA polymerase promoter that is sufficiently complementary to a sequence in the mRNA so as to prime first strand cDNA synthesis, and (ii) reverse transcriptase under conditions sufficient for reverse transcriptase

activity to occur;

(b) synthesizing double-stranded cDNA by contacting the first strand cDNA with (i) a second random primer not comprising an RNA polymerase promoter wherein said second random primer is sufficiently complementary to a sequence in the first strand cDNA so to prime second strand cDNA synthesis, and (ii) a DNA polymerase under conditions sufficient for RNA polymerase activity to occur;

(c) isolating and purifying said double-stranded cDNA; and

(d) subjecting said purified double-stranded cDNA to *in vitro* transcription under conditions sufficient for template driven DNA transcription to occur, such that cRNA is produced.

35. The method of claim 34 wherein said sample contains total RNA or total mRNA from cells.

36. The method of claim 35 wherein said cells are prokaryotic cells.

37. The method of claim 36 wherein said prokaryotic cells are yeast cells.

38. The method of claim 35 wherein said cells are eukaryotic cells.

39. The method of claim 38 wherein said eukaryotic cells are mammalian cells.

40. The method of claim 39 wherein said mammalian cells are human cells.

41. The method of claim 34 wherein said sample contains total RNA from 1×10^6 cells or less.

42. The method of claim 34 wherein said sample contains at least 10,000 different mRNAs.

43. The method of claim 34, wherein the RNA polymerase promoter is a T7 RNA polymerase promoter sequence and the RNA polymerase is T7 RNA polymerase.

44. The method of claim 34, wherein the RNA polymerase promoter is a T3 RNA polymerase promoter sequence and the RNA polymerase is T3 RNA polymerase.

45. The method of claim 34, wherein the RNA polymerase promoter is an SP6 RNA polymerase promoter sequence and the RNA polymerase is SP6 RNA polymerase.

46. The method of claim 34, wherein said DNA polymerase is E. coli DNA polymerase I, Klenow fragment of E. coli DNA polymerase I, or T4 DNA polymerase.

47. The method of claim 34, wherein the primers are each 4-50 nucleotides in length.

48. The method of claim 34, wherein said mRNA further comprises a polyadenylated tail and said mRNA is contacted with one or more random oligonucleotide primers comprising an RNA polymerase promoter in combination with an oligo (dT) primer comprising an RNA

polymerase promoter under conditions sufficient for reverse transcriptase activity to occur.

49. The method of claim 48, wherein said oligo (dT) primer is stably associated with the surface of a solid support.

50. The method of claim 49, which further comprises labeling the transcribed cRNA with a label.

51. The method of claim 50, wherein the label is a fluorescent, radioactive, enzymatic, hapten, biotin, or digoxigenin label.

52. The method of claim 51, wherein the label is fluorescent.

53. The method of claim 52 wherein the fluorescent label is fluorescein isothiocyanate, lissamine, Cy3, Cy5, or rhodamine 110.

54. The method of claim 34, further comprising, after the *in vitro* transcription step, determining the presence or absence of a target mRNA in said sample.

55. The method of claim 34, further comprising, after the *in vitro* transcription step, determining the effect of drug treatment on a target mRNA in said sample.

56. The method of claim 34, wherein the mRNA is extracted from at least one cell of interest, and further comprising (e) contacting the cRNA produced in step (d) with an array containing one or more species of polynucleotide positioned at preselected sites on the array, under conditions conducive to hybridization; and (f) detecting any hybridization that occurs between said one or more species of polynucleotide and said cRNA.

57. The method of claim 56 wherein the array comprises a support with at least one surface and at least 100 different polynucleotide probes, each different polynucleotide probe comprising a different nucleotide sequence and being attached to the surface of the support in a different, selected location on said surface.

58. The method of claim 57 wherein the nucleotide sequence of each of the different polynucleotide probes is in the range of 40-80 nucleotides in length.

59. The method of claim 57 wherein the polynucleotide probes comprise oligodeoxyribonucleotides sequences.

60. The method of claim 57 wherein the polynucleotide probes comprise cDNA sequences.

61. The method of claim 57 wherein the array comprises polynucleotide probes of at least 2,000 different sequences.

62. The method of claim 57 wherein the polynucleotide probes hybridize to at least 10,000 different polynucleotide sequences.

63. The method of claim 57 wherein the array has at least 1,000 polynucleotide probes per 1

cm².

64. The method of claim 57 wherein the nucleotide sequence of the polynucleotide probes are specific for at least 50% of the genes in the genome of the cells of interest.

65. A method of producing double-stranded deoxyribonucleic acid molecules from a single-stranded ribonucleic acid molecule, said method comprising:

(a) contacting said ribonucleic acid molecule with (i) one or more first random oligonucleotide primers comprising an RNA polymerase promoter, and (ii) reverse transcriptase under conditions sufficient for template driven enzymatic deoxyribonucleic acid synthesis to occur, whereby a first deoxyribonucleic acid molecule is produced;

(b) contacting said first deoxyribonucleic acid molecule (i) with a second random primer which does not comprise an RNA polymerase promoter, and (ii) DNA polymerase under conditions sufficient for template driven enzymatic deoxyribonucleic acid synthesis to occur, whereby a second deoxyribonucleic acid molecule is produced; and

(c) isolating and purifying the resulting double-stranded deoxyribonucleic acid molecule, such that a double-stranded deoxyribonucleic acid molecule is produced.

66. The method of claim 65, wherein said single-stranded ribonucleic acid molecule is an mRNA.

67. The method of claim 65, wherein said deoxyribonucleic acid molecule is cDNA.

68. The method of claim 65, wherein the RNA polymerase promoter is a T7 RNA polymerase promoter sequence and the RNA polymerase is T7 RNA polymerase.

69. The method of claim 65, wherein the RNA polymerase promoter is a T3 RNA polymerase promoter sequence and the RNA polymerase is T3 RNA polymerase.

70. The method of claim 65, wherein the RNA polymerase promoter is an SP6 RNA polymerase promoter sequence and the RNA polymerase is SP6 RNA polymerase.

71. The method of claim 65, wherein the DNA polymerase is E. coli DNA polymerase I, Klenow fragment of E. coli DNA polymerase I, or T4 DNA polymerase.

72. The method of claim 65, wherein the primers are each 4-50 nucleotides in length.

73. The method of claim 65, wherein said method further comprises introducing said double-stranded deoxyribonucleic acid molecule into a vector.

74. The method of claim 65, wherein said single-stranded ribonucleic acid molecule further comprises a polyadenylated tail and said ribonucleic acid molecule is contacted with one or more first random oligonucleotide primers comprising an RNA polymerase promoter in combination with an oligo (dT) primer comprising an RNA polymerase promoter under conditions sufficient for template driven enzymatic deoxyribonucleic acid synthesis to occur.

75. The method of claim 74, wherein said single-stranded ribonucleic acid molecule is an mRNA.

76. The method according to claim 74, wherein said deoxyribonucleic acid molecule is cDNA.

77. The method of claim 74, wherein the RNA polymerase promoter sequence is a T7 RNA polymerase promoter sequence and the RNA polymerase is T7 RNA polymerase.

78. The method of claim 74, wherein the RNA polymerase promoter sequence is a T3 RNA polymerase promoter sequence and the RNA polymerase is T3 RNA polymerase.

79. The method of claim 74, wherein the RNA polymerase promoter sequence is an SP6 RNA polymerase promoter sequence and the RNA polymerase is SP6 RNA polymerase.

80. The method of claim 74, wherein the primers are each 4-50 nucleotides in length.

81. The method of claim 74, wherein said method further comprises introducing said double-stranded deoxyribonucleic acid molecule into a vector.

82. The method of claim 74, wherein said oligo (dT) primer is stably associated with the surface of a solid support.

83. A method for producing a 5' enriched cDNA library from a sample of mRNA molecules, said method comprising:

(a) contacting said mRNA with (i) one or more first random oligonucleotide primers comprising an RNA polymerase promoter, and (ii) reverse transcriptase under conditions sufficient for template driven enzymatic deoxyribonucleic acid synthesis to occur, whereby a first population of cDNA molecules is produced;

(b) contacting said first population of cDNA molecules (i) with a second random primer not comprising an RNA polymerase promoter, and (ii) DNA polymerase under conditions sufficient for template driven enzymatic deoxyribonucleic acid synthesis to occur, whereby a second population of cDNA molecules is produced; and

(c) isolating said second population of double-stranded cDNA molecules; such that said 5' enriched library is produced.

84. A kit comprising in one or more containers:

(a) a mixture of first random oligonucleotide primers, each first primer comprising a 3' end sequence of 10-50 nucleotides, said 3' end sequence comprising an A, a G, a T, or a C nucleotide randomly present in each position of said 3' end sequence; and each first primer comprising a 5' end sequence of 10-50 nucleotides, said 5' end sequence comprising an RNA polymerase promoter;

(b) a mixture of oligo (dT) primers, each oligo (dT) primer comprising a 3' end sequence of 10-50 nucleotides, said 3' end sequence comprising a T nucleotide present in each position of

said 3' end sequence; and each oligo (dT) primer comprising a 5' end sequence of 10-50 nucleotides, said 5' end sequence comprising an RNA polymerase promoter; and

(c) a mixture of second random oligonucleotide primers, each second primer comprising a sequence of 10-50 nucleotides, said sequence comprising an A, a G, a T, or a C nucleotide randomly present in each position of said sequence.