

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
8 November 2001 (08.11.2001)

PCT

(10) International Publication Number
WO 01/83696 A2

(51) International Patent Classification⁷: C12N

(21) International Application Number: PCT/US01/13807

(22) International Filing Date: 27 April 2001 (27.04.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/560,845 28 April 2000 (28.04.2000) US

(71) Applicant: **DIGITAL GENE TECHNOLOGIES, INC.**
[US/US]; Suite 110, 11149 North Torrey Pines Road, La Jolla, CA 92037 (US).

(72) Inventors: **MULLER, Rolf**; 12744 Via Donada, Del Mar, CA 92014 (US). **RIDDLE, Gretchen, H.**; 1450 Harbor Island Drive, San Diego, CA 92101 (US). **GLASS, James, R.**; 1467 Kings Cross Drive, Cardiff, CA 92007 (US).

(74) Agent: **ZIMMERMAN, Roger, P.**; McDonnell Boehnen Hulbert & Berghoff, 32nd floor, 300 South Wacker Drive, Chicago, IL 60606 (US).

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

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

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHODS FOR RAPID ISOLATION AND SEQUENCE DETERMINATION OF GENE-SPECIFIC SEQUENCES

(57) Abstract: The present invention relates to methods for the rapid isolation and sequencing of novel gene-specific sequences. Additionally, the present invention relates to novel oligonucleotide primer sequences and compositions thereof and kits comprising such oligonucleotide primer sequences and compositions.

WO 01/83696 A2

- 1 -

METHODS FOR RAPID ISOLATION AND SEQUENCE DETERMINATION OF
GENE-SPECIFIC SEQUENCES

(MBHB 99,023-A)

5 TECHNICAL FIELD OF INVENTION

The present invention relates to a method for the rapid isolation and sequencing of novel gene-specific sequences. Additionally, the present invention relates to novel oligonucleotide primer sequences and compositions thereof and kits comprising such oligonucleotide primer sequences and compositions.

10

BACKGROUND OF THE INVENTION

Nucleotide sequence fragments of unknown genes can be obtained by various molecular biology techniques including protein sequencing, cDNA library screening, EST sequencing, and cloning of DNA fragments generated using polymerase chain reaction (PCR). On occasion, some information regarding the structure and function of the gene may be inferred from the nucleotide sequence of the gene fragment. However, in order to characterize gene structure and function decisively, it is often necessary to extend the nucleotide sequence beyond that of the fragment or to determine the full-length sequence of the gene. In practice, the challenge is to isolate a single particular sequence extension from within a large mixture of other unwanted sequences.

20

Previous approaches for obtaining the nucleotide sequence of extended or full-length DNA from a gene sequence fragment have involved the construction and screening of cDNA libraries or the use of PCR-based sequence extension methods. However, these commonly used methods are limited due to high background levels of non-specific PCR products. Most previous methods have required the cloning of the obtained DNA products in order to have sufficient quantities to determine the sequences. Such methods requiring screening and cloning are expensive, labor-intensive, and can provide only low sample throughput.

25

The traditional approach of screening of a cDNA library has the advantage of creating long extended sequences, but requires time-consuming library construction in either plasmids (Peacock, S.L., et al., , *Biochim. Biophys. Acta* 66: 243-250 (1981)) or bacteriophage (Young , R.A. and Davis R.W., *Proc. Natl. Acad. Sci. USA* 80:

30

- 2 -

1194-1198 (1983)). The libraries have to be plated on filters (Hanahan, D. and Meselson, M., , *Methods Enzymology*, 100: 333-342 (1983)) and screened by various hybridization or clonal selection methods. The likelihood of non-specific hybridization and false positives are significant problems that require extensive screening and analysis of clones. Furthermore, because limitations inherent to hybridization-based detection, low-abundance mRNAs are hard to detect using traditional DNA library screens if compared to PCR based techniques that allow DNA amplification.

PCR is a powerful technique for amplifying specific nucleic acid sequences (U.S. Pat No. 4,683, 202; U.S. Pat No. 4,683, 195; Erlich, *PCR Technology*, (Stockton Press, 1989)) and is becoming more widely used as a basis for isolating gene sequences and their characterization.

One PCR-based method for extended or full-length sequence generation, commonly known as Rapid Amplification of cDNA Ends (RACE) technology (Frohman et al., *Proc. Natl. Acad. Sci. USA*, 85:8998-9002 (1988) overcomes some of the traditional cDNA library problems like low level of detection and filter hybridization of cDNA libraries. The RACE method is based upon using mRNA as a template for the preparation of gene-specific cDNA, adapting this cDNA with a universal 5' priming site and subsequent PCR amplification with 3' gene specific and 5' universal primers. RACE allows the detection of low abundance mRNAs, but the original protocol is still associated with significant problems. These problems include increased background created by the 5' modification, low amplification efficiency and a requirement for large amounts of mRNA template, a reagent often in short supply. The high background of nonspecific PCR products that is generated by RACE requires labor intensive screening of numerous clones. While the currently available methods are useful in the amplification of specific transcripts from cDNA and bacterial genome templates, several shortcomings remain. Modifications to improve product homogeneity often sacrifice product yield, making direct sequencing and other analyses difficult. Furthermore, the detection of low abundance or rare transcripts is difficult to achieve using these methods.

Recent modifications of the RACE technology have overcome some of those problems. Several modifications of the RACE technique involve the use of specially adapted primers ligated to both the 3' and 5' ends of a double stranded cDNA. The

- 3 -

target gene-specific sequence is amplified using known gene-specific primers and the specially adapted primers (see, e.g. Chenchik et al., *BioTech.*, 21:526-534 (1996)). However, due to primer mismatch and amplification of mismatched sequences, these RACE techniques can result in high background, especially in the case of rare transcripts. Further, in methods in which the initial cDNA library is not amplified, the number of template DNA molecules for rare transcripts remains extremely low. Consequently, especially in the case of rare transcripts, these RACE techniques do not produce a pure population of the target sequence in sufficient concentration to be directly sequenced and, instead, necessitate cloning the target product into a sequencing vector.

One method uses RNA ligase to add a known sequence to the end of the mRNA (Liu, X., and Gorovsky, M.A. (1993) *Nucleic Acids Research*, 21: 4954-60 (1993). Other methods have added known sequences to the 5' end of cDNA libraries by using adapters or adapter like molecules (Chenchik et al., *BioTech.*, 21:526-534 (1996) and Tavtigian et al. (U.S. Pat.No. 5,789,206) and the resulting cDNA population is used as a PCR amplification template. To obtain the 3' end of a cDNA, the 3'-RACE technique uses an adapter primer containing a unique sequence and polydT residues to reverse transcribe an mRNA population, producing a first strand cDNA. A gene-specific primer is annealed with the first strand cDNA and extended to generate a complementary second strand. The target cDNA is amplified using a gene-specific primer and the described adapter primer. To obtain the 5' end of a cDNA, a gene-specific primer is used to reverse transcribe the mRNA, after which a poly(dA) tail (or other polydeoxynucleotide tail) is added to the primer extended product. The second strand synthesis is carried out by annealing the adapter poly(dT) primer with the poly(dA) tail. Similarly, the target cDNA is amplified using the gene-specific primer and the adapter primer.

The methods described above still require labor intensive cloning of the PCR generated fragments and screening methods to isolate the extended sequences. Other PCR based techniques make use of biotinylated gene specific primers and streptavidin based capture to improve and enrich for gene-specific cDNA fragments (Lankiewicz et al. *Nuc. Acids Res.*, 25:2037-2038 (1997) and thus decrease background. It has to be noted that using a single step capture approach is not sufficient to generate a pure sequencing template. This is particularly true for the sequence extension of rare genes.

- 4 -

The method described by Lankiewics et al. also addresses the problem of using large amounts of mRNA as templates for RACE. They create a gene specific cDNA and add known sequences to the cDNA ends, but still required homopolymer tailing at the 5' end. Again, this protocol requires time consuming cloning for the determination of the extended cDNA sequences. Others have also used biotinylated gene specific primers (Gräslund et al, *BioTech.*, 27:488-498 (1999) for the separation of the gene of interest from the non-specific background. However, while this method may complement existing RACE strategies and reduces background efficiently, it has several drawbacks, including the inability to generate the absolute 5' end of the gene sequence. This is due to the use of 5' non-specific internal priming sites during the amplification process.

Despite such recent advances in PCR-based methods, the rapid and efficient isolation of a partially extended or full-length cDNA clone in sufficient homogeneity and concentration for sequencing remains a considerable challenge. A method for the isolation of rare transcripts remains an even greater challenge. Accordingly, there exists a need for improved methods of isolating and sequencing extended gene-specific sequences and, in particular, low abundance transcripts. Specifically, there remains a need for a method which provides an efficient means of producing a purified gene-specific sequence from a collection of nucleic acid sequences sufficiently homogenous and in sufficient quantity to be directly sequenced. The need of cloning and clonal selection using screening methods like PCR or hybridization prior to sequence determination does not allow the method to be introduced in a robotics high-throughput environment needed in our genomics age. The method described herein allows direct sequencing and identification of gene specific sequences and provide a description for a fully automatable process for obtaining extended and full-length gene sequences.

BRIEF SUMMARY OF THE INVENTION

The present invention provides novel methods and composition for isolating and sequencing gene-specific polynucleotides. The present invention provides several advantages over the available methods. The elimination of cloning and screening clones provides quicker sequence results. The capture of the gene-specific first strand

- 5 -

significantly reduces the background of non-specific PCR products. Amplification of the entire adapted cDNA template increases the ability to detect and sequence rare transcripts. The method of the present invention is adaptable to multiwell formats, providing a high-throughput system for generation of extended and full-length sequences. The method of the present invention is adaptable to automation using techniques of robotics, fluid handling and numeric control known to the skilled artisan, further increasing the throughput of a system for generation of extended and full-length sequences.

The present invention provides a novel method for isolating a gene-specific polynucleotide, which method comprises the steps of: (a) synthesizing a population of double stranded DNA molecules, wherein each strand has a 5' end and a 3' end, using an anchor primer having a first captureable moiety, wherein the anchor primer is at the 3' end of each double stranded DNA molecule; (b) ligating a double stranded adapter rare sequence molecule to the 5' end of each double stranded DNA molecule; (c) synthesizing a single stranded gene-specific polynucleotide using a gene-specific primer having a second captureable moiety, wherein the captureable moiety of the gene-specific primer is different from the captureable moiety of the anchor primer; (d) purifying the single stranded gene-specific polynucleotide; and (e) amplifying the purified gene-specific polynucleotide using both a gene-specific primer and a primer that hybridizes to either a sequence located in the anchor primer or a sequence located in the adapter molecule.

Typically, the method comprises the additional step of sequencing the amplified isolated gene-specific polynucleotide.

In general, the anchor primer comprises the sequence 5'- N'(N)_a(T)_b-3' where N' is a nucleotide to which a captureable moiety is attached, a is an integer of from about 15 to about 50, and b is an integer of from about 12 to about 25. In one embodiment, N' is a biotinylated nucleotide. In another embodiment, N' is an aminated nucleotide.

In one embodiment, the anchor primer comprises the sequence 5'- NAATTCAACTGGAAGCGGCCGCAGGAA(T)₁₈-3' where N is a biotinylated guanidylate nucleotide (SEQ ID NO: 1). In another embodiment, the anchor primer comprises the sequence 5'- NAATTCAACTGGAAGCGGCCGCAGGAA(T)₁₈-3' where N is an aminated guanidylate nucleotide (SEQ ID NO: 2).

- 6 -

Typically, the double-stranded adapter molecule comprises the sequence of at least one rare restriction enzyme site wherein the rare restriction enzyme site occurs in a mammalian genome less than of about one site per 15,000 nucleotides. Suitably, the double-stranded adapter molecule comprises the sequence of two or more rare
5 restriction enzyme sites which are sequentially arranged, selected from the group consisting of NotI, AscI, BaeI, FseI, SfiI, SgfI, SgrAI, SrfI, PacI, PmeI, PpuMI, RsrII, SapI, SexAI, Sse8387I, SbfI, SpeI, SalI, RsrI, and NheI restriction enzyme sites. In one embodiment, the double-stranded adapter molecule comprises the sequence of contiguous SbfI, SpeI, SalI, RsrI, and NheI restriction enzyme sites.

10 Typically, the double-stranded adapter molecule comprises an RNA polymerase promoter sequence selected from the group consisting of T3, SP6, and T7 RNA polymerase promoter sequences. In one embodiment, the double-stranded adapter molecule comprises the T7 polymerase promoter sequence. In one
15 embodiment, the double-stranded adapter molecule comprises the sequence of 5'-AT-AGCCTGCAGGTAATACGACTCACTATAGGGACTAGTCGACGGACCGCTAGCATCAGATC - 3' (SEQ ID NO: 13). In one embodiment, the double stranded adapter molecule comprises SEQ ID NO: 15.

Typically, method comprises the step of amplifying the double stranded DNA molecules comprising both an anchor primer and an adapter molecule before the step
20 of synthesizing the single stranded gene-specific polynucleotide, wherein the double stranded DNA molecules are amplified using a 3' end primer that hybridizes to a sequence located in the anchor primer and a 5' end primer that hybridizes to a sequence located in the adapter molecule. In one embodiment, the 3' end amplification primer comprises the sequence 5'- NAATTCAACTGGAAGCGGCCGAGGA -3'
25 (SEQ ID NO: 20 or 21) and the 5' end amplification primer comprises the sequence of 5' - CCTGCAGGTAATACGACTCACTATAGG - 3' (SEQ ID NO: 17). In one embodiment, the 5' end amplification primer comprises the sequence of SEQ ID NO: 18.

In one embodiment, the method comprises the additional steps of:

30 synthesizing a population of single-stranded RNA molecules from the population of double stranded DNA molecules comprising an anchor primer and an adapter molecule, and

- 7 -

digesting the double stranded DNA template molecules, before the step of synthesizing the single stranded gene-specific polynucleotide.

In one embodiment, the gene specific primer comprises the sequence 5'- N'(N)_c -3' where N' is a nucleotide to which a captureable moiety is attached and c is an integer of from about 17 to about 40. In one embodiment, N' is a biotinylated nucleotide. In another embodiment, N' is an aminated nucleotide. Typically, the first captureable moiety of the anchor primer of the double stranded DNA molecule is affixed to a substrate comprising a coating of streptavidin. Alternatively, the first captureable moiety of the anchor primer of the double stranded DNA molecule is affixed to a substrate comprising a coating of N-oxysuccinimide ester.

In one embodiment, the second captureable moiety of the gene-specific primer of the gene-specific single stranded polynucleotide is affixed to a substrate comprising a coating of streptavidin. In another embodiment, the second captureable moiety of the gene-specific primer of the gene-specific single stranded polynucleotide is affixed to a substrate comprising a coating of N-oxysuccinimide ester.

Typically, the purified single stranded gene-specific polynucleotide is amplified using a 5' end gene-specific primer and a 3'end amplification primer, the 3'end amplification primer hybridizing to a portion of the sequence of the anchor primer.

In one embodiment, the 3' end amplification primer comprises SEQ ID NO: 23. In another embodiment, the 5' end amplification primer comprises SEQ ID NO: 17. In one embodiment, the 5' end amplification primer comprises SEQ ID NO: 18.

In another embodiment, the present invention provides a method for isolating a gene-specific polynucleotide comprising the steps of:

synthesizing a population of double stranded DNA molecules having a 3' end and a 5' end;

ligating a double stranded adapter molecule to the 5' end of each double stranded DNA molecule;

attaching at least one gene-specific oligonucleotide primer having a captureable moiety to a solid substrate, wherein the gene-specific primer is attached via the captureable moiety;

- 8 -

synthesizing at least one single stranded gene-specific polynucleotide using the attached gene-specific primer(s);

purifying the single-stranded gene-specific polynucleotide(s); and
amplifying the gene-specific polynucleotide(s) using a gene-specific primer and an
5 primer that hybridizes to a sequence located in the adapter molecule.

In one embodiment, the second captureable moiety of the gene-specific primer of the gene-specific single stranded polynucleotide is affixed to a substrate comprising a coating of streptavidin. In another embodiment, the second captureable moiety of the gene-specific primer of the gene-specific single stranded polynucleotide is affixed
10 to a substrate comprising a coating of N-oxysuccinimide ester. In one embodiment, the purified single stranded gene-specific polynucleotide is amplified using a 3' end gene-specific primer and a 5' end amplification primer that hybridizes to a sequence located in the adapter molecule. Typically, the 5' end amplification primer comprises SEQ ID NO: 17. Alternatively, the 5' end amplification primer comprises SEQ ID
15 NO: 18.

The invention also provides an isolated polynucleotide having an attached captureable moiety chosen from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:45 and SEQ ID NO:46. In another embodiment, the invention provides an isolated double stranded polynucleotide chosen from the group consisting
20 of SEQ ID NO: 13 and complementary sequence SEQ ID NO: 14, and SEQ ID NO: 15 and complementary sequence SEQ ID NO: 16. In another embodiment, the invention provides an isolated polynucleotide comprising the sequence chosen from the group consisting of SEQ ID NO: 17-19.

The invention also provides an isolated polynucleotide comprising an anchor
25 amplification primer oligonucleotide having an attached biotin moiety. Suitably, the anchor primer is an isolated polynucleotide selected from the group consisting of:

NAATTCAACTGGAAGCGGCCGAGGAATTTTTTTTTTTTTTTTTT
(SEQ ID NO: 1&2);

NAATTCAACTGGAACCTGCAGGAGGAATTTTTTTTTTTTTTTTTT
30 (SEQ ID NO: 3&4);

NAATTCAACTGGAACTAGTAGGAATTTTTTTTTTTTTTTTTT (SEQ
ID NO: 5&6);

- 9 -

NAATTCAACTGGAAGTCGACAGGAATTTTTTTTTTTTTTTTTT (SEQ ID NO: 7&8);

NAATTCAACTGGAACGGATCCGAGGAATTTTTTTTTTTTTTTTTT (SEQ ID NO: 9&10); and

5 NAATTCAACTGGAAGCTAGCAGGAATTTTTTTTTTTTTTTTTT (SEQ ID NO: 11&12),

where N is chosen from A, G, C, and T, and the N is covalently bound to a captureable moiety chosen from the group consisting of biotin, primary amines, avidin, streptavidin, neutravidin, primary carboxylates, thiol alcohols, thiol carboxylates, carbonyls, sugars, lipids, and peptides. In one embodiment, the anchor amplification primer oligonucleotide has an attached primary amine comprising the sequence consisting of SEQ ID NO: 21. In another embodiment, the anchor amplification primer oligonucleotide comprises the sequence consisting of SEQ ID NO: 23.

15 In another embodiment, the invention provides a composition comprising the oligonucleotide and an aqueous carrier. In another embodiment, the invention provides an article of manufacture comprising the oligonucleotide and a streptavidin-coated substrate. In another embodiment, the invention provides a kit suitable for isolating and sequencing a gene-specific polynucleotide from a population of DNA or RNA polynucleotides comprising at least one single stranded anchor primer oligonucleotide having an attached captureable moiety and at least one double stranded adapter oligonucleotide in an amount sufficient for at least one assay and suitable containers.

Typically, the kit further comprises at least one single stranded amplification anchor primer oligonucleotide and at least one single stranded amplification adapter primer oligonucleotide. Typically, the kit further comprises at least one single stranded gene-specific primer oligonucleotide having a captureable moiety, wherein the captureable moiety of the anchor primer oligonucleotide is different from the captureable moiety of the gene-specific primer oligonucleotide.

30 The kit typically comprises the gene-specific primer oligonucleotide is attached to a substrate comprising a coating. In one embodiment, the substrate is a multiwell plate.

- 10 -

In another embodiment, the present invention provides a method for determining without cloning an extended sequence of a gene-specific polynucleotide that has a partial gene-specific sequence known, comprising the steps of:

- 5 synthesizing a population of double stranded DNA molecules having a 3' end and a 5' end using an anchor primer having a first captureable moiety, wherein the anchor primer is at the 3' end of each double stranded DNA molecule;
- ligating a double stranded adapter molecule to the 5' end of each double stranded DNA molecule producing double stranded DNA molecules comprising both an anchor primer and an adapter molecule;
- 10 synthesizing a single stranded gene-specific polynucleotide using a gene-specific oligonucleotide primer having a second captureable moiety, wherein the captureable moiety of the gene-specific oligonucleotide primer is different from the captureable moiety of the anchor primer;
- isolating the single stranded gene-specific polynucleotide; and
- 15 amplifying the isolated gene-specific polynucleotide using both a gene-specific primer that hybridizes to either a sequence located in the anchor primer or a sequence located in the adapter molecule; and
- sequencing the amplified isolated gene-specific polynucleotide without cloning, thereby determining the sequence of a polynucleotide.

20

- In another embodiment, the present invention provides a method for the simultaneous isolation and sequence determination of multiple novel genes, comprising the steps of: (a) synthesizing a population of double stranded DNA molecules, wherein each strand has a 5' end and a 3' end; (b) ligating a double
- 25 stranded adapter molecule to the 5' end of each double stranded DNA molecule; (c) attaching at least one gene-specific oligonucleotide primer having a captureable moiety to a solid substrate, wherein the gene-specific primer is attached via the captureable moiety; (d) synthesizing at least one single stranded gene-specific polynucleotide using the attached gene-specific primer(s) and a double stranded DNA
 - 30 having a 5' end adapter as a template; (e) purifying the single stranded gene-specific polynucleotide(s); and (f) amplifying the gene-specific polynucleotide(s) using both a gene-specific primer and an primer that hybridizes to a sequence located in the adapter molecule.

- 11 -

Additionally, the present invention provides novel oligonucleotide primer sequences which can be used to practice preferred embodiments of the present invention. Also provided are compositions which comprise a novel oligonucleotide primer and an aqueous carrier. The present invention also provides kits for the rapid
5 isolation and sequencing of novel genes from DNA and RNA libraries which comprise novel oligonucleotide primers.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the production of a 3' end capture modified cDNA (cDNA_M)
10 using a captureable anchor primer.

Figure 2 shows the production of a specially adapted cDNA (acDNA_M) by attachment of a double stranded adapter molecule to the 5' end of the cDNA_M.

Figure 3 shows the production of a gene-specific sequence oriented in either the mRNA sense direction or the mRNA antisense direction.

15 Figure 4 summarizes anchor amplification primers and adapter amplification primers used.

Figure 5 shows the purification of a gene-specific sequence from a biotinylated cDNA library using aminated capture.

20 Figure 6 shows the purification of a gene-specific sequence from an aminated cDNA library using biotinylated capture.

Figure 7 shows the purification of a gene-specific sequence from a synthetic RNA library.

25 Figure 8 is a schematic diagram showing the adapter and gene-specific primers as well as the sequencing strategy used to determine the sequence of the cyclophilin gene.

Figure 9 is a schematic diagram showing the adapter and gene-specific primers as well as the sequencing strategy and primer walking used to determine the sequence of a novel M41 gene.

30 Figure 10 is a schematic diagram showing the adapter and gene-specific primers as well as the sequencing strategy and primer walking used to determine the sequence of a novel M51 gene.

- 12 -

DETAILED DESCRIPTION OF THE INVENTION

The orientation of a nucleic acid is described in terms of reference to the mRNA ("sense" RNA). Thus, reference to the 5' end and 3' end of a double stranded DNA is meant to describe those ends corresponding to the 5' end and 3' end, respectively, of the mRNA. Using this terminology, a primer that hybridizes to the 3' end of a double stranded DNA is oriented in the 5' or antisense direction. A primer that hybridizes to the 5' end of a double stranded DNA is directed in the 3' or sense direction.

As used herein, unless otherwise specified, the term "polynucleotide" is defined to encompass DNA and RNA of both synthetic and natural origin. The polynucleotide may exist as single or double stranded DNA or RNA, or an RNA/DNA heteroduplex. Thus, the polynucleotide of the present invention can be composed of any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. A polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms.

As used herein, the term "oligonucleotide" refers to a single stranded DNA or RNA molecule. An "oligonucleotide primer" refers to an oligonucleotide that is hybridized to a nucleic acid template for sequencing purposes or to prime enzymatic synthesis of a second nucleic acid strand. Thus, an "oligonucleotide" refers to a molecule having a nucleic acid sequence contained in SEQ ID NO: 1-50. An "oligonucleotide" of the present invention also includes those oligonucleotides capable of hybridizing, under stringent hybridization conditions, to the complementary sequences contained in SEQ ID NO: 1-50. "Stringent hybridization conditions" refers to an overnight incubation at 42° C in a solution comprising 50%

- 13 -

formamide, 5x SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing in 0.1x SSC at about 65°C.

5 An "oligonucleotide" of the present invention also includes those oligonucleotides that are substantially identical to the oligonucleotide sequences of SEQ ID NO: 1-50. As that term is used herein, "substantially identical" refers to a nucleotide sequence that is at least about 80% identical to the reference oligonucleotide. More preferably, a "substantially identical" sequence is at least about
10 90% identical, and most preferably at least about 95-99% identical to a sequence contained in SEQ ID NO:1-50.

"Identity" per se has an art-recognized meaning and can be calculated using published techniques. (See, e.g. Lesk, A.M., Ed., *Computational Molecular Biology*, (Oxford University Press, New York, (1988)); Smith, D.W., Ed., *Biocomputing: Informatics and Genome Projects*, (Academic Press, New York, (1993)); Griffin et al., Eds., *Computer Analysis of Sequence Data, Part I*, (Humana Press, New Jersey, (1994)); von Heinje, G., *Sequence Analysis in Molecular Biology*, (Academic Press, (1987)); and Gribskov et al., Eds., *Sequence Analysis Primer*, (M Stockton Press, New York, (1991)). While there exists a number of methods to measure identity between
20 two polynucleotide or polypeptide sequences, the term "identity" is well-known to skilled artisans (Carillo, et al., *SIAM J Applied Math* 48:1073 (1988)). Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in Bishop, M.J., Ed., *Guide to Huge Computers*, (Academic Press, San Diego, (1994)), and Carillo (1988), *supra*.
25 Methods for aligning polynucleotides or polypeptides are codified in computer programs, including the GCG program package (Devereux, et al., *Nuc. Acids Res.* 12(1):387 (1984)), BLASTP, BLASTN, FASTA (Atschul et al., *J. Molec. Biol.* 215:403 (1990), and Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, Madison, WI) using
30 the local homology algorithm of Smith and Waterman, *Advances in Applied Mathematics*, 2:482-489 (1981).

When using any of the sequence alignment programs to determine whether a particular sequence is, for instance, 95% identical to a reference sequence, the parameters are set so that the percentage of identity is calculated over the full length

- 14 -

of the reference polynucleotide and that gaps in identity of up to 5% of the total number of nucleotides in the reference polynucleotide are allowed.

A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (*Comp. App. Biosci.*, 6:237-245 (1990)). The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB search of a DNA sequence to calculate percent identity are: Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization Group Length=0, and Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, and Window Size=500 or query sequence length in nucleotide bases, whichever is shorter. Preferred parameters employed to calculate percent identity and similarity of an amino acid alignment are: Matrix=PAM 150, k-tuple=2, Mismatch Penalty= 1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty=0.05, and Window Size=500 or query sequence length in amino acid residues, whichever is shorter. The term "double stranded" refers to a pair of nucleotide molecules that exist in a hydrogen-bonded arrangement typically associated with DNA. Included are paired polynucleotides and oligonucleotides that are essentially double stranded, meaning that they may contain short regions of mismatch, such as from one to about three nucleotides, resulting from design or error.

As used herein, the term "substantially complementary" refers to the hybridization or base pairing between nucleotides, such as, for example, between the two strands of a double stranded DNA molecule or between an oligonucleotide primer and a primer binding site on a single stranded nucleotide to be amplified or sequenced. Two single stranded nucleotide molecules are said to be substantially complementary when the nucleotides of one strand, optimally aligned with appropriate nucleotide insertions, deletions or substitutions, pair with at least about 80% of the nucleotides of the other strand.

The "captureable anchor primer" as that term is used herein refers to the modified polydT primer used to generate the second strand nucleic acid from the template strand nucleic acid. The captureable anchor primer comprises a single

- 15 -

stranded oligonucleotide sequence wherein a polydT sequence is found at the 3' end and a captureable moiety is found at the 5' end of the oligonucleotide.

The "captureable moiety" as that term is used refers to any functional group or molecule that can be joined to a nucleotide and can be attached to a substrate or solid-phase support. The captureable moiety can be directly or indirectly attached to a nucleotide and can also be directly or indirectly attached to a substrate or solid-phase support.

The term "spacer" is intended to encompass any suitable means that can be used to link a nucleotide of the anchor primer to the captureable moiety or any suitable means by which to link the captureable moiety to the solid-phase support. The spacer should not adversely affect the function of the anchor primer or the captureable moiety.

As used herein, the term "substrate" or "solid-phase support" is defined as any material having a rigid or semi-rigid surface. The substrate or solid support should have a surface chemistry that allows the substrate to form a direct attachment with a captureable moiety. Alternatively, the substrate or solid support should have a surface that can be coated or derivatized such that, upon coating or derivatization, the substrate can form an attachment with a captureable moiety.

The "adapter molecule" as that term is used herein refers to the double stranded polynucleotide which is ligated to the 5' end of each cDNA_M having a 3' end modified with a captureable anchor primer, thereby generating an adapted cDNA_M (acDNA_M).

As used herein, the term "rare restriction enzyme site" refers to a restriction enzyme site that has at least six bases and occurs in a mammalian genome with a maximum frequency of about one site per 10,000 nucleotides.

As used herein, the term "RNA polymerase promoter site" refers to the consensus sequence of the promoter site to which an RNA polymerase binds to initiate RNA synthesis.

The "anchor amplification primer" (An) refers to the 3' end oligonucleotide of an oligonucleotide primer pair used to amplify the acDNA_M library. The An primer

- 16 -

hybridizes with a sense anchor primer sequence for the synthesis of a product extended in the 5' or antisense orientation.

The "adapter amplification primer" (Ap) refers to the 5' end oligonucleotide of an oligonucleotide primer pair used to amplify the acDNA_M library. The Ap primer
5 hybridizes with an antisense adapter molecule sequence for the synthesis of a product extended in the 3' or sense orientation.

The "captureable gene-specific primer" as that term is used herein refers to the oligonucleotide primer used to generate the first strand nucleic acid of a gene-specific sequence. The captureable gene-specific primer comprises a single stranded
10 oligonucleotide sequence wherein a captureable moiety is attached to the 5' end of the oligonucleotide.

The "gene-specific amplification primer" (Gs) refers to the oligonucleotide primer used to amplify isolated gene-specific cDNA. The Gs primer is selected to be fully or substantially complementary to a known gene-specific sequence on the sense
15 or antisense strand of a purified gene-specific cDNA. The Gs primer may comprise the same nucleotide sequence as the gene-specific primer used to generate the first strand of a gene-specific cDNA.

As used herein, the term "nested oligonucleotide primer" refers to an oligonucleotide primer that hybridizes with an internal sequence in the first-reaction
20 PCR product. Thus, if the first-reaction PCR product is a acDNA_M molecule, a nested oligonucleotide primer would hybridize with an internal site located in the anchor primer or the adapter molecule.

In accordance with the present invention, novel methods and compositions are provided for the rapid isolation and sequence determination of extended or full-length
25 sequences of novel genes. The present invention is based on the generation of a non-cloned, specially adapted cDNA library (acDNA_M) comprising double stranded DNA molecules that have a first captureable moiety at their 3' end and a known adapter sequence at their 5' end. A modified gene-specific primer having a second captureable moiety is used to generate a gene-specific polynucleotide. Both the
30 cDNA library and the gene-specific primer are uniquely modified such that the cDNA library can be separated from the gene-specific sequence, allowing the differential capture of the gene-specific sequence. Accordingly, in the utilization of PCR

- 17 -

methodology, the present invention eliminates the time consuming cloning and library screening steps associated with conventional methods of library construction and screening. Additionally, the present invention provides an improvement over current PCR RACE methods by producing a purified gene-specific product in sufficient quantity and purity to allow direct sequencing of the gene-specific PCR product.

In view of the above, the present invention provides a novel method for isolating a gene-specific polynucleotide. In a typical embodiment, the method comprises: (a) synthesizing a population of double stranded DNA molecules, wherein each strand has a 5' end and a 3' end, using an anchor primer having a first captureable moiety, wherein the anchor primer is at the 3' end of each double stranded DNA molecule; (b) ligating a double stranded adapter molecule to the 5' end of each double stranded DNA molecule; (c) synthesizing a single stranded gene-specific polynucleotide using a gene-specific primer having a second captureable moiety, wherein the captureable moiety of the gene-specific primer is different from the captureable moiety of the anchor primer; (d) purifying the single stranded gene-specific polynucleotide; and (e) amplifying the purified gene-specific polynucleotide using both a gene-specific primer and a primer that hybridizes to either a sequence located in the anchor primer or a sequence located in the adapter molecule. In a preferred embodiment, the amplified PCR product is then directly sequenced without cloning of the product.

I. Production of a specially adapted cDNA library ($acDNA_M$)

The first step of the present invention involves the generation of a specially adapted double stranded cDNA library ($acDNA_M$). The $acDNA_M$ library is a population of double stranded cDNA molecules wherein each cDNA molecule has a captureable anchor primer at the 3' end and a specially adapted DNA molecule at the 5' end. A double stranded cDNA library may be produced from a population of mRNA molecules using methods well-known in the art. The population of mRNA molecules used to generate the cDNA library may be isolated from any tissue or cell population, including cells grown in culture. Preferably, the mRNA is isolated from tissue or cells actively transcribing a potential gene of interest. Methods of extraction of RNA are well-known in the art and are described, for example, in Sambrook et al., *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Press, Cold Spring

- 18 -

Harbor, New York, 1989), vol. 1, ch.7, "Extraction, Purification, and Analysis of Messenger RNA from Eukaryotic Cells", incorporated herein by this reference. Other isolation and extraction methods are also well-known.

Typically, RNA isolation is performed in the presence of chaotropic agents
5 such as guanidinium chloride or guanidinium thiocyanate, although other detergents and extraction agents can alternatively be used. The mRNA (i.e. poly(A)⁺ RNA) is typically isolated from the total extracted RNA by chromatography over oligo(dT)-cellulose or other chromatographic media that have the capacity to bind the polyadenylated 3'-end portion of mRNA molecules. Preferably, the cDNA library is
10 prepared using poly(A)⁺ selected mRNA. Also preferably, the poly(A)⁺ selected mRNA is representative for all the expressed genes in the prepared sample. Typically, about 2 µg of mRNA is used to prepare the cDNA library. Alternatively, but less preferably, total RNA may be used to synthesize cDNA.

Double stranded cDNAs are prepared from a population of mRNA molecules
15 using a polythymidylate (polydT) primer and reverse transcriptase to generate the first strand cDNA, thereby creating a mRNA:cDNA hybrid. The polydT primer hybridizes to the polyadenylate (polydA) sequence present on the 3' terminus of each mRNA molecule. The first strand of cDNA is synthesized through the process of reverse transcription whereby the enzyme reverse transcriptase adds deoxyribonucleotides to
20 the 3' terminus of the polydT primer (Varmus, Science, 240, 1427-1435 (1988); Sambrook et al., *Molecular Cloning: A Laboratory Manual*, vol. 2, "Construction and Analysis of cDNA Libraries"). The second strand cDNA is generated using an RNase to break the RNA:cDNA hybrid and a DNA polymerase to synthesize a complementary DNA strand from the template DNA strand.

25 The double stranded acDNA_M library of the present invention is generated with a modified polydT primer, called a captureable anchor primer, using the well-known methods of cDNA library construction described above. The captureable anchor primer comprises the single stranded sequence 5'- N(N)_a(T)_b-3', where N is a nucleotide to which a captureable moiety is attached, (N)_a is a polynucleotide where a
30 is an integer of from about 15 to about 50, and (T)_b is a polythymidylate where b is an integer of from about 12 to about 25. N and N may be any nucleotide independently selected from the group consisting of adenylate (A), guanidylate (G), cytidylate (C), uridylate (U) and thymidylate (T) nucleotides. N and N may also include derivitized

- 19 -

or modified nucleotides. Preferably, N is the terminal nucleotide located at the 5' end, however, N may be an internal nucleotide as long as the attached captureable moiety does not interfere with anchor primer hybridization with the mRNA.

5 The captureable moiety can be any functional group or molecule that, first, can be joined to a nucleotide N and, second, can be attached to a substrate or solid-phase support, such that N remains affixed to the solid-phase support by the attachment of the captureable moiety. While the captureable moiety may form any type of attachment with the substrate or solid-phase support, the attachment must be of sufficient strength to maintain the nucleotide N on the solid-phase support.

10 Preferably, the captureable moiety is attached to the substrate or solid-phase support by affinity attachment or by covalent bonding. In accordance with the present invention, suitable captureable moieties include, for example, avidin, streptavidin, neutravidin, biotin, primary amines, primary carboxylates, thiol alcohols, thiol carboxylates, carbonyls, sugars, lipids, and peptides.

15 In one preferred embodiment, the captureable moiety is a biotin molecule. Biotin is a vitamin which allows affinity attachment with the avidin family of proteins. The biotin-avidin attachment is the strongest non-covalent attachment known, having an association constant of 10^{15} M^{-1} . In another equally preferred embodiment, the captureable moiety is a primary amine. A primary amine allows
20 covalent attachment, for example, to an N-oxysuccinimide ester (NOS) by displacing the N-oxysuccinimide group, resulting in the formation of a very specific covalent bond.

In addition, amines can be attached to several other molecules useful for capture onto solid-phase supports, such as those described in U.S. Patent No.
25 5,677,276, which is incorporated herein by reference. Other molecules that react with amines are known to the skilled artisan and include, for example, imidoesters and N-hydroxysuccinimidyl esters (NHS-esters). Imidoesters react with primary amines, resulting in an amidine bond (Staros et al., J. Biol. Chem., 256:5890-5893 (1981); Browne et al., Biochem. Biophys. Res. Comm., 67:126-132 (1975)). Examples of
30 suitable imidoesters include dimethyladipimide-2-HCl (Pierce #20664), dimethylpimelimide-HCl (Pierce #20666), dimethylsuberimide-2HCl (Pierce #20668), dimethyl 3,3'-dithiobispropionimide-2HCl (Pierce #20665).

- 20 -

NHS-esters react with primary and secondary amines, resulting in a covalent amide bond. Examples of suitable NHS-esters include disuccinimidyl glutarate (Pierce #29592), disuccinimidyl suberate (Pierce #21555), bis(sulfosuccinimidyl)suberate (Pierce #21579), dithiobis(succinimidyl propionate) (Pierce #22585), dithiobis(sulfosuccinimidyl propionate) (Pierce #21577), ethylene glycobis(succinimidylsuccinate) (Pierce #21565), ethylene glycobis(sulfosuccinimidylsuccinate) (Pierce #21566), disuccinimidyl tartarate (Pierce #20590), disulfosuccinimidyl tartarate (Pierce #20591), bis[2-(succinimidylloxycarbonyloxy)ethyl]sulfone (Pierce # 21554), bis[2-(sulfosuccinimidylloxycarbonyloxy)ethyl]sulfone (Pierce # 21556), and N-hydroxysuccinimidyl 2,3-dibromopropionate (Pierce #22340).

Other NHS-esters include NHS-ester maleimide compounds. Suitable NHS-ester maleimide compounds include, for example, succinimidyl 4-(N-maleimidomethyl) cyclohexane-1-carboxylate (Pierce #22320), sulfosuccinimidyl 4-(N-maleimidomethyl) cyclohexane-1-carboxylate (Pierce #22322), m-maleimidobenzoyl-N-hydroxysuccinimide ester (Pierce #22310), m-maleimidobenzoyl-N-hydroxysulfosuccinimide ester (Pierce #22312), succinimidyl 4-(p-maleimidophenyl)-butyrate (Pierce #22315), sulfosuccinimidyl 4-(p-maleimidophenyl)-butyrate (Pierce #22317), bismaleimidoheptane (Pierce #22319), N-(γ -maleimidobutyryloxy)succinimide ester (Pierce #22314), N-(γ -maleimidobutyryloxy)sulfosuccinimide ester (Pierce #22324).

Still other compounds that are reactive with primary amines are known in the art and include NHS-ester haloacetyls, NHS-ester pyridyl disulfides, carbodiimides, arylhalides and arylazides. Suitable NHS-ester haloacetyls include N-succinimidyl(4-iodoacetyl)aminobenzoate (Pierce #22325, 22326) and sulfosuccinimidyl(4-iodoacetyl)aminobenzoate (Pierce #22327, 22328). Suitable NHS-ester pyridyl disulfides include 1,4-Di-[3'-2'-pyridyldithio(propionamido)butane] (Pierce #21701), 4-succinimidylloxycarbonyl- α -(2-pyridyldithio)toluene (Pierce #21558, 21458), sulfosuccinimidyl-6-[α -methyl- α -(2-pyridyldithio)-toluamido]hexane (Pierce #21568, 21569), N-succinimidyl-3-(2-pyridyldithio)-propionate (Pierce #21557, 21657, 21757), succinimidyl 6-[3-(2-pyridyldithio)propionamido]hexanoate (Pierce #21651, 21652), and sulfosuccinimidyl 6-[3-(2-pyridyldithio)propionamido]hexanoate (Pierce

- 21 -

#21651, 21652). Carbodiimides react with primary amines to form amide or hydrazone bonds. Examples of suitable carbodiimides are 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (Pierce #22980, 22981) and N,N'-dicyclohexylcarbodiimide (Pierce #20320). One example of a suitable arylhalide is
5 1,5-difluoro-2,4-dinitrobenzene (Pierce #21524).

As stated, several arylazides are reactive with amines, including N-5-azido-2-nitrobenzoyloxysuccinimide (Pierce #21551), N-hydroxysuccinimidyl-4-azidobenzoate (Pierce #21560), N-hydroxysulfosuccinimidyl-4-azidobenzoate (Pierce #21561), N-hydroxysuccinimidyl-4-azidosalicylic acid (Pierce #27715), N-
10 hydroxysulfosuccinimidyl-4-azidosalicylic acid (Pierce #27725), sulfosuccinimidyl-(4-azidosalicylamido)-hexanoate (Pierce #27735), p-nitrophenyl-2-diazo-3,3,3-trifluoropropionate (Pierce #20669), 2-diazo-3,3,3-trifluoropropionylchloride (Pierce #20669), N-succinimidyl-(4-azidophenyl)1,3'-dithiopropionate (Pierce #21552), sulfosuccinimidyl-(4-azidophenyldithio)propionate (Pierce #21553),
15 sulfosuccinimidyl 2-(7-azido-4-methylcoumarin-3-acetamide) ethyl-1,3'-dithiopropionate (Pierce #33030), sulfosuccinimidyl 7-azido-4-methylcoumarin-3-acetate (Pierce #33025), sulfosuccinimidyl 2(m-azido-o-nitrobenzamido)-ethyl-1,3'-dithiopropionate (Pierce #21549), N-succinimidyl-6-(4'-azido-2'-nitrophenylamino)hexanoate (Pierce #22588), sulfosuccinimidyl-6-(4'-azido-2'-
20 nitrophenylamino)hexanoate (Pierce #22589), sulfosuccinimidyl 2-(p-azidosalicylamido)ethyl-1,3'-dithiopropionate (Pierce #27716), and sulfosuccinimidyl 4-(p-azidophenyl)butyrate (Pierce #21562).

Other captureable moieties, such as sugar groups and thiol groups, are known in the art. Compounds reactive with sugar groups, such as phenyl azide-hydrazide,
25 can be useful for capture onto solid-phase supports. Compounds reactive with thiol groups include maleimides, haloacetyls and pyridyl disulfides, such as the NHS-ester maleimide, NHS-ester haloacetyl, and NHS-ester pyridyl disulfide compounds provided above. Still other possible captureable moieties include non-selective photoreactive compounds, such as azidobenzoyl Hydrazide (Pierce #21510, 21509),
30 N-[4-azidosalicylamido]butyl]-3'-(2'-pyridyldithio)propionamide (Pierce 21512), p-Azidophenylglyoxal monohydrate, 4-(p-Azidosalicylamido)butylamine, 1-(p-Azidosalicylamido)-4-(iodoacetamido)butane (Pierce #21511), and Bis-[β-4-azidosalicylamido)ethyl]disulfide (Pierce #21564).

- 22 -

The captureable moiety may be attached to an anchor primer using methods well-known in the art (see Hermansson, *Bioconjugate Techniques* (Academic Press, New York, (1996); Wong, S., *Chemistry of Protein Conjugation and Cross-linking* (CRC Press, Florida (1991)). For example, reagents for incorporating a primary
5 amine into oligonucleotides are commercially available (Clontech, #5203-1, #5207-2, #5202-1) and described in the art (Connolly, B., *Nuc. Acid Res.*, 15:3131 (1987); Agrawal et al., *Nuc. Acid Res.*, 14:6227 (1986); Smith et al., *Nuc. Acid Res.*, 12:2399 (1985); Sproat et al., *Nuc. Acid Res.*, 15:6181 (1987); Sinha et al., *Nuc. Acid Res.*, 16:2659 (1988)). Also, reagents for incorporating other captureable moieties, such as
10 a biotin molecule or a thiol functional group, into oligonucleotides are commercially available (Clontech, #5024-1, #5021-1, #5211-1). For example, biotinylated UTPs, which are commercially available with different sized linkers (i.e. Biotin-11-dUTP, Enzo Biochemicals; Biotin-21-UTP, Clontech; Biotin-16-UTP, Boehringer Mannheim) can be used to produce biotinylated nucleic acids (Ampliscribe T7 High
15 Transcription Kit, Epicentre). One skilled in the art would realize additional methods and reagents for incorporating captureable moieties into oligonucleotides, such as, for example, introducing a thiol group into an amine-modified oligonucleotide through reaction with the amine group using 2-iminothiolane or Traut's reagent (Pierce #26101), SATA (Pierce #26102), or SPDP (Pierce #21757, 21657, 21557).

20 Additionally, an anchor primer having an attached captureable moiety may be purchased commercially such as, for example, a biotinylated anchor primer (Gibco Life Technologies, Grand Island, NY) or an amine-modified anchor primer (Midland Certified Reagents, Midland, TX).

The captureable moiety may be attached directly to a nucleotide of the anchor
25 primer. Alternatively, the captureable moiety may be attached indirectly to a nucleotide of the anchor primer via a spacer. In addition, the captureable moiety may be attached directly or indirectly to the substrate or solid-phase support. The spacer may encompass any suitable means that can be used to link a nucleotide of the anchor primer to the captureable moiety or any suitable means to link the captureable moiety
30 to the solid-phase support. However, the spacer should not adversely affect the function of the anchor primer or the captureable moiety.

- 23 -

The polynucleotide of the anchor primer (N_a) may comprise any nucleotide sequence consisting of A, G, T, U, and C nucleotides, as well as derivatized or modified nucleotides. Preferably, the anchor primer polynucleotide comprises a sequence that shares minimal homology with any sequence present in the cDNA library. The low homology design is important to minimize non-specific background generation during the gene-specific PCR amplification step discussed below. For example, a sequence that likely shares minimal homology to a sequence in the cDNA library is a rare restriction enzyme site. Examples of rare restriction enzyme sites that occur with a maximum frequency of about one site per 10,000 nucleotides (nt) in the mouse genome include: Not I (GCGGCCGC); Sbf I (CCTGCAGG, average restriction 1/15,000 nt); Spe I (ACTAGT, 1/15,000 nt); Sal I (GTCGAC, 1/20,000 nt); Rsr II (CGGA/TCCG, 1/60,000 nt); and Nhe I (GCTAGC, 1/10,000 nt). Preferably, the polynucleotide of the anchor sequence comprises a sequence containing at least one rare restriction enzyme site having at least six bases. More preferably, the polynucleotide contains a Not I restriction enzyme site.

The polynucleotide of the anchor primer comprises a sequence of at least 7 nucleotides. Preferably, the anchor primer polynucleotide comprises a sequence of from about 10 to about 40 nucleotides. More preferably, the anchor primer polynucleotide comprises a sequence of from about 20 to about 30 nucleotides. Most preferably, the anchor primer polynucleotide comprises a sequence of about 25 nucleotides. The polythymidylate of the anchor primer comprises a sequence of at least 8 thymidylate nucleotides. Preferably, the polythymidylate comprises a sequence of from about 10 to about 30 thymidylate nucleotides. More preferably, the polythymidylate comprises a sequence of from about 12 to about 25 thymidylate nucleotides. Most preferably, the polythymidylate comprises a sequence of about 18 thymidylate nucleotides. A particularly preferred captureable anchor primer in accordance with the present invention is a captureable anchor primer comprising the sequence 5'- NAATTCAACTGGAAGCGGCCGCAGGAA(T)₁₈-3' where N is a biotinylated guanidylate (SEQ ID NO:1). Another equally preferred captureable anchor primer in accordance with the present invention is a captureable anchor primer comprising the sequence 5'- NAATTCAACTGGAAGCGGCCGCAGGAA(T)₁₈-3' where N is an aminated guanidylate (SEQ ID NO: 2).

- 24 -

In other embodiments, an anchor primer comprising an RNA polymerase site is useful. A suitable anchor primer comprising a T3 site is SEQ ID NO: 45 or SEQ ID NO: 46. A suitable adapter molecule having a T7 site is SEQ ID NO: 13 or SEQ ID NO: 15. An acDNA_M library constructed using an anchor primer comprising a T3 site can be amplified using corresponding anchor amplification primers, e.g., chosen from the group consisting of SEQ ID NO: 47 or SEQ ID NO: 48. Alternatively, an acDNA_M library constructed using an anchor primer comprising a T7 site can be amplified using corresponding adapter amplification primers chosen from the group consisting of SEQ ID NO: 17, 18, or 19.

The 3' end modified cDNAs of the cDNA_M library are synthesized from a mRNA template using a captureable anchor primer under conditions for the preparation of double stranded cDNA that are well-known in the art. Such techniques are described, for example, in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, vol. 2, "Construction and Analysis of cDNA Libraries". One skilled in the art would realize the precise reaction conditions depend on the length and specific sequence of the captureable anchor primer. The reaction conditions for one preferred anchor primer is described in Example 1. The captureable anchor primer, 5'-N(N)_a(T)_b-3', hybridizes via its polydT sequence to the polydA sequence present on the 3' terminus of each mRNA molecule and serves as a template for first strand cDNA_M synthesis. The first cDNA_M strand is synthesized in the 5' or antisense orientation using a reverse transcriptase. Suitable reverse transcriptases include those from avian myeloblastosis virus (AMV) and Moloney murine leukemia virus (MMLV). The second cDNA_M strand is generated using an RNase and DNA polymerase to synthesize the complementary cDNA strand from the first cDNA_M strand, resulting in a double-stranded cDNA_M having a captureable anchor primer at its 3' end. Suitable RNases include, for example, RNase H and RNase One (Promega, #M4261). Suitable DNA polymerases include *E. coli* DNA Polymerase I (Promega). In one preferred embodiment of the present invention, a biotinylated anchor primer is used to generate a 3' end biotinylated cDNA library (cDNA_B). In another preferred embodiment, an aminated anchor primer is used to generate a 3' end aminated cDNA library (cDNA_A).

The captureable anchor primer selected for first strand cDNA production is dependent upon the purification procedure chosen to purify the gene-specific

- 25 -

sequence (discussed below). It is critical to the present invention that the captureable anchor primer be incapable of capture by the procedure used to capture the gene-specific sequence. As shown in Fig.1, following cDNA production, the 3' end (mRNA sense) of each double stranded DNA molecule in the cDNA library will be modified with a captureable moiety resulting in a modified double-stranded cDNA library (cDNA_M). This modification allows the capture of the complete cDNA_M library through the attachment of the captureable moiety to a substrate or solid-phase support. For example, a biotinylated cDNA library (cDNA_B) may subsequently be captured through attachment to an avidin-coated substrate. Accordingly, the selection of a biotinylated anchor primer and subsequent generation of a cDNA_B library is useful for the purification of a gene-specific sequence having a captureable moiety that does not form an attachment with avidin. An aminated cDNA library (cDNA_A) may subsequently be captured, for example, through attachment to an N-oxysuccinimide ester (NOS) activated substrate. Accordingly, the selection of an aminated anchor primer is useful for the purification of a gene-specific sequence having a captureable moiety that does not form an attachment with NOS.

As shown in Fig. 2, following construction of the 3' end modified double stranded cDNAs (cDNA_M), a double stranded adapter molecule is ligated to each cDNA_M thereby generating an adapted cDNA_M (acDNA_M). The double stranded adapter molecule may comprise any polynucleotide sequence consisting of A, G, T, U, and C nucleotides as well as derivatized and modified nucleotides. Preferably, the adapter molecule shares minimal homology with the anchor primer sequence. Also, the double stranded DNA adapter molecule preferably comprises a sequence which shares minimal homology with any sequence present in the cDNA library. The low homology adapter sequence design is important to minimize non-specific background generation during the gene-specific PCR amplification step. For example, the adapter molecule may be designed such that it comprises a sequence containing at least one rare restriction enzyme site as described above.

Since the objective is to construct an adapter molecule that shares minimal homology with cDNA library sequences, the number of rare restriction sites used to construct the adapter sequence should be based on the frequency with which the restriction enzyme site occurs in the genome of the relevant species. For example, a single rare restriction enzyme site sequence may be used to construct the adapter

- 26 -

molecule when the restriction enzyme site has eight or more bases. However, to ensure minimal homology, the adapter molecule preferably comprises a sequence containing two or more rare restriction enzyme sites that are sequentially arranged and preferably contiguous. More preferably, the adapter molecule comprises a sequence

5 containing about five rare restriction enzyme sites that are sequentially arranged and contiguous. Examples of suitable restriction enzymes sites are Not I, Asc I, Bae I, Fse I, Sfi I, Sgf I, SgrA I, Srf I, Pac I, Pme I, PpuM I, Rsr II, Sap I, Sex AI, Sse8387 I, Sbf I, Spe I, Sal I, Nhe I, and Swa I restriction enzymes sites. An example of a preferred adapter molecule comprises a sequence containing the Sbf I, Spe I, Sal I, Rsr II, and

10 Nhe I restriction enzyme sites which are sequentially arranged.

The double stranded adapter molecule may also be designed such that it comprises the sequence of an RNA polymerase promoter site. The RNA polymerase promoter site is the consensus sequence of the promoter site to which an RNA polymerase binds to initiate RNA synthesis. Suitable RNA polymerase promoter sites

15 include the T7, T3 and SP6 RNA polymerase promoter sites. A preferable RNA polymerase promoter site is the T7 RNA polymerase promoter site.

Preferably, the adapter molecule comprises the sequence of an RNA polymerase promoter site and at least one rare restriction enzyme site. Examples of suitable rare restriction enzyme sites include those listed previously. More preferably,

20 the adapter molecule comprises the sequence of an RNA polymerase promoter site and two or more rare restriction enzyme sites. For example, the adapter molecule may comprise the sequence of the T7 polymerase promoter site and Sbf I, Spe I, Sal I, Rsr II and Nhe I restriction enzyme sites. Thus, in one preferred embodiment, the double stranded adapter molecule comprises the sequence 5' - Sbf I - T7 promoter sequence -

25 Spe I - Sal I - Rsr I - Nhe I, corresponding with the sequence of 5' - ATAGCCTGCAGGTAATACGACTCACTATAGGGACTAGTCGACGGACCGCT AGCATCAGATC - 3' (SEQ ID NO: 13) and the complementary sequence (SEQ ID NO: 14). In yet another preferred embodiment, the double stranded adapter molecule comprises the sequence of 5' - ACGAGCGGATAACAATTTACACAGGGCGG-

30 CCGCTAATACGACTCACTATAGGGGTCGAC-3' (SEQ ID NO: 15) and the complementary sequence (SEQ ID NO: 16).

In accordance with the preferred embodiments of the present invention, the adapter molecule comprises a sequence of at least 15 nucleotides. More preferably,

- 27 -

the adapter molecule comprises a sequence of from about 20 nucleotides to about 70 nucleotides. Most preferably, the adapter molecule comprises a sequence of about 60 nucleotides.

Although it is not necessary, the double stranded adapter molecule is preferably constructed such that it has a cohesive end at its 5' end and a blunt end at its 3' end as shown in Fig. 2. This construction allows the adapter molecule to be ligated to the cDNA_M molecule in the proper orientation and also prevents multiple adapter ligations to one cDNA_M molecule. The duplex adapter molecule may be prepared by synthesizing complementary single strand adapter polynucleotides such that one of the single-strands is from about four nucleotides to about ten nucleotides longer than its complementary strand. The two single strands are subsequently annealed, producing a 5' end overhang. For example, in one embodiment shown in Fig. 2, the single strand adapter has a sequence of 5'-ATAGCCTGCAGGTAATACGACTCACTATAGGGA-CTAGTCGACGGACCGCTAGCATCAGATC - 3' (SEQ ID NO: 13) and the reverse complementary single strand has a sequence of 5' -GATCTGATGCTAGCGGTCCG-TCGACTAGTCCCTATAGTGAGTCGTATTACCTGCAGG- 3' (SEQ ID NO: 14) producing a four nucleotide 5' end overhang when the single strands are annealed.

The 3' blunt end of the adapter molecule is ligated to the 5' blunt end of each modified double stranded cDNA_M molecule using a blunt-end ligase and blunt-end ligation conditions well-known in the art (Sambrook et al., *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Press, Cold Spring Harbor, N.Y., 1989). Suitable blunt-end ligases include T4 ligase. Preferably, the adapter molecule is ligated to the cDNA_M molecules using T4 ligase. The captureable moiety at the 3'-end of the cDNA_M molecule prevents ligation of the adapter molecule to the 3' end of the cDNA_M.

In one preferred embodiment, the double stranded adapter molecule may be ligated to modified cDNAs having a captureable anchor primer which is biotinylated (cDNA_B), creating an adapted biotinylated cDNA library (acDNA_B). For example, as shown in Fig. 3, a double stranded adapter molecule comprising the sequence 5' - ATAGCCTGCAGGTAATACGACTCACTATAGGGA-CTAGTCGACGGACCGCT AGCATCAGATC - 3' and the complementary sequence (SEQ ID NO: 13 and 14) may be ligated to modified cDNAs having a captureable anchor primer comprising the sequence 5'- NAATTCAACTGGAAGCGGCCGCAGGAA(T)₁₈-3' where N is a

- 28 -

biotinylated guanidylate (SEQ ID NO:1). Also, for example, a double stranded adapter molecule comprising the sequence 5' –ACGAGCGGATAACAATTTTCAC-ACAGGGCGGCCGCTA ATACGACTCACTATAGGGGTCGAC-3' and the complementary sequence (SEQ ID NO: 15 and 16) may be ligated to modified

5 cDNAs having a captureable anchor primer comprising the sequence 5'-NAATT-CAACTGGAAGCGGCCGCAGGAA(T)₁₈-3' where N is a biotinylated guanidylate (SEG ID NO: 1). In another preferred embodiment, the double stranded adapter molecule may be ligated to modified cDNAs having a captureable anchor primer that is aminated, creating an adapted aminated cDNA library (acDNA_A). For example, a

10 double stranded adapter molecule comprising the sequence 5' –ATAGCCTGCAGGTAATACGACTCACTATAGGGACTAGTCGACGGACCG-CTAGCATCAGATC - 3' and the complementary sequence (SEQ ID NO: 13 and 14) may be ligated to modified cDNAs having a captureable anchor primer comprising the sequence 5' – NAATTCAACTGGAAGCGGCCGCAGGAA(T)₁₈ - 3' where N is an

15 aminated guanidylate (SEQ ID NO: 2). In another example, a double stranded adapter molecule comprising the sequence 5' – ACGAGCGGATAACAATTTTCACACAG-GGCGGCCGCTAATACGACTCACTATAGGGGTCGAC - 3' and the complementary sequence (SEQ ID NO: 15 and 16) may be ligated to modified cDNAs having a captureable anchor primer comprising the sequence 5'-

20 NAATTCAACTGGAAGCGGCCGCAGGAA(T)₁₈ - 3' where N is an aminated guanidylate (SEQ ID NO: 2).

Optionally, the complete acDNA_M library may be amplified using well-known PCR methods. The library amplification can increase the total amount of the non-cloned acDNA_M library. An additional advantage of the total acDNA_M amplification

25 is the up-regulation of rare transcripts that subsequently allow easier detection and manipulation of rare cDNA molecules. The acDNA_M library may be amplified using a set of two nested oligonucleotide amplification primers having opposing orientation, one of which is located in the adapter molecule sequence and the other of which is located in the anchor primer sequence. The adapter amplification primer (sense)

30 hybridizes with an antisense adapter molecule sequence and is extended in the 3' or sense orientation of the cDNA. The anchor amplification primer (anti-sense) hybridizes with a sense anchor primer sequence and is extended in the 5' or antisense orientation of the gene.

- 29 -

The adapter amplification primer (Ap) and anchor amplification primer (An) used to amplify the acDNA_M library are each preferably an oligonucleotide primer of sufficient length to act as a template for the synthesis of extension products.

Preferably, the Ap or An primer is an oligodeoxynucleotide of at least 12 nucleotides.

- 5 More preferably, the Ap or An primer is an oligodeoxynucleotide of from about 15 nucleotides to about 50 nucleotides. Most preferably, the Ap or An primer is an oligodeoxynucleotide of from about 20 nucleotides to about 40 nucleotides.

- The Ap and An amplification primers may be prepared using any suitable method known in the art or may be purchased commercially (MWG Biotech, Inc.,
10 Highpoint, N.C.). For example, the oligonucleotide primers may be prepared using the well-known phosphotriester, phosphoramidite, and H-phosphonate methods or automated embodiments thereof, as described in Engels et al., *Angew. Chem. Intl. Ed.*, 28, 716-734 (1989). In one known automated embodiment, diethylphosphoramidites are used as starting materials and may be synthesized as
15 described by Beaucage et al., *Tetrahedron Letters*, 22, 1859-1962 (1981). Alternatively, oligonucleotide primers may be isolated from a biological source, such as a restriction endonuclease digest.

- The Ap and An amplification primers are selected to be fully or substantially complementary to the different strands of the acDNA_M in the regions of the adapter
20 molecule and anchor primer sequences, respectively. Two single stranded nucleotide molecules are substantially complementary when the nucleotides of one strand, optimally aligned with appropriate nucleotide insertions, deletions or substitutions, pair with at least about 80% of the nucleotides of the other strand, preferably, at least about 90% to 95%, and more preferably, from about 98% to about 100%. Preferably,
25 the Ap and An amplification primers are fully complementary to the respective adapter molecule and anchor primer sequences.

- While any adapter molecule and anchor primer sequence can be selected for use as a primer, care should be taken to avoid sequences that have significant secondary structure, particularly at the 3' end of the primer. Thus, primers should be
30 selected which do not have the potential to form primer-dimer structures or be self-complementary. Computer programs, such as Squiggles or Circles (available from Univ. of Wisconsin) are useful for revealing these structures (see also, Rychlik, W., J.

- 30 -

NIH Res., 6, 78 (1994). Also, the adapter molecule and anchor primer sequences used for amplification should not be complementary with each other.

Thus, in accordance with the present invention, the complete acDNA_M library may be amplified using one Ap amplification primer which is fully or substantially
 5 complementary to an antisense sequence located in the adapter molecule and a second An amplification primer which is fully or substantially complementary to a sense sequence located in the anchor primer. Importantly, the An amplification primer must have at its 5' end the same captureable moiety as the anchor primer. For example, in one preferred embodiment shown in Fig. 3, in which the adapter molecule comprises
 10 the sequence 5' -ATAGCCTGCAGGTAATACGACTCACTATAGGGACTAGTC-GACGGACCG CTAGCATCAGATC - 3' (SEQ ID NO: 13), one Ap primer preferably comprises 5'- CCTGCAGGTAATACGACTCACTATAGG -3' (SEQ ID NO: 17). In another preferred embodiment, in which the adapter molecule comprises the above sequence, a different Ap primer preferably comprises 5'-TAGTCGACG-
 15 GACCGCTAGCATCAGATC - 3' (SEQ ID NO: 18). In yet another preferred embodiment in which the adapter molecule comprises the sequence 5'-ACGAGCG-GATAACAATTTACACAGGGCGGCCGCTAATACGACTCACTATAGGGGTC GAC-3' (SEQ ID NO: 15), one Ap primer preferably comprises the sequence 5'-CCGCTAATACGACTCACTATAGGGGTCGAC-3' (SEQ ID NO: 19).

20

Additionally, in a preferred embodiment in which the anchor primer comprises the sequence 5'- NAATTCAACTGGAAGCGGCCGCAGGAA(T)₁₈-3' (SEQ ID NO: 1 or 2). A preferred amplification primer comprises 5'-NAATTCAACTGGAAGC-GGCCGCAGGAAT -3' (SEQ ID NO: 20), where N is a biotinylated G nucleotide.
 25 In yet another preferred embodiment where the anchor primer comprises the sequence 5' - NAATTCAACTGGAAGCGGCCGCAGGAA(T)₁₈-3' (SEQ ID NO: 1 or 2), one An primer preferably comprises 5'-NAATTCAACTGGAAGCGGCCGCAGGA -3' (SEQ ID NO: 21), where N is an aminated G nucleotide.

30

Following hybridization of the Ap and An primers, the cDNAs are amplified using PCR amplification methods well-known in the art (see, e.g. Erlich, *PCR Technology*, Chap. 1, Stockton Press, New York, (1989); Rapley, R. and Walker, J., *Molecular Biomethods Handbook*, Humana Press, Totowa, New Jersey, (1998)).

- 31 -

Although PCR amplification conditions vary in accordance with the specific primers and template DNA present in the sample, the standard PCR reaction is typically done in a 50 or 100 μ l volume and usually contains, in addition to the template DNA, 10 mM TrisHCl (pH 8.4), 1.5-3.0 mM $MgCl_2$, 0.05-0.5 μ M of each primer, 50-200 μ M of each deoxytriphosphate (dATP, dCTP, dGTP, and dTTP), and 2.5 units of *Taq* polymerase. In addition, the PCR reaction may also include 50 mM KCl, gelatin, glycerol, or DMSO. For example, using the novel primers provided, the PCR reaction is preferably done in a 50 μ l volume and contains 10 mM TrisHCl (pH 8.4), 2.0-3.0 mM $MgCl_2$, 0.25 μ M of each primer, 200 μ M of each deoxytriphosphate (dATP, dCTP, dGTP, and dTTP), and 2.5 units of *Taq* polymerase. The amplification is typically performed automatically using a DNA Thermal Cycler. Suitable thermal Cyclers can be obtained from MJ Research, Inc. (Watertown, MA), PE Biosystems (Foster City, CA), Stratagene (San Diego, CA) and several other suppliers known in the art. Amplification typically consists of repeated cycles made up of three defined steps, termed denaturation (i.e., strand separation), primer annealing, and primer extension. In the first step, the template DNA (i.e., the acDNA_M library) is heated in excess of 90 degrees Celsius for 15-120 seconds to separate the double stranded DNA. In the next step, the temperature is reduced to 40-60 degrees Celsius for about 30-120 seconds to allow the oligonucleotide primers to anneal with the template DNA. In the final step, an enzymatic primer reaction is carried out at about 68°-72 degrees Celsius for about 1-5 minutes. The three separate stages are usually repeated between 20 and 40 times. In some cases, the primer annealing and extension can be done in one step at the same temperature (Kim et al., *Nuc. Acids Res.*, 16:8887-8903 (1988)). In addition to simplifying the procedure to a two-temperature cycle, simultaneously annealing and extending primers at a temperature greater than 55 degrees Celsius may improve the specificity of the reaction by minimizing the extension of primers that are mismatched with the template. The temperature at which the annealing (and extension) is done depends on the length and GC content of the primers. Also, the incubation time for the annealing/extension step may vary according to the length of the target gene being amplified. Preferably, the PCR reaction conditions and amplification parameters are such that optimal amplification of long cDNA molecules is achieved. Thus, for example, the extension step time may be increased to 10-20 minutes and the denaturation step time may be decreased so as to minimize

- 32 -

depurination of the template DNA. Also, glycerol or DMSO may be added to the reaction mixture to help lower the melting and strand separation temperature, thus minimizing template depurination. In addition, reducing the potassium concentration may improve the efficiency of amplification of longer PCR products. Using the
5 amplification primers provided in the present invention, the cycling program may be, for example: 94 degrees Celsius for 2 minutes, followed by 20 cycles of 94 degrees Celsius for 45 seconds (strand separation); 68 degrees Celsius for 5 minutes (primer annealing and extension). As mentioned, using a higher temperature for the primer annealing/extension step provides overall greater stringency which promotes primer
10 specificity. Also, allowing longer extension times promotes the synthesis of longer extended sequences and also allows for greater amplification (i.e., the production of a greater number of DNA molecules).

If shorter primers or degenerate primers (e.g. 80-95% complementary to the adapter molecule or anchor primer) are used, the reactions conditions and/or thermal
15 profile of the reaction may be adapted to reflect the lower stability of the primed template, in accordance with known methods. For example, shorter primers and degenerate primers may require lower annealing temperatures.

Synthetic RNA Alternative

In an alternative embodiment, wherein the adapter molecule of the acDNA_M
20 comprises an RNA polymerase promoter site, the complete acDNA_M library may be amplified using an RNA polymerase. This amplification method involves producing a synthetic RNA (sRNA) library from the acDNA_M library using an RNA polymerase to transcribe sRNA molecules *in vitro*. Since RNA polymerases produce about 100 molecules per one DNA template molecule, use of an RNA polymerase in this manner
25 results in an amplification of the acDNA_M library. This molar up-regulation is advantageous in that less starting material in the form of purified polyA⁺ selected mRNA is required for the detection of gene-specific sequences. In addition, the RNA amplification facilitates the detection of rare mRNA transcripts. This embodiment also provides the possibility to purify the gene specific DNA strand from the
30 synthetic RNA library. Quality control of the adapter ligation process is facilitated by monitoring the amount of transcript driven from the ligated promoter.

Any suitable RNA promoter consensus sequence and corresponding RNA polymerase may be used to produce the sRNA library. Examples of such suitable

- 33 -

RNA promoter sequences and RNA polymerases can be found in Watson et al., *Molecular Biology of the Gene*, 4th ed., ch. 13-15, Benjamin/Cummings Publishing Co., Menlo Park, CA. As discussed previously, preferably the adapter molecule comprises an RNA polymerase promoter site selected from the group consisting of T3, T7 and SP6 RNA polymerase promoter sites. Accordingly, the sRNA library is preferably generated using the corresponding RNA polymerase selected from the group consisting of T3, T7 and SP6 RNA polymerases. More preferably, the T7 promoter sequence and T7 RNA polymerase are used to generate the sRNA library. Thus, in one preferred embodiment using this amplification method, the adapter molecule comprises a sequence containing 5' - TAATACGACTCACTATAGGGA - 3' (SEQ ID NO 22), which is the consensus sequence for the T7 promoter. In an even more preferred embodiment, the adapter molecule comprises the sequence of 5' - AT-AGCCTGCAGGTAATACGACTCACTATAGGGACTAGTCGACGGACCGCTAGCATCAGATC - 3' (SEQ ID NO: 13). In an equally preferred embodiment, the adapter molecule comprises the sequence 5' -ACGAGCGGATAACAATTTCACACAGGGCGGCCGCTA ATACGACTCACTATAGGGGTCGAC-3' (SEQ ID NO: 15).

The synthetic RNA is synthesized from the acDNA_M template using well-known methods of *in vitro* transcription (Sambrook et al., pp. 10.5, 10.27-10.37). Given that the adapter molecule containing the RNA promoter sequence is at the 5' end of the acDNA_M, the synthetic RNA is transcribed in the sense orientation of the gene. Following synthetic RNA production, the synthetic RNA library is separated from the acDNA_M library by digestion of the acDNA_M molecules using a deoxyribonuclease, such as DNase I. The synthetic RNA library can be further purified using well-known methods of RNA column chromatography to remove the small DNA fragments. The resultant purified synthetic RNA library acts as the amplified template for gene-specific extended sequence generation discussed below.

II. Gene-Specific First Strand Sequence Generation

The second step of the present invention involves the generation of a gene-specific first strand cDNA sequence using a captureable gene-specific primer and either an acDNA_M library or a sRNA library. Each acDNA_M or sRNA molecule in

- 34 -

their respective libraries comprises a gene-specific sequence (i.e. the gene of interest), a captureable anchor primer at the 3' end and an adapter molecule at the 5' end. A captureable gene-specific primer is hybridized with a known sequence found in the gene-specific sequence of the acDNA_M or sRNA molecule and used to synthesize the partial or full-length gene-specific sequence. If an acDNA_M library is used, the gene-specific cDNA sequence is generated with a captureable gene-specific primer employing well-known methods of primer extension. If a sRNA library is used, the gene-specific cDNA sequence is generated with a captureable gene-specific primer through the process of reverse transcription.

10 A captureable gene-specific primer can be synthesized for a gene of interest after the sequence determination of a short fragment of the gene of interest. Such gene-specific sequence determinations may be deduced, for example, from the amino acid sequence. Additionally, gene-specific sequence determinations may be obtained by screening cDNA libraries to obtain short portions of cDNA clones (i.e. expressed
15 sequence tags or "ESTs") or by using existing EST databases (Venter et al., *Science*, 252, 1651-1656 (1991); Adams et al., *Nature*, 355, 632-634 (1992)).

The captureable gene-specific primer comprises the single stranded sequence 5'- Y(N)_c - 3', where Y is a nucleotide to which a captureable moiety is attached and N_c is a oligodeoxynucleotide in which c is an integer of from about 17 to about 40. Y
20 may be any nucleotide selected from the group consisting of A, G, C, U and T nucleotides, including derivatized and modified nucleotides. Importantly, the gene-specific primer has a captureable moiety attached to a 5'-end nucleotide that allows capture of the extended gene-specific sequence. Preferably, the captureable moiety is attached to the terminal nucleotide although it may be attached to internal nucleotides
25 as well. As described previously, the captureable moiety refers to any molecule that can be joined to a nucleotide Y and can also be attached to a substrate or solid support, such that Y is maintained on the solid support by the attachment of the captureable moiety. While the captureable moiety may form any type of attachment with the solid support, the attachment must be of sufficient strength to maintain the
30 nucleotide Y on the solid support. Preferably, the captureable moiety is attached to the substrate or solid support by affinity attachment or by covalent bonding. Suitable captureable moieties include, for example, avidin, streptavidin, neutravidin, biotin, primary amines, primary carboxylates, sugars, thiol alcohols, thiol carboxylates,

- 35 -

lipids, and peptides. Examples of suitable captureable moieties include those compounds listed previously.

Preferably, the captureable moiety is a biotin molecule. An equally preferable captureable moiety is that of a primary amine. As discussed previously, the
5 captureable moiety may be attached directly to a nucleotide of the gene-specific primer or, alternatively, may be attached indirectly to such nucleotide via a linker using methods well-known in the art (Hermansson, *Bioconjugate Techniques* (Academic Press, New York, (1996)). Alternatively, a gene-specific primer having a captureable moiety, such as a biotinylated gene-specific primer or an aminated gene-
10 specific primer may be purchased commercially.

While any suitable captureable moiety may be joined to the gene-specific primer, the captureable moiety must be such that it is different from the captureable moiety of the anchor primer found in the acDNA_M. The duality of the captureable moieties of the modified anchor primer and modified gene-specific primer allows
15 subsequent solid phase separations of the acDNA_M library from the gene-specific sequence. Thus, if the anchor primer of the acDNA_M library has, for example, an attached biotin moiety, the gene-specific primer may have, for example, an attached primary amine. Alternatively, if the anchor primer of the acDNA_M library has, for example, an attached primary amine, the gene-specific primer may have, for example,
20 an attached biotin molecule.

Once the sequence of a short fragment of a gene of interest is known, the oligodeoxynucleotide (N_c) sequence of the captureable gene-specific primer is selected to be fully or substantially complementary to the known sequence of the gene-specific fragment as that term has been previously defined. While the
25 oligodeoxynucleotide sequence of the captureable gene-specific primer may be any nucleotide sequence, the gene-specific primer should have a random nucleotide distribution and should not contain stretches of polypurines or polypyrimidines. Also, the gene-specific primer sequence should be such that the primer will not anneal with the anchor primer or the adapter molecule primer and will not anneal internally to
30 form hairpin loops or any other secondary structure (Erlich, Ed., *PCR Technology*, Chap. 1, (Stockton Press, New York (1989)).

- 36 -

The captureable gene specific primer is preferably a single stranded oligodeoxynucleotide of sufficient length to act as a template for the synthesis of the gene-specific sequence. The exact length of the gene-specific primer and the quantities used will depend on many factors, including temperature, degree of
5 homology and other reaction conditions. As known in the art, the annealing temperature will vary with the GC content of the particular primer and the composition of the hybridization solution. Preferably, the length of the gene-specific primer is sufficient to avoid non-specific priming under high annealing temperatures (i.e., temperatures above 60°). Thus, in accordance with the present invention, the
10 gene-specific primer is an oligodeoxynucleotide of at least 17 nucleotides. More preferably, the gene-specific primer is an oligodeoxynucleotide of from about 20 nucleotides to about 40 nucleotides. Most preferably, the gene-specific primer is an oligodeoxynucleotide of from about 25 nucleotides to about 30 nucleotides.

The gene-specific primer may be prepared using any suitable known method
15 for preparing single-stranded oligodeoxynucleotides and may be purchased commercially as well (MWG BioTech, Inc., High Point, N.C.) For example, the oligonucleotide primers may be prepared using the well-known phosphotriester, phosphoramidite, and H-phosphonate methods or automated embodiments thereof, as described in Engels et al., *Angew. Chem. Intl. Ed.*, 28, 716-734 (1989).

20 A. Gene-Specific Sequence Generation from the acDNA_M Library

In accordance with the present invention, the gene-specific cDNA of interest may be generated by primer extension of the captureable gene-specific primer following hybridization to the known gene-specific cDNA in the acDNA_M library. Although any suitable captureable gene-specific primer may be used, preferably the
25 captureable gene-specific primer is a biotinylated gene-specific primer or an aminated gene-specific primer. As shown in Fig. 3, the gene-specific sequence may be extended in either the 3' (sense) or 5' (antisense) orientation of the gene, depending on the gene-specific primer design. A gene-specific cDNA extended in the sense orientation will comprise the gene-specific sequence at its 5' end and the anchor
30 primer sequence at its 3' end. A gene-specific cDNA extended in the antisense orientation will comprise the gene-specific sequence at its 5' end and the adapter molecule sequence at its 3' end.

- 37 -

Thus, for example, a biotinylated gene-specific primer may be constructed such that it hybridizes to the antisense strand of the known gene-specific sequence of the acDNA_M molecule, allowing the gene-specific sequence to be extended in the sense orientation of the gene. The resultant biotinylated gene-specific cDNA
5 comprises the biotinylated gene-specific sequence at its 5' end and the anchor primer sequence at its 3' end. Alternatively, a biotinylated gene-specific primer may be constructed such that it hybridizes to the sense strand of the known gene-specific sequence of the acDNA_M molecule, allowing the gene-specific sequence to be extended in the antisense orientation of the gene. The resultant biotinylated gene-
10 specific cDNA comprises the biotinylated gene-specific sequence at its 5' end and the adapter sequence at its 3' end (corresponds to having the adapter sequence at the 5' end and gene-specific sequence at the 3' end of the gene). Fig. 4 summarizes diagrammatically the amplification primers used.

The captureable gene-specific primer may be extended with any suitable
15 thermostable DNA polymerase in the presence of dNTPs using well-known primer extension methodology. Also, the amplification of extended gene-specific sequences may be achieved using well-known PCR methods (Erich, Ed., *PCT Technology* (Stockton Press, New York (1989)). For example, a two step linear PCR amplification method, changing the temperature from 68 degrees Celsius to 94
20 degrees Celsius for multiple cycles (i.e. about 35), may be used to increase the production of gene-specific elongated DNA strand synthesis. The elongation time should be long enough to allow synthesis of the complete gene sequence to its 5' or 3' end. Typical elongation times are from about 0.5 minutes to about 15 minutes. Since DNA polymerase synthesizes about 1kb/min the skilled artisan will choose the time
25 appropriate for the possible maximum length of the template. After completion of the linear elongation and amplification, the PCR reaction contains a mixture of acDNA_M library molecules, the amplified elongated gene-specific single cDNA strands and the gene-specific primers.

30 B. Gene-Specific Sequence Generation from the sRNA Library

In another embodiment of the present invention, the gene-specific sequence of interest may be generated by reverse transcription following hybridization of a

- 38 -

captureable gene-specific primer to a known gene-specific sequence in the sRNA library. Preferably, the captureable gene-specific primer is a biotinylated gene-specific primer or an aminated gene-specific primer. Construction of the sRNA library results in a population of single-stranded sRNA molecules synthesized in the sense orientation. Thus, each sRNA molecule comprises a sense strand sequence containing the 5' end adapter molecule, the gene-specific sequence of interest, and the 3' end anchor primer. Using the previously described methods, a captureable gene-specific primer may be constructed which hybridizes to the known gene-specific sequence of the sRNA molecule. As shown in Fig. 7, the gene-specific cDNA is reverse transcribed in the antisense orientation, such that the resultant gene-specific cDNA comprises the gene-specific sequence at its 5' end and the adapter sequence at its 3' end (corresponds to having the adapter sequence at the 5' end and gene-specific sequence at the 3' end of the sense gene).

The gene-specific cDNA is reversed transcribed using reverse transcriptase and well-known methods for first strand cDNA synthesis (Sambrook et al., *Molecular Cloning: A Laboratory Manual*, vol. 2, "Construction and Analysis of cDNA Libraries"). Suitable reverse transcriptase include those from avian myeloblastosis virus (AMV) and Moloney murine leukemia virus (MMLV). A preferred reverse transcriptase is the MMLV reverse transcriptase.

III. Purification of the Gene-Specific cDNA

The next step of the present invention involves purifying the gene-specific sequence of interest by separating the gene-specific cDNA from the acDNA_M or, in other embodiments, the sRNA library. The separation of the gene-specific cDNA from the acDNA_M is achieved by the fact that the gene-specific cDNA and the acDNA_M library have different captureable moieties, which moieties allow for the differential capture of the gene-specific sequence. The separation of the gene-specific cDNA from the sRNA library is achieved, in part, through the different biochemical properties of the gene-specific cDNA sequence and the sRNA library which allow the sRNA library to be degraded without affecting the integrity of the gene-specific cDNA.

A. Purification of the Gene-Specific cDNA from an acDNA_M Library

- 39 -

The successful purification of a gene-specific cDNA may be accomplished by constructing any combination of captureable acDNA_M libraries and captureable gene-specific cDNAs of interest, as long as the combination allows for the differential capture of the gene-specific cDNAs. In general, the gene-specific purification may be achieved by capturing either the gene-specific cDNA or the acDNA_M library such that the unbound cDNA population can be separated from the bound or captured cDNA population. Thus, for example, the gene-specific purification may be achieved by capturing the gene-specific cDNA and removing the unbound acDNA_M library.

Alternatively, the gene-specific purification may be achieved by capturing the acDNA_M library and obtaining the unbound gene-specific cDNA. Additionally, the purification of a gene-specific cDNA may be accomplished through a series of capture steps. For example, the purification may be accomplished by first capturing one cDNA population (i.e. gene-specific cDNA or acDNA_M library) on a substrate and subsequently capturing the other cDNA population on a second substrate. The purification can be measured in terms of the quality of sequenceable gene-specific material, defined by: (1) the length of the sequence and (2) the number of readable sequences from an ABI 377 sequence trace. The two step procedure produces substantially better results than the one step method, judged by quality of sequence.

The substrate may be any substrate or solid support that has a rigid or semi-rigid surface suitable for capturing the intended molecule. The substrate has a surface chemistry that forms an attachment between the substrate and captureable moiety. Alternatively, the substrate has a surface that can be coated or derivatized such that the coated or derivatized substrate forms an attachment with a captureable moiety.

The substrate may comprise biological, nonbiological, organic, or inorganic materials, or a combination of these. Examples of suitable materials include agarose, nitrocellulose or other membrane materials, glass, modified silicon, (poly)tetrafluoroethylene, (poly)vinylidenedifluoride, polystyrene, polycarbonate, or combinations thereof. Other suitable substrates may be used and will be readily apparent to the ordinary skilled artisan. Substrates may exist as particles, strands, precipitates, gels, sheets, tubing, spheres, beads, pellets, containers, capillaries, films, plates or slides. Preferably, the solid support takes the form of a multiwell plate, such as a PCR plate, or a small bead or pellet, or other convenient form.

- 40 -

In one preferred embodiment shown in Fig. 5, a gene-specific cDNA may be purified by constructing a biotinylated acDNA (acDNA_B) library and an aminated gene-specific primer using previously described methods. The aminated gene-specific primer is hybridized with the acDNA_B library and used to generate aminated gene-specific cDNA, as previously described. Following linear PCR elongation and amplification of the aminated gene-specific cDNA, the biotinylated acDNA_B can be separated from the aminated gene-specific cDNA by contacting the acDNA_B: aminated gene-specific cDNA mixture with a streptavidin-coated substrate. The biotinylated acDNA_B molecules bind to streptavidin, thereby capturing the acDNA_B molecules on the streptavidin-coated substrate and allowing for the separation of the unbound aminated gene-specific cDNA which remains in the liquid phase. Any substrate having a surface chemistry that can be derivatized by streptavidin may be used. Suitable streptavidin-coated substrates include, for example, multi-well microtitre plates, PCR tubes, polystyrene beads, paramagnetic polymer beads, superparamagnetic polymer beads, and paramagnetic porous glass particles. A preferred streptavidin-coated substrate is a suspension of superparamagnetic avidin- and oligodT- coated polymer beads (Dynal, Inc., Lake Success, NY) which allows the removal of the bound acDNA_B library from the liquid phase by magnetic forces.

In some instances, where a one-step capture procedure may not result in a sufficiently purified population of gene-specific cDNAs, a second capture step may be performed using a different capture substrate. For example, following the first reaction of the acDNA_B: aminated gene-specific cDNA with a streptavidin-coated substrate, the unbound cDNA mixture containing aminated gene-specific cDNA and any residual unbound acDNA_B may be subsequently contacted with a NOS-coated substrate. The aminated gene-specific cDNA molecules bind to NOS, thus capturing the gene-specific cDNA on the NOS-coated substrate and further separating the aminated gene-specific cDNA from any unbound acDNA_B. Any substrate having a surface chemistry that can be actively derivatized by NOS may be used. Suitable NOS-coated substrates include, for example, multi-well PCR plates (DNA-BIND, Costar). Such PCR plates are especially advantageous because the attached gene-specific single stranded cDNA is stable at 4 degrees Celsius or at -20 degrees Celsius and can be used for repeated gene-specific PCR amplifications (see Example 16).

- 41 -

In another preferred embodiment shown in Fig. 6, a gene-specific (Gs) cDNA may be purified by constructing an aminated acDNA (acDNA_A) library and a biotinylated gene-specific primer using previously described methods. The biotinylated gene-specific primer is hybridized with the acDNA_A library and used to generate biotinylated gene-specific cDNA, as previously described. Following linear PCR elongation and amplification of the biotinylated gene-specific cDNA, the acDNA_A can be separated from the biotinylated gene-specific cDNA by contacting the acDNA_A: biotinylated gene-specific cDNA mixture with a streptavidin-coated substrate. The biotinylated gene-specific cDNA binds to streptavidin, thus capturing the gene-specific cDNA on the streptavidin-coated substrate and allowing for the separation of the unbound aminated acDNA_A library from the captured biotinylated gene-specific cDNA. Suitable streptavidin-coated substrates are previously described. A preferred streptavidin-coated substrate is a suspension of superparamagnetic avidin- and oligodT- coated polymer beads (Dynal, Inc., Lake Success, NY) which allows the removal of the bound gene-specific cDNAs from the liquid phase by magnetic forces. Another possible streptavidin-coated substrate is a streptavidin-coated multi-well PCR plate. These PCR plates are especially advantageous because the attached gene-specific single stranded cDNA is stable at 4 degrees Celsius or at -20 degrees Celsius and can be used for repeated gene-specific PCR amplifications.

Additionally, the above purification may be accomplished in two steps. Prior to the above reaction of the acDNA: biotinylated gene-specific cDNA with a streptavidin-coated substrate, the mixture may be contacted with a NOS-coated substrate to remove the acDNA_A library.

It should be noted that the discarded acDNA_M library can be re-used to isolate additional gene-specific sequences and may also be re-amplified.

B. Purification of the Gene-Specific cDNA from a sRNA Library

The successful purification of a gene-specific cDNA may also be accomplished by constructing a sRNA library that can be used as a template for gene-specific sequence extension by reverse transcriptase. As discussed previously, a sRNA is a single stranded RNA comprising a gene-specific sequence of interest, a 5' end adapter sequence, and a 3' end anchor sequence. Using a captureable gene-

- 42 -

specific primer that hybridizes to the known gene-specific sequence in the sRNA template, a captureable gene-specific cDNA is reverse transcribed in the antisense orientation. In accordance with this embodiment, the captureable gene-specific cDNA is purified by reacting the sRNA: captureable gene-specific cDNA mixture with an
5 RNAse to degrade the sRNA library. Any suitable RNAse may be used to degrade the sRNA library. Examples of preferred RNAases include RNAse H and RNAse One.

Additionally, if further purification is desired, the captureable gene-specific cDNA may be further purified using the methods described above. For example,
10 biotinylated gene-specific cDNA may be further purified by contacting the biotinylated gene-specific cDNA and any residual sRNA with a streptavidin-coated substrate. The biotinylated gene-specific cDNA molecules bind to streptavidin, thus capturing the gene-specific cDNA on the streptavidin-coated substrate and allowing for the separation of the unbound sRNA library from the captured biotinylated gene-specific cDNA. Any suitable streptavidin-coated substrates described above may be
15 used to capture the biotinylated gene-specific cDNA. A preferred streptavidin-coated substrate is a suspension of superparamagnetic avidin- and oligodT- coated polymer beads. Another preferred streptavidin-coated substrate is a streptavidin-coated multi-well PCR plate which can be used for repeated gene-specific PCR amplifications.

Also, for example, following RNAse treatment, aminated gene-specific cDNA
20 may be further purified by contacting the aminated gene-specific cDNA and any residual sRNA with a NOS-coated substrate. The aminated gene-specific cDNA binds to NOS, thus capturing the gene-specific cDNA on the NOS-coated substrate and further separating the aminated gene-specific cDNA from any residual sRNA.
25 Preferably, the NOS-coated substrate is a NOS-coated multi-well PCR plate which can be used for repeated gene-specific PCR amplifications.

IV. Gene-Specific cDNA Amplification

The above purification methods produce highly purified, gene-specific single
30 stranded extended cDNA sequences. The extended cDNA is defined at its 5' end by a known gene-specific primer sequence and defined at its 3' end by either a known anchor sequence or a known adapter sequence, depending on the orientation of the sequence extension. In a preferred embodiment, the gene-specific cDNA may be

- 43 -

amplified by standard PCR methods using a set of two nested primers having opposing orientation, one of which is located in the known gene-specific sequence and the other of which is located in either an adapter sequence or an anchor primer sequence.

5 As previously discussed, during the gene-specific cDNA generation step (Fig. 3), a gene-specific sequence may be extended in either the 5' (sense) or 3' (antisense) orientation of the gene, depending on the gene-specific primer design. A gene-specific cDNA extended in the sense orientation will comprise a captureable gene-specific sequence at its 5' end and an anchor primer sequence at its 3' end. A gene-specific cDNA extended in the antisense orientation will comprise a captureable gene-specific sequence at its 5' end and an adapter molecule sequence at its 3' end. Thus, in one preferred embodiment in which the gene-specific sequence is extended in the sense orientation, the gene-specific cDNA comprises a biotinylated known gene-specific sequence at its 5' end and an anchor primer sequence at its 3' end. In another equally preferred embodiment, the gene-specific cDNA comprises an aminated known gene-specific sequence at its 5' end and an anchor primer sequence at its 3' end. Additionally, where the gene-specific sequence is extended in the antisense orientation, preferred embodiments include a gene-specific cDNA comprising a biotinylated known gene-specific sequence at its 5' end and an adapter sequence at its 3' end, as well as a gene-specific cDNA comprising an aminated known gene-specific sequence at its 5' end and an adapter sequence at its 3' end.

 Accordingly, a purified gene-specific cDNA comprising a captureable gene-specific sequence at its 5' end and an anchor primer sequence at its 3' end may be amplified using a set of two nested primers, one of which is located in the known gene-specific sequence (sense primer) and the other of which is located in the anchor sequence (anti-sense primer). Amplification is achieved through anchor primer extension in the 5' or antisense orientation and gene-specific primer extension in the 3' or sense orientation of the gene. Likewise, a purified gene-specific cDNA comprising a captureable gene-specific sequence at its 5' end and an adapter molecule sequence at its 3' end may be amplified using a set of two nested primers, one of which is located in the known gene-specific sequence (anti-sense primer) and the other of which is situated in the adapter sequence (sense primer). Amplification is

- 44 -

achieved through adapter primer extension in the 3' or sense orientation and gene-specific primer extension in the 5' or antisense orientation of the gene.

The gene-specific amplification primer (Gs) is selected to be fully or substantially complementary, as that term has been defined, to the sense or antisense strand of the known gene-specific sequence depending on the orientation of the amplification. Preferably, the Gs amplification primer may comprise the same sequence as the gene-specific primer used to generate the single stranded gene-specific cDNA (without the attached captureable moiety). The An amplification primer (antisense) is selected to be fully or substantially complementary to the sense strand in the region of the anchor sequence. Also, if an acDNA_M library amplification step was performed, the An amplification primer used to amplify the purified gene-specific cDNA may comprise the same sequence as the An amplification primer used to amplify the acDNA_M library. The Ap amplification primer (sense) is selected to be fully or substantially complementary to the antisense strand in the region of the adapter sequence. Likewise, the Ap amplification primer used to amplify the purified gene-specific cDNA may have the same sequence as the Ap amplification primer used to amplify the acDNA_M library.

The known gene-specific (Gs), anchor (An) and adapter (Ap) amplification primers used to amplify the gene-specific cDNA sequences are each preferably a single stranded oligodeoxynucleotide of sufficient length to act as a template for the synthesis of extension products. The length of the Gs, An, or Ap primer should be sufficient to avoid non-specific priming under high annealing temperatures. Preferably, the Gs, An or Ap primer is an oligodeoxynucleotide of at least 18 nucleotides. More preferably, the Gs, An, or Ap primer is an oligodeoxynucleotide of from about 20 nucleotides to about 40 nucleotides. Most preferably, the Gs, An, or Ap primer is an oligodeoxynucleotide of from about 25 nucleotides to about 30 nucleotides. The Gs, An, and Ap amplification primers may be prepared using any suitable method known in the art, including the methods of oligonucleotide synthesis previously discussed.

In one preferred embodiment in which the purified gene-specific sequence comprises a captureable gene-specific sequence and an anchor primer comprising the sequence 5' – NAATTCAACTGGAAGCGGCCGCAGGAA(T)₁₈-3' (SEQ ID NO: 1),

- 45 -

one preferred An amplification primer is 5'-NAATTCAACTGGAAGCGGCCGCAG-GA -3' (SEQ ID NO: 20 or 21). In another preferred embodiment in which the purified gene-specific sequence comprises a captureable gene-specific sequence and an adapter molecule comprising the sequence 5' -ATAGCCTGCAGGTAATACGA-CTCACTATAGGGACTAGTCGACGGACCG CTAGCATCAGATC - 3' (SEQ ID NO: 13), one Ap primer is preferably 5'-CCTGCAGGTAATACGACTCACTATAGG -3' (SEQ ID NO: 17). In another preferred embodiment, in which the adapter molecule has the above sequence, a different Ap primer is preferably 5'- TAGTCGACGGACCGCTAGCATCAGATC - 3' (SEQ ID NO: 18). In yet another preferred embodiment in which the purified gene-specific sequence comprises a captureable gene-specific sequence and an adapter molecule comprising 5' - ACGAGCGGATAACAATTTACACAGGGCGGCCGCTA ATACGACTCACTATAGGGGTCGAC-3' (SEQ ID NO: 15), one Ap primer preferably comprises the sequence 5' - CCGCTAA-TACGACTCACTATAGGGGTCGAC-3' (SEQ ID NO: 19).

15

The gene-specific cDNA sequence may be amplified using standard PCR amplification methods known in the art and described previously. This amplification results in the generation of a gene-specific cDNA fragment population where all gene-specific cDNAs have one homogenous end which sequence is defined by the gene-specific primer and one end of heterogeneous length terminating with either the anchor primer or adapter molecule sequence, depending on the amplification primer choice. The heterogeneity in fragment length of the gene-specific cDNA population is due to the generation of truncated cDNA molecules during the initial cDNA_M library construction and also the difference in length of the mRNA polyA tail.

In a related embodiment, the present invention also provides a method for the simultaneous isolation and sequence determination of multiple novel genes. The method involves the utilization of several gene-specific primers, each having a captureable moiety at its 5' end and in a separate reaction chamber. The gene-specific primers are attached to a solid substrate via their captureable moieties. The captured gene-specific primers are each hybridized with a cDNA or RNA (sRNA or mRNA) library and used to generate gene-specific first strand cDNAs. Following removal of the library, the captured gene-specific first strands are amplified using a gene-specific primer and a library specific primer.

30

- 46 -

Thus, in a typical embodiment, the method comprises: (a) synthesizing a population of double stranded DNA molecules having a 5' end and a 3' end; (b) ligating a double stranded adapter molecule to the 5' end of each double stranded DNA molecule; (c) attaching at least one gene-specific oligonucleotide primer having a captureable moiety to a solid substrate, wherein the gene-specific primer is attached via the captureable moiety; (d) synthesizing at least one single stranded gene-specific polynucleotide using the attached gene-specific primer(s); (e) purifying the single-stranded gene-specific polynucleotide(s); and (f) amplifying the gene-specific polynucleotide(s) using a gene-specific primer and an primer that hybridizes to a sequence located in the adapter molecule or anchor primer, depending on the orientation of the gene specific primers.

In accordance with the present invention, an adapted cDNA library or an adapted sRNA library is constructed. A cDNA library is generated from mRNA with a polydT primer using methods well known in the art. Following construction of the cDNA library, a double stranded adapter molecule is ligated to each cDNA, thereby generating an adapted cDNA library. Suitable double stranded adapter molecules have been previously described. Preferably, the adapter molecule is designed such that it comprises a sequence containing at least one rare restriction enzyme site, as that term has been previously described. Also preferably, the double stranded adapter molecule comprises the sequence of an RNA polymerase promoter site. A preferable RNA polymerase promoter site is the T7 RNA polymerase promoter site. Thus, in one preferred embodiment, the adapter molecule comprises the sequence of SEQ ID NO: 13. In another preferred embodiment, the double-stranded adapter molecule comprises the sequence of SEQ ID NO: 15.

The double stranded adapter molecule is preferably constructed such that it has a cohesive end at its 5' end and a blunt end at its 3' end. The 3' blunt end of the adapter molecule is ligated to the 5' blunt end of the cDNA using a blunt end ligase and well-known blunt end ligation conditions.

The adapted cDNA library may be amplified using a pair of amplification primers consisting of an anchor amplification primer and an adapter molecule amplification primer and the previously described PCR amplification methods.

- 47 -

Alternatively, a modified cDNA_M library, i.e. a cDNA library having a captureable anchor sequence, or having an anchor sequence without the captureable moiety, can be used.

5 In another embodiment, wherein the adapter molecule comprises an RNA polymerase promoter site, a sRNA library is constructed as previously described.

Captureable gene-specific oligonucleotide primers are synthesized for genes of interest after the sequence determination of a short fragment of the genes of interest. Such gene-specific sequence determinations may be deduced, for example, from the amino acid sequence or may be obtained using existing EST databases. As previously
10 described, the captureable gene-specific primer comprises the single stranded sequence 5'- Y(N)_c -3', where Y is a nucleotide to which a captureable moiety is attached and N_c is a oligodeoxynucleotide in which c is an integer of from about 17 to about 40. Each of the gene-specific primers has a captureable moiety attached to the
15 5'-terminal nucleotide that allows capture of the extended gene-specific sequence to a substrate or solid support. Preferably, the captureable moiety is affixed to the solid support by affinity attachment or by covalent bonding. Suitable captureable moieties have been previously described above and include, for example, avidin, streptavidin, neutravidin, biotin, primary amines, primary carboxylates, sugars, thiol alcohols, thiol carboxylates, lipids, peptides, and the other compounds previously listed. Preferably,
20 the captureable moiety is a biotin or primary amine.

The gene-specific primer is attached to a substrate or solid support via the captureable moiety. Suitable substrate materials and forms have been previously described. In one preferred embodiment, a biotinylated gene-specific primer is affixed to a streptavidin-coated substrate, such as streptavidin-coated thin wall PCR
25 tube. In another preferred embodiment, an aminated gene-specific primer is attached to a NOS-activated thin wall PCR tube.

The gene-specific first strand cDNAs are synthesized with the appropriate enzyme using the cDNA or RNA as a template. First, the adapted cDNA, synthetic RNA molecules are hybridized with the attached gene-specific primers using well-
30 known methods. Next, the captured gene-specific primers are extended with any suitable thermostable DNA polymerase in the presense of dNTPs using previously described primer extension methods. The resultant gene-specific cDNAs each

- 48 -

comprise a known gene-specific primer sequence at the 5' end and an adapter sequence at the 3' end.

Upon completion of the gene-specific first-strand cDNA synthesis, the gene-specific cDNAs are purified by dissociating the cDNA or RNA library. The dissociation of the library from the gene-specific cDNAs is accomplished by heat
5 denaturation and several washing steps. If a sRNA library is used, an additional RNase treatment step may be employed to remove the sRNA. In another embodiment, the modified cDNA library described above can be removed using the suitable capture substrate, e.g. aminated gene-specific sequences were captured on a
10 NOS-coated substrate and avidin beads were used to remove a biotinylated library.

The purified gene-specific cDNAs are amplified using standard PCR technology with two nested primers, one of which is located in the gene-specific primer and the other of which is located in the adapter sequence. In one preferred embodiment, in which the adapter molecule comprises the sequence 5' -
15 ATAGCCTGCAGGTAATACGACTCACTATAGGGACTAGTCGACGGACCGCT AGCATCAGATC-3' (SEQ ID NO: 13), the adapter amplification primer (antisense) preferably comprises the sequence 5' - CCTGCAGGTAATACGACTCACTATAGG-3' (SEQ ID NO: 17).

In another preferred embodiment in which the adapter molecule comprises the
20 same adapter sequence, the adapter amplification primer preferably comprises the sequence 5' - TAGTCGACGGACCGCTAGCATC-3' (SEQ ID NO: 18). In yet another preferred embodiment, in which the adapter molecule comprises the sequence 5' - ACGAGCGGATAACAATTTACACAGGGCGGCCGCTAATACGACTCAC TATAGGGGTCGAC-3' (SEQ ID NO: 15), the adapter amplification primer
25 preferably comprises the sequence 5' - CCGCTAATACGACTCACTATAGGGGTC-GAC-3' (SEQ ID NO: 19). The gene-specific amplification primer is unique to each gene-specific cDNA and may comprise the same sequence as the gene-specific primer used to generate the first strand cDNA.

The double-stranded gene-specific cDNAs are present in the liquid phase after
30 completion of the PCR amplification. The single stranded gene-specific cDNA templates remain attached to the solid support which can be washed and stored at -20 degrees Celsius for further use.

- 49 -

V. Direct Sequencing of the Amplified Gene-Specific cDNA

One advantage of the present invention is that the purified gene-specific cDNA can be directly sequenced without cloning and screening steps. Thus, in another embodiment of the present invention, the method comprises the additional step of sequencing the isolated gene-specific cDNA product. The purified amplified gene-specific cDNA can be directly sequenced from the end containing the gene-specific sequence using the same gene-specific primer as used for the PCR amplification. Typical annealing and sequencing methods are discussed in the Examples.

Additionally, if further sequence information is necessary, the gene-specific cDNA can be sequenced using standard primer walking methodology in conjunction with the present invention (see Examples 15 and 16). For example, typically up to about 600-700 nucleotides of direct sequence can be obtained through one sequencing reaction for short, abundant genes (see Fig. 8). Additional sequence information may be obtained by designing a new gene-specific primer corresponding with an internal sequence further downstream or upstream of the gene-specific 5' end. The internal gene-specific primer hybridizes with the downstream or upstream sequence, allowing the determination of additional cDNA sequences. This process of using internal gene-specific sequencing primers to elucidate downstream or upstream sequences can be repeated until the complete sequence of the gene-specific cDNA has been determined (see Fig. 10).

Additionally, the full-length of a gene-specific cDNA may be obtained using primer walking in conjunction with the present invention. Depending on the size of the gene-specific cDNA of interest and the efficiency of the initial cDNA_M library construction, the gene-specific cDNA may be truncated at its 5' end rather than embody a full-length sequence. After the generation of a purified, amplified gene-specific cDNA sequence, the corresponding full-length cDNA may be obtained by repeating the invention using a modified gene-specific primer rather than a modified anchor primer. The modified gene-specific primer has a captureable moiety and corresponds with an internal sequence further upstream of the 3' end polydA tail. Following hybridization of the modified gene-specific primer to the mRNA, the first cDNA_M strand is synthesized in the previously described manner, resulting in an extended single strand cDNA comprising additional 5' end sequence (see Example 16). The same gene-specific template can be used for primer walking.

- 50 -

Since no cloning steps are involved, the method of the present invention can be completely automated. The automation of the PCR system provides several advantages. First, automation facilitates the determination of the optimal PCR conditions for the extended sequence generation of a particular gene-specific sequence. DNA fragments derived from cDNA libraries or differential cDNA display systems provide only partial gene sequence information and, in most cases, no additional information about the full-length gene size, gene sequence composition or the abundance of its mRNA transcript is known. Such limited information creates a challenge for the determination of the full-length gene sequence, given that the success of the sequence extension of a cDNA depends on several unknown factors, including the abundance of the gene in the adapted cDNA library, its length, the choice of gene-specific primers for the generation of the gene-specific first strand, the primers for the amplification of the gene-specific product and the primers for the direct sequencing. In addition, the success of the amplification of the gene specific sequence depends on the annealing temperature and the magnesium concentration used in the PCR reaction. Rather than determine the optimal PCR reaction conditions in individual experiments, the reaction variables can be modified and tested simultaneously using a 96 well PCR plate for the experimental design. Thus, even without prior knowledge of the gene-specific characteristics, automation allows the optimal extension and amplification conditions to be determined in an efficient manner. In addition, automation of the PCR reaction allows the extension and amplification of multiple gene-specific sequences simultaneously.

Another advantage of the present invention is that the purified gene-specific cDNA can be directly size-fractionated. Thus, in another embodiment of the present invention, the method comprises the additional step of size-fractionating the isolated gene-specific cDNA product. Suitable methods of size-fractionation include gel electrophoresis in its various forms, capillary electrophoresis and column (size) fractionation.

Yet another advantage of the present invention is that the purified gene-specific sequence can be directly cloned into a vector if desired. Thus, in another embodiment of the present invention, the method comprises the additional step of cloning the isolated gene-specific cDNA product into a vector. Preferably, the vector is an expression vector. The vector may be, for example, a phage, plasmid, viral, or

- 51 -

retroviral vector. Retroviral vectors may be replication competent or replication defective.

The gene-specific sequence should be operatively linked to an appropriate promoter such as the phage lambda PL promoter, the *E. coli lac, trp, pho A and tac* promoters, the SV40 early and late promoters, and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs should further contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation.

10 The coding portion of the transcripts expressed by the constructs should preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA, or UAG) appropriately positioned at the end of the polypeptide to be translated.

The expression vector should also preferably include at least one selectable marker for propagation in a host. Suitable markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture and tetracycline, kanamycin, or ampicillin resistance for culturing in *E. coli* and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as *E. coli*, *Streptomyces* and *Samonella typhimurium* cells, fungal cells, such as yeast cells, insect cells, such as *Drosophila* S2 and *Spodoptera* Sf9 cells, mammalian cells such as CHO, COS, 293, Bowles melanoma cells and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

15 20

Vectors preferred for use in bacteria include, for example, pQE70, pQE60, and pQE-9 (Quiagen, Inc.), pBluescript vectors, Phagescript vectors, pNH16A, pNH18A, pNH46A (Stratagene Cloning Systems, Inc.), and ptrc99a, pKK223-3, pDR540, pRIT5 (Pharmacia Biotech, Inc.). Preferred eukaryotic vectors include pWLNEO, pSV2CAT, pOG44, pXTI and pSG (Stratagene) and pSVK3, pBPV, pMSG, and pSVL (Pharmacia). Other suitable vectors will be readily apparent to the skilled artisan.

25

Generally, a plasmid vector is introduced into a host cell in a precipitate or in a complex with a charged lipid. For example, introduction of the vector construct can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods known in the art. Such methods are described in standard laboratory

30

- 52 -

manuals such as Davis et al., Basic Methods in Molecular Biology (1986). If the vector is a virus, it may be packaged in vitro using an appropriate packaging cell line and then transduced into host cells.

5 A polypeptide can be recovered and purified from recombinant cell cultures by well-known methods, including ammonium sulfate or ethanol precipitation, acid, extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography, and lectin chromatography. Preferably, high performance liquid chromatography (HPLC) is employed for purification.

10 The methods described above are used for the determination of novel gene sequences without prior knowledge of mRNA length and abundance. Most of the experimental steps are PCR based and therefore directly dependent on the parameters influencing the quality of PCR amplification, such as template concentration, buffer conditions, cycling temperature, cycle number, and cycle duration. The present
15 invention can be also be performed using a multi-well matrix, which would allow standardization of the PCR protocol for many different genes simultaneously.

Additionally, the present invention provides novel oligonucleotide primer sequences and compositions of oligonucleotide primer sequences which can be used to practice preferred embodiments of the invention. For example, a novel anchor
20 primer comprising SEQ ID NO: 1 is provided which can be used to generate biotinylated cDNA_B using an mRNA template. Additionally, a novel anchor primer comprising SEQ ID NO: 2 is provided which can be used to generate aminated cDNA_A. Also provided are two novel double-stranded adapter molecules comprising SEQ ID NO: 13 and SEQ ID NO: 15 and their respective complementary sequences
25 comprising SEQ ID NO: 14 and SEQ ID NO: 16, which can be used to generate an adapted acDNA_M library. Additionally, novel amplification primers which can be used to amplify the adapted acDNA_M library are provided. Such primers include anchor amplification primers comprising SEQ ID NO: 20 and SEQ ID NO: 21 and adapter amplification primers comprising SEQ ID NO: 17, SEQ ID NO: 18, and SEQ
30 ID NO: 19 and 44. Also, provided are amplification primers which can be used to amplify a gene-specific sequence. These include an anchor amplification primer comprising SEQ ID NO: 23 and above described adapter amplification primers (SEQ ID NO: 17-19).

- 53 -

The present invention also includes those oligonucleotide sequences which are substantially identical to the sequences of SEQ ID NO: 1-23, as that term has been defined. The oligonucleotide sequences may also include those sequences capable of hybridizing, under stringent hybridization conditions, to sequences that are
5 complementary to the sequences of SEQ ID NO: 1-23, as those conditions have been defined.

In another embodiment, the present invention also provides a composition comprising one or more of the novel DNA primer sequences described above and an aqueous carrier. For example, the composition may comprise an anchor primer
10 having a captureable moiety, an adapter molecule, an anchor amplification primer, an adapter amplification primer or any combination thereof. Any suitable aqueous carrier can be used in the composition. Preferably, the carrier renders the composition stable at a desired temperature, such as room temperature or storage temperature (i.e., 4 degrees Celcius to -20 degrees Celsius), and is of approximately neutral pH.
15 Examples of suitable carriers are known to those of ordinary skill in the art and include Tris-EDTA Buffer, and DEPC-H₂O.

In another embodiment, the present invention provides an article of manufacture comprising an anchor primer having captureable moiety and an attached substrate. Preferably, the anchor primer is a biotinylated anchor primer or an
20 aminated anchor primer. Also preferably, the substrate is an avidin substrate or a NOS-coated substrate.

The present invention also includes an isolation and sequencing system, preferably in kit form, for the rapid isolation and sequencing of novel genes from DNA and RNA libraries. The isolation and sequencing system includes, in an amount
25 sufficient to perform at least one assay, a single-stranded anchor primer oligonucleotide having an attached captureable moiety and a double-stranded adapter oligonucleotide as a separately packaged reagent for use in construction of the adapted acDNA_M library. In this embodiment, the user provides the gene-specific primer oligonucleotides corresponding with the gene of interest.

30 In another embodiment, the isolation and sequencing system additionally includes a single stranded amplification anchor primer oligonucleotide and a single stranded amplification adapter primer oligonucleotide as a separately packaged reagent. The user provides the gene-specific primer oligonucleotides corresponding

- 54 -

with the gene of interest. The PCR primers are utilized in pairs, as is well-known, based on the 5' end sequence of the gene-specific cDNA and the 3' end sequence of the adapter molecule or the 3' end of the anchor sequence.

In another embodiment, the isolation and sequencing system includes, in an amount sufficient to perform at least one assay, a single stranded anchor primer oligonucleotide having an attached captureable moiety, a double stranded adapter oligonucleotide, a single-stranded amplification anchor primer oligonucleotide, a single stranded amplification adapter primer oligonucleotide, and at least one single stranded gene-specific primer oligonucleotide having an attached captureable moiety as a separately packaged reagent. In accordance with the present invention, the captureable moiety of the anchor primer oligonucleotide is different from that of the gene-specific primer oligonucleotide. In this embodiment, at least one captureable gene-specific oligonucleotide having a sequence corresponding to that of a known gene or gene fragment, such as an EST, is provided. The gene-specific oligonucleotide is desirably obtained from a specific tissue of the user's choice. The gene-specific oligonucleotide may also be obtained from tissue that was treated by exposure to an agent that may alter gene expression, including chemical agents drugs, or growth factors.

In another embodiment, the isolation and sequencing kit includes at least one gene-specific primer that is captured on a substrate or solid support. Suitable substrates have been previously discussed and include glass, modified silicon, (poly) tetrafluoroethylene, (poly) vinylidene difluoride, polystyrene and polycarbonate materials which exist as particles, strands, precipitates, gels, sheets, tubing, spheres, beads, pellets, containers, capillaries, films, plates, or slides. Preferably, the gene-specific primers are captured on a PCR plate, a thin-wall PCR tube, or spheres (beads).

Preferred anchor primer oligonucleotides having an attached captureable moiety, double stranded adapter oligonucleotides, amplification anchor primer oligonucleotides and amplification adapter primer oligonucleotide are described previously and also described in the Examples.

Instructions for use of the packaged reagent(s) are also typically included. As used herein, the term "package" refers to a solid matrix or material such as glass,

- 55 -

plastic (e.g., polyethylene, polypropylene or polycarbonate), paper, foil and the like capable of holding within fixed limits an oligonucleotide of the present invention. Thus, for example, a package can be a plastic or glass vial used to contain up to microgram quantities of a contemplated oligonucleotide. Also, it can be a PCR plate
5 well or other solid phase substrate to which microgram quantities of a gene-specific primer oligonucleotide have been operatively affixed, i.e., attached so as to be capable of serving as a template for gene-specific generation and sequencing.

“Instructions for use” typically include a tangible expression describing a sequence of actions to be taken by the user, including at least one method parameter
10 such as the relative amounts of reagent and sample to be admixed, at least one reagent concentration, time periods for which a reagent sample admixture is to be maintained at a specified temperature, or suitable buffer characteristics.

The present invention for rapid, high throughput DNA sequencing has several diagnostic and research applications, including genetic and infectious disease
15 diagnosis, toxicology testing, individual identification, genetic research, occupational hazard screening, and agriculture and pharmaceutical development. It is estimated that there are 4000-5000 genetic diseases in humans, in which a mutational change in a gene hinders the function of the gene product, leading to a serious medical condition. Examples of genetic diseases for which associated mutations have been
20 identified include cystic fibrosis, phenylketonuria, Alzheimer’s disease, cancer, Duchenne muscular dystrophy, and familial hypercholesterolemia. Recent studies show that, while a genetic disease may be associated with one or a few specific mutations, the majority of genetic diseases are likely caused by any of numerous mutations scattered along the affected gene. Such cases necessitate a rapid, high
25 throughput DNA sequencing technique to search the entire affected gene for mutations that may be associated with a disease. By appropriate choice of gene-specific oligonucleotide primers, the present invention can be used to detect any mutation in a target gene, facilitating the diagnosis of genetic diseases as well as the identification of carriers of recessive genetic disorders. Moreover, the high
30 throughput nature of the present invention and its potential for automation makes it useful as a research tool in population studies aimed at determining which mutations within a target gene are associated with disease and which mutations represent

- 56 -

harmless polymorphisms. Such information is instrumental in the development of therapeutic agents.

In addition, the present invention may be useful in the diagnosis and identification of infectious agents. Each species or strain of micro-organism is expected to have unique DNA sequences which can be used as markers to identify that particular species or strain. Using the appropriate gene-specific primers, rapid high throughput DNA sequencing can determine a particular species or strain of virus or bacteria, facilitating diagnosis and treatment of the infectious disease. Also, the present invention may prove useful in the identification of new species or strains of micro-organisms.

Other diagnostic uses include DNA typing of individuals for use in forensics and paternity testing. Rapid, high throughput DNA sequencing can be useful for the analysis of polymorphic differences in the DNA of individuals. Furthermore, the potential for automation would allow the simultaneous screening of large numbers of polymorphic markers in the individual. Accordingly, the present invention offers significant advantages over the current technique of restriction fragment polymorphism analysis (RFLP), which is time-consuming, labor-intensive and does not provide sequence analysis. DNA typing will also be useful to verify the source and ownership of agricultural products, as new plants and livestock are developed by genetic engineering.

In addition, the present invention is useful as a research tool for the rapid wide-spread determination of novel full-length DNA sequences and for the characterization of complex multi-gene disorders and other genetic traits. For example, high throughput sequencing can be used in conjunction with known EST sequences and other data generated by the Human Genome Project to determine corresponding full-length cDNA sequences. Rapid, high throughput DNA sequencing can also be used to perform repetitive DNA sequence analysis of large numbers of individuals to determine the DNA sequences associated with multi-gene diseases and genetic traits. Additionally, the present invention can be used for wide-scale sequencing of plant and animal genomes. The sequence data derived from the above sequencing projects will aid in the development of pharmaceutical drugs and the creation of improved crops and livestock.

- 57 -

Another research use includes the detection of DNA mutations induced by environmental agents. Rapid sequencing can be used to screen the DNA of individuals exposed to high or chronic levels of chemicals or radiation to identify the induced mutations in specific genes as a way of predicting the mutagenic risk of such agents.

The following examples further illustrate the present invention but should not be construed to limit the present invention in any way.

10

EXAMPLE 1

This example demonstrates the construction of a modified cDNA library (cDNA_M), such as using a biotinylated anchor primer or an animated anchor primer. The resultant cDNA_B or cDNA_A library is constructed using the ThermoScript™ System for cDNA Synthesis (Cat. # 11146-016, 11146-032; Gibco Life Technologies, Grand Island, NY).

mRNA Isolation:

The poly A+ selected mRNA is prepared using standard techniques known in the art. See, e.g., Sambrook, J., et al., *Molecular Cloning, A Laboratory Manual*, 2d Ed., Cold Spring Harbor Press (1989), pp. 7.1 – 7.29.

First Strand cDNA Synthesis:

To make a biotinylated library, the isolated poly A+ selected mRNA (about 2 µg) is reacted with 2 µl of 500 ng/µl 5' biotinylated anchor primer (SEQ ID NO: 1 or 45) (MWG Biotech., Inc.) in a total volume of 10 µl DEPC-H₂O. The resulting mRNA/ anchor primer mixture is incubated at 70 degrees Celsius for 10 minutes and then quickly chilled on ice to stop the reaction. The mRNA/ anchor primer mixture is centrifuged at 14,000 x g for 10 seconds at 4 degrees Celsius and placed on ice.

The first strand reaction mixture is prepared by combining the mRNA/ anchor primer mixture with a mixture comprising the following components that is held on ice and vortexed gently to mix: 4 µl 5X cDNA Synthesis Buffer (250 mM Tris acetate, pH 8.4), 375 mM potassium acetate, 40 mM magnesium acetate, stabilizer - vortex 5 seconds just prior to use), 1 µl 0.1 DTT, 1 µl RNaseOUT (40 U/µl), 2 µl

- 58 -

10mM dNTPs, 1µl DEPC H₂O, 1µl Thermoscript Reverse Transcriptase (15 U/µl).
The first strand reaction mixture is incubated at 55 degrees Celsius for about 1 hour.
The first strand synthesis is terminated by incubating the mixture at 85 degrees
Celsius for about 5 minutes. The first strand reaction mixture is centrifuged at 14,000
5 x g for about 10 seconds and placed on ice.

Second Strand cDNA Synthesis:

The second strand cDNA synthesis is performed by adding the following reaction
mixture to the centrifuged first strand reaction mixture: 91µl DEPC-H₂O, 30µl 5X
Second Strand Buffer (Thermoscript cDNA synthesis kit, BRL), 3µl 10mM dNTPs,
10 1µl *E. coli* DNA ligase (10 units/µl), 4µl *E. coli* DNA polymerase I (10 U/µl). The
resulting mixture is vortexed gently and incubated at 16 degrees Celsius for about 2
hours. Next, 2µl (10 units) of T4 DNA Polymerase is added and the reaction mixture
is incubated at 16 degrees Celsius for 15 minutes to synthesize the second strand. The
second strand synthesis reaction is terminated by the addition of 10µl of 0.5M EDTA
15 to the reaction mixture, and the result is placed on ice.

Purification of Double-Stranded cDNA:

The double-stranded cDNA synthesized in the previous step is extracted and
purified by adding 150µl of Tris-buffered phenol:chloroform:isoamyl alcohol
(25:24:1) to the reaction mixture and vortexing for 30 seconds. The reaction mixture
20 is then centrifuged at 16,000 x g for about 5 minutes. The aqueous phase is removed
and mixed with 75µl 7.5 M NH₄OAc (2.5 M final) in 550µl 100% cold ethanol. The
resulting ethanolic mixture is vortexed thoroughly and incubated at about -20 degrees
Celsius for at least overnight. The ethanolic mixture is then centrifuged at 16,000 x g
at 4 degrees Celsius for about 30 minutes. The supernatant is discarded and the cDNA
25 pellet is dried at room temperature until all the residual ethanolic mixture is
evaporated. The cDNA is resuspended in the appropriate volume of DEPC-H₂O to
produce a solution of cDNA having a final concentration of 50 ng/µl.

Construction of a cDNA_A library is achieved using the same methods of
mRNA synthesis, first- and second-strand cDNA synthesis, and cDNA purification,
30 except that an aminated anchor primer (SEQ ID NO: 2, 46) is used for the first-strand
cDNA synthesis.

EXAMPLE 2

This example demonstrates the construction of an adapted modified cDNA library (acDNA_M), such as an adapted biotinylated or aminated cDNA library (acDNA_B or acDNA_A) by ligation of a double-stranded adapter molecule to the modified cDNA library.

A modified cDNA library, such as a biotinylated or aminated cDNA library, is constructed using the methods described in Example 1. The adapter is produced by annealing the oligonucleotides of SEQ ID NO: 13 and 14, creating a double-stranded DNA molecule with one cohesive end (5' sense) and one blunt end (3' sense). The two complementary primers were first denatured at 94 degrees Celsius for 5 minutes, followed by annealing steps through a gradual decrease in temperature from 94 degrees Celsius to 25 degrees Celsius in about 2 hours. The annealing step was performed in a 250mM NaCl solution. The 3' blunt-end of the double-stranded adapter molecule (4.5 µg) is ligated to the 5' end of the modified double stranded cDNA (1 µg) using T4 DNA ligase in a total volume of 50 µl ligase buffer (BRL Biolabs). The reaction is continued for 16 hours at 16 degrees Celsius.

An alternative adapter molecule is produced by annealing the oligonucleotides of SEQ ID NO: 15 and 16 using the same annealing conditions. This adapter molecule is blunt-end ligated to a modified cDNA library as described above.

EXAMPLE 3

This example demonstrates the amplification of an acDNA_M library using anchor amplification primers (An) and adapter amplification primers (Ap) to generate an amplified acDNA_M (aacDNA_M).

An acDNA_M library is made according to the methods described in Examples 1 and 2. The complete acDNA_M is amplified using a pair of primers, one of which is located in the anchor sequence and one of which is located in the adapter sequence. For example, an oligonucleotide of either SEQ. ID NO: 17 or 18 is used to amplify an acDNA_M having an adapter molecule of SEQ ID NO: 13. Likewise, an oligonucleotide of either SEQ ID NO: 19 or 44 is used to amplify an acDNA_M having an adapter molecule of SEQ ID NO: 15. In addition, an oligonucleotide of either SEQ

- 60 -

ID NO: 20 or 21 is used to amplify an acDNA_M having an anchor primer of SEQ ID NO: 1 or 2.

A small aliquot of the acDNA_M library (50 ng) is amplified using the appropriate adapter amplification primer and anchor amplification primer in the following reaction mixture: Platinum Taq Polymerase High Fidelity 1 U (GIBCO BRL cat # 11304-011), 1 x High fidelity PCR buffer, 2 mM MgSO₄, 0.2 mM dNTP mix. The acDNA_M is amplified by PCR at low cycle number and at high annealing temperature using the following 2-step PCR program: 1 cycle: 94°C -2 minutes; 20 cycles: 94°C -45 seconds; 68°C- 5 minutes. The amplified adapted cDNA library (aacDNA_M) is purified using spin columns (Qiagen).

The cDNA synthesis, adapter ligation and adapted cDNA amplification can be accomplished within a two-day time period. The cDNA library produced from 2 µg of polyA⁺ selected RNA is sufficient to extend the sequence of 100 different DNA fragments. Amplification of the adapted cDNA library produced enough template to extend the sequence of 10,000 different DNA fragments without the need to recreate or amplify the cDNA library.

EXAMPLE 4

This example demonstrates the construction of a sRNA library using an acDNA_M as a template.

An adapted cDNA_M library containing either the T3 and /or T7 DNA-dependent RNA polymerase consensus site in either the anchor primer or adapter molecule is used to construct a sRNA library. An acDNA_M library is produced according to the methods discussed in Examples 1 and 2; using an anchor primer and/or an adapter molecule having a T3 or T7 site. A suitable anchor primer having a T3 site is SEQ ID NO: 45 or 46. A suitable adapter molecule having a T7 site is SEQ ID NO: 13 or 15. The acDNA_M library can be amplified using anchor amplification primers of SEQ ID NO: 47 or 48 and adapter amplification primers of SEQ ID NO: 17, 18, or 19.

The acDNA_M or aacDNA_M library is used as a template for *in vitro* transcription using the RiboMAX Large Scale RNA production system (Promega, T₇ - #P1300; T₃ - #P1290) according to the manufacturer's instructions. The generated

- 61 -

sRNA library is separated from the acDNA_M or aacDNA_M library template by DNASE I treatment (RQ1 RNase-free DNase, Promega #M6101) for one hour at 37°C, followed by a purification step using a CENTRI-SEP spin column (Princeton Separations, #CS-901).

5

EXAMPLE 5

This example demonstrates the generation of an aminated or biotinylated gene-specific first strand cDNA using a synthetic RNA template.

10 Gene-specific first Strand cDNA Synthesis:

The sRNA is made according to the method of Example 4. The sRNA is reacted with 1.5µl of 100ng/µl aminated or biotinylated gene-specific primer in a total volume of 10µl DEPC-H₂O. The annealing reaction is incubated at 70°C for 10 minutes and then quick chilled on ice to stop the reaction. The reaction mixture is
15 then centrifuged at 13,000 x g for 10 seconds at 4°C and placed on ice. The following master reaction mix is prepared on ice and vortexed gently to mix: 4µl 5X cDNA Synthesis buffer (250 mM Tris acetate (pH 8.4), 375 mM potassium acetate, 40 mM magnesium acetate, stabilizer - vortex 5 seconds just prior to use), 1µl 0.1 DTT, 1µl RNaseOUT (40u/µl), 2µl 10mM dNTPs, 1µl DEPC H₂O, 1µl Thermoscript Reverse
20 Transcriptase (15 U/µl). The master reaction mix (10 µl) is added to the sRNA on ice. The cDNA synthesis reaction is incubated at 55°C for 1 hour and terminated by incubating the mixture at 85°C for 5 minutes. The first strand reaction mixture is centrifuged at 14 000 x g for 15 seconds and placed on ice.

Purification of the gene-specific first strand cDNA:

25 The gene-specific first strand cDNA is separated from the sRNA library by digesting the RNA using 5 Units of Rnase ONE (Promega cat# M4265) at 37°C for 30 minutes. The gene-specific first strand cDNA is further purified from RNA fragments, enzymes and buffers by spin column purification (Qiagen QIAquick spin column: cat #28104) using the protocol of the manufacturers.

30 The biotinylated or the aminated gene-specific first strand is ready for attachment to a solid phase as described in Example 8 or Example 9.

- 62 -

EXAMPLE 6

This example demonstrates the generation of an aminated gene-specific first strand cDNA using an acDNA_B or aacDNA_B template and an aminated gene-specific primer. Platinum Taq DNA Polymerase High Fidelity was used (Cat. No. 11304-011,
5 Gibco Life Technologies, Grand Island, NY) to generate the aminated gene-specific first strand cDNA.

As shown in Fig. 3, the single stranded gene-specific cDNA is synthesized in either the sense or antisense orientation depending on the orientation of the gene-specific primer. A gene specific cDNA synthesized from a gene-specific primer
10 having sense orientation contains the anchor sequence. A gene specific cDNA synthesized from a gene-specific primer having antisense orientation contains the adapter sequence. A linear amplification reaction is prepared using 50ng of biotinylated acDNA_B or amplified biotinylated aacDNA_B library as template, one aminated gene-specific primer (final concentration 200ng/ul) and Platinum Taq
15 Polymerase High Fidelity 1 U, 1 x High fidelity PCR buffer, 2 mM MgSO₄, 0.2 mM dNTP mix. (GIBCO BRL cat # 11304-011). (The reaction is cycled in a MJ Research Cycler using the following program: 94°C for 2 minutes; 94°C for 30 seconds, 60°C for 30 seconds, 68°C for 5 minutes, (40 cycles); 68°C for 5 minutes; 14°C, forever. The biotinylated adapted cDNA: aminated gene-specific first strand cDNA is purified
20 from the reaction mix (primers, buffer and enzyme) using QIAquick spin columns (Qiagen: cat #28104).

EXAMPLE 7

25 This example demonstrates the generation of a biotinylated gene-specific first strand cDNA using an acDNA_A or aacDNA_A template and a biotinylated gene-specific primer. Platinum Taq DNA Polymerase High Fidelity was used (Cat. No. 11304-011, Gibco Life Technologies, Grand Island, NY) to generate the biotinylated gene-specific first strand cDNA.

30 A linear amplification reaction is prepared using 50ng of an aminated acDNA_A or an amplified aminated aacDNA_A library as template, one biotinylated gene-specific primer (final concentration 200ng/ul) and Platinum Taq Polymerase High Fidelity 1 U, 1 x High fidelity PCR buffer, 2 mM MgSO₄, 0.2 mM dNTP mix. (GIBCO BRL cat

- 63 -

11304-011). The reaction is cycled in a MJ Research Cyclor using the following program: 94°C for 2 minutes; 94°C for 30 seconds, 68°C for 5 minutes, (40 cycles); 60°C for 30 seconds, 68°C for 5 minutes; 14°C, forever. The aminated adapted cDNA: biotinylated gene-specific first strand cDNA is purified from the reaction mix
5 (primers, buffer and enzyme) using QIAquick spin columns (Qiagen: cat #28104).

EXAMPLE 8

This example demonstrates the isolation of an aminated gene-specific cDNA from a biotinylated adapted library (acDNA_B or aacDNA_B).

10 The aminated gene-specific cDNA is isolated from the reaction products of Example 6 (i.e., a complete acDNA_B library and aminated the gene-specific single strand cDNA) using two independent purification steps. The first step involves the association of the biotinylated adapted cDNA library to streptavidin- and oligodT-coated Dynabeads (Dyna) and physical separation of the acDNA_B library through
15 magnetic forces from the aminated gene-specific single stranded cDNA. The second step involves the attachment of the aminated gene-specific cDNA to NOS-coated PCR plates and elimination of the remaining acDNA_B library molecules through a stringent washing procedure.

acDNA_B Library removal by streptavidin bead capture:

20 Streptavidin Dynabeads (Dyna Dynabeads M-280 Streptavidin cat# 112.05/06) and oligo dT beads (Dyna Dynabeads oligo (dT25) cat# 610.02) are washed and prepared as were used as recommended by the supplier. The aminated gene-specific single stranded cDNA and acDNA_B mixture (product of EXAMPLE 6) is adjusted to 1M NaCl and mixed with 10µl of the washed Streptavidin/ oligo dT
25 bead suspension. The gene-specific cDNA, acDNA_B library, and streptavidin/oligo dT bead mixture is incubated at 43°C with moderate shaking for 30 minutes. The beads are centrifuged at 12,000 x g for one minute to pellet the streptavidin/oligo dT beads with the attached aDNA_B library. The supernatant containing the aminated gene-specific single stranded cDNA is removed. The supernatant is purified from the
30 buffer solution using a QIAquick spin column.

Solid phase attachment of the gene-specific single stranded cDNA:

The aminated gene-specific cDNA obtained in Step 1 is further purified through the solid phase attachment to the wall of a NOS- coated PCR plate well. The

- 64 -

aminated gene-specific cDNA is suspended in 50 μ l of 500 mM NaPhosphate, pH 8.5. The gene-specific single stranded cDNA is added to a thin wall PCR plate having a N-oxysuccinimide surface (NOS) (Costar DNA-bind Thermowell M cat# 6573) and incubated at 37 degrees Celsius for 15 minutes (or overnight at 4 degrees Celsius) .

- 5 The gene-specific cDNA is blocked at 37 degrees Celsius for 30 minutes using blocking buffer (TTBS, 3% Bovine Serum Albumin, 0.05% Tween, pH 8.0). The blocking agent is removed and the gene-specific cDNA is re-suspended in TE buffer. The plate can be stored at 4 degrees Celsius for future amplification or can be washed for immediate amplification. The attached gene-specific cDNA is washed to
- 10 eliminate non-bound aDNA_B library molecules according to the following wash steps: (1) High wash buffer (3M SSC, 0.1% Sarkosyl) - repeat 3 times; (2) 0.3 NaOH - one time; (3) 0.1 m SSC - one time; (4) High Wash Buffer - repeat twice; and (5) H₂O - repeat 3 times..

EXAMPLE 9

- 15 This example demonstrates the isolation of a biotinylated gene-specific cDNA from an aminated adapted library (acDNA_A or aacDNA_A).

- The biotinylated gene-specific cDNA is isolated from the reaction products of Example 7 (i.e., a complete acDNA library and the biotinylated gene-specific single strand) by using two independent purification steps. The first step involves the
- 20 attachment of the aminated acDNA_A library to NOS-coated plates and recovery of the biotinylated gene-specific cDNA in solution. The second step involves the association of the biotinylated gene-specific cDNA to streptavidin coated Dynabeads (Dyna) and physical separation from the contaminating aminated acDNA_A through magnetic forces.

- 25 Solid phase attachment of the aminated acDNA_A library:

- The biotinylated gene-specific first strand cDNA:acDNA_A is suspended in 50 μ l of 500 mM NaPhosphate, pH 8.5. The mixture is added to a thin wall PCR plate having a N-oxysuccinimide surface (NOS) (costar DNA-bind Thermowell M cat# 6573) and incubated at 37 degrees Celsius for about 15 minutes (or overnight at 4
- 30 degrees Celsius). The biotinylated gene-specific cDNA that is not able to attach to the NOS-coated plate stays in solution and is transferred to a fresh tube for streptavidin capture.

- 65 -

Gene-specific cDNA isolation by streptavidin bead capture:

Streptavidin Dynabeads (DynaL Dynabeads M-280 Streptavidin cat# 112.05/06) are washed and prepared as recommended by the supplier. The product obtained from Step 1 is adjusted to 1M NaCl and mixed with 10µl of the washed Streptavidin suspension. The mixture is incubated at 43°C with moderate shaking for 30 minutes. The tube containing the mixture of the biotinylated gene-specific cDNA attached to the beads and the aminated acDNA_A library are centrifuged at 12,000 x g for one minute. The beads with the attached gene-specific cDNA are resuspended in 10mM TRIS pH7.5 and can be stored at 4°C.

10

EXAMPLE 10

This example demonstrates the amplification of a purified gene-specific first strand cDNA sequence.

A purified gene-specific first strand cDNA is prepared according to the method described in Example 8 or 9. An attached gene specific first strand cDNA directed in the antisense orientation is amplified using an antisense gene-specific primer and a sense adapter sequence primer (i.e., SEQ ID NO: 17, 18, 19, or 44). An attached gene-specific first-strand cDNA directed in the sense orientation is amplified using a sense gene specific primer and an antisense anchor sequence primer (i.e., SEQ ID NO: 23 or 49). The PCR amplification is performed using the Platinum Taq-High Fidelity System (cat # 11304-011).

20

PCR Amplification of Gene-Specific first strand cDNA directed in Antisense Orientation:

The gene-specific primer 200ng in a 50 µl reaction and adapter primer (SEQ ID NO: 17, 18, 19, 44) (200 ng) are added to a reaction mixture containing 1 U Platinum Taq Polymerase High Fidelity, 1 x High fidelity PCR buffer, 2 mM MgSO₄, 0.2 mM dNTP mix. (GIBCO BRL cat # 11304-011). The reaction mixture is added to a PCR well containing an attached aminated gene specific first strand cDNA (Example 8) or to a PCR tube containing a biotinylated gene-specific first strand cDNA associated with streptavidin coated beads (Example 9). The reaction is cycled in an MJ Research Cyler using the following program: 94 degrees Celsius for 2 minutes; 94 degrees Celsius 20 for seconds, 55 degrees Celsius for 30 seconds, 68 degrees Celsius for 4 minutes, (35 cycles); 68 degrees Celsius for 5 minutes; 14

30

- 66 -

degrees Celsius, forever. Following the PCR reaction, the supernatant containing soluble amplified gene specific products are purified from dNTPs, enzymes and buffers by spin column purification (Qiagen QIAquick spin column: #28104) using the protocol recommended by the supplier.

5 PCR Amplification of Gene-Specific cDNA Directed in the Sense Orientation:

The gene-specific primer (200ng in a 50 µl reaction) and anchor primer (200ng) (SEQ ID NO: 23 or 49) are added to a reaction mixture containing 1 U Platinum Taq Polymerase High Fidelity, 1x High fidelity PCR buffer, 2 mM MgSO₄, 0.2 mM dNTP mix. (GIBCO BRL # 11304-011). The reaction mixture is added to a
10 PCR well containing an attached aminated gene-specific first strand cDNA or to a PCR tube containing a biotinylated gene-specific first-strand cDNA associated with streptavidin beads. The PCR cycling program described above is used to amplify the gene-specific cDNA.

15 EXAMPLE 11

This example demonstrates the direct sequencing of a PCR product template synthesized from a captured gene-specific first strand cDNA.

A purified, amplified gene-specific PCR product is generated according to the method described in Example 10. The gene-specific PCR product is cycle sequenced
20 using the PRISMA Sequenase Terminator single-stranded DNA sequencing kit (Applied Biosystem, # 401458). Gene-specific sequencing primers located in the known DNA sequence fragment are used for sequence elongation. The sequencing reaction containing 200-500 ng gene-specific PCR template, 100 ng gene-specific sequencing primer and BigDye terminator are mixed and cycled in a MJ Research
25 Cyclor for 31 cycles (96 degrees Celsius for 2 minutes; 96 degrees Celsius for 15 seconds, 50 degrees Celsius for 15 seconds, 60 degrees Celsius for 4 minutes (31 cycles); 14 degrees Celsius, (forever)). The sequencing reaction is ethanol precipitated, subjected to gel or capillary electrophoresis on an ABI 377 or an ABI 3700, respectively (Applied Biosystems), sequenced and compared to databases using
30 functional analysis.

EXAMPLE 12

This example demonstrates the rapid sequence determination of a known gene, such as the cyclophilin gene, using the aminated capture approach.

- 5 Using the method described in Example 1, a capture modified cDNA_B library was constructed using mRNA isolated from whole mouse brain and a biotinylated anchor primer comprising SEQ ID NO: 1. Using the method described in Example 2, an adapted acDNA_B library was constructed by ligating an adapter molecule comprising SEQ ID NO: 13 and 14 to the 5' end of the cDNA_B. The acDNA_B library
- 10 was amplified according to the method described in Example 3 using an anchor amplification primer comprising SEQ ID NO: 20 and an adapter amplification primer comprising SEQ ID NO: 17. The cyclophilin-specific cDNA was synthesized in the antisense orientation according to the method described in Example 6 using an aminated cyclophilin primer comprising SEQ ID NO: 24. The resultant cyclophilin
- 15 specific single-stranded cDNA in antisense orientation comprised an aminated cyclophilin-specific sequence at its 5' end and an adapter sequence at its 3' end. The cyclophilin cDNA was purified using the method described in Example 8. The purified cyclophilin cDNA was amplified according to the method described in Example 10 using a cyclophilin-specific amplification primer comprising SEQ ID
- 20 NO: 50 and an adapter amplification primer comprising SEQ ID NO: 18. Finally, the cyclophilin cDNA was directly sequenced according to the method described in Example 11 using a cyclophilin sequencing primer comprising SEQ ID NO: 29 (see Fig. 8). The sequence obtained by directly sequencing the cyclophilin specific PCR product matched with 100% identity the sequence reported in Genbank (X52803).
- 25 One sequence read allowed the determination of 509 nucleotides following the cyclophilin specific sequencing primer.

EXAMPLE 13

- 30 This example demonstrates the rapid sequence determination of the known cyclophilin gene using the biotinylated capture approach.

Using the method described in Example 1, a capture modified cDNA_A library was constructed using mRNA isolated from whole mouse brain and an aminated

- 68 -

anchor primer comprising SEQ ID NO: 2. Using the method described in Example 2, an adapted, aminated acDNA_A library was constructed by ligating an adapter molecule comprising SEQ ID NO: 15 and 16 to the 5' end of the cDNA_A. The acDNA_A library was amplified according to the method described in Example 3 using an anchor
5 amplification primer comprising SEQ ID NO: 21 and an adapter amplification primer comprising SEQ ID NO: 44. The cyclophilin-specific cDNA was synthesized in the antisense orientation according to the method described in Example 7 using a biotinylated cyclophilin primer comprising SEQ ID NO: 25. The resultant cyclophilin specific single-stranded cDNA in antisense orientation comprised a biotinylated
10 cyclophilin-specific sequence at its 5' end and an adapter sequence at its 3' end. The cyclophilin cDNA was purified using the method described in Example 9. The purified cyclophilin cDNA was amplified according to the method described in Example 10 using a cyclophilin-specific amplification primer comprising SEQ ID NO: 50 and an adapter amplification primer comprising SEQ ID NO: 44. Finally, the
15 cyclophilin cDNA was directly sequenced according to the method described in Example 11 using a cyclophilin sequencing primer comprising SEQ ID NO: 29 (see Fig. 8). The sequence obtained by directly sequencing the cyclophilin specific PCR product matched with 100% identity the sequence reported in GenBank (X52803). One sequence read allowed the determination of 527 nucleotides following the
20 cyclophilin specific sequencing primer.

EXAMPLE 14

This example demonstrates the rapid sequence determination of the known cyclophilin gene using the sRNA capture approach.

25 Using the method described in Example 1, a capture modified cDNA_A library was constructed using mRNA isolated from whole mouse brain and an aminated anchor primer comprising SEQ ID NO: 46. Using the method described in Example 2, an adapted, aminated acDNA_A library was constructed by ligating an adapter molecule comprising SEQ ID NO: 13 and 14 to the 5' end of the cDNA_A. The sRNA
30 was produced by *in vitro* transcription with T7 polymerase according to the protocol described in Example 5. The biotinylated cyclophilin-specific cDNA was synthesized in the antisense orientation according to the method described in Example 5 using an biotinylated cyclophilin primer comprising SEQ ID NO: 25. The resultant cyclophilin

- 69 -

specific single-stranded cDNA in antisense orientation comprised a biotinylated cyclophilin-specific sequence at its 5' end and an adapter sequence at its 3' end. The cyclophilin cDNA was purified using the method described in Example 9. The purified cyclophilin cDNA was amplified according to the method described in

5 Example 10 using a cyclophilin-specific amplification primer comprising SEQ ID NO: 50 and an adapter amplification primer comprising SEQ ID NO: 44. Finally, the cyclophilin cDNA was directly sequenced according to the method described in

Example 11 using a cyclophilin sequencing primer comprising SEQ ID NO: 29 (see Fig. 8). The sequence obtained by directly sequencing the cyclophilin specific PCR

10 product matched with 100% identity the sequence reported in Genbank (X52803). One sequence read allowed the determination of 447 nucleotides following the cyclophilin specific sequencing primer.

EXAMPLE 15

15 This example demonstrates the rapid sequence determination of a first unknown gene by extension from a short DNA fragment of known sequence, DST M41, into adjacent sequence regions of unknown sequence, using the aminated capture approach (Fig. 9).

Using the method described in Example 1, a capture modified cDNA_B library

20 was constructed using mRNA isolated from whole brain tissue and a biotinylated anchor primer comprising SEQ ID NO: 1. Using the method described in Example 2, an adapted acDNA_B library was constructed by ligating an adapter molecule comprising SEQ ID NO: 13 and 14 to the 5' end of the cDNA_B. The acDNA_B library was amplified according to the method described in Example 3 using an anchor

25 amplification primer comprising SEQ ID NO: 20 and an adapter amplification primer comprising SEQ ID NO: 17.

A gene-specific fragment of 268 nucleotides was generated by the TOGA differential display system (U.S. Patent 5,459,037). The fragment contained a polyA stretch at its 3' end, indicating that the DNA fragment was located at the 3' end of a

30 gene. Since no open reading frame could be detected in the DNA fragment it was likely that the cDNA sequence was derived from the 3' untranslated region of a gene. Homology searches to DNA databanks (GeneBank) using the NCBI blast server did not reveal homology to any known gene sequence. In order to learn more about M41,

- 70 -

especially concerning the coding region of the gene, extended sequence was generated using the aminated solid phase method described in Example 6, 8, 10, 11. An aminated gene-specific primer corresponding to the M41 gene-specific fragment was synthesized (SEQ ID NO: 30). The M41 cDNA was generated in the antisense orientation according to the method described in Example 6 using the aminated M41 primer comprising SEQ ID NO: 30. The resultant M41 cDNA comprised an aminated M41-specific sequence at its 5' end and an adapter sequence at its 3' end. The M41 cDNA was purified using the method described in Example 8. The purified M41 cDNA was amplified according to the method described in Example 10 using a M41-specific amplification primer comprising SEQ ID NO: 32 and an adapter amplification primer comprising SEQ ID NO: 18. Finally, the M41 cDNA was directly sequenced according to the method described in Example 11 using an M41 sequencing primer comprising SEQ ID NO: 33. The extended sequence was obtained and could be analyzed within 2 days after starting the experimental protocols. The sequence was extended from the DNA fragment in the antisense orientation by 300 nucleotides of novel sequence, and overlapping the known sequence by 40 nucleotides. Sequence analysis revealed a stop codon that is located 7 nucleotides 5' of the M41 DNA fragment. The extended sequence revealed an open reading frame over the entire extended sequence. This protocol allows rapid sequence extension into the coding region of the gene. The coding sequence can be further explored using DNA and protein analysis tools to determine possible gene functions.

EXAMPLE 16

This example demonstrates the rapid sequence extension of a second unknown gene, M51, (Fig. 10) using the aminated capture approach and primer walking on the gene specific PCR product. This example also shows the reusability of the solid phase attached gene specific first strand as template for multiple rounds of gene specific PCR amplification.

Using the method described in Example 1, a capture modified cDNA_B library was constructed using mRNA isolated from whole brain tissue and a biotinylated anchor primer comprising SEQ ID NO: 1. Using the method described in Example 2, an adapted acDNA_B library was constructed by ligating an adapter molecule comprising SEQ ID NO: 13 to the 5' end of the cDNA_B. The acDNA_B library was

- 71 -

amplified according to the method described in Example 3 using an anchor amplification primer comprising SEQ ID NO: 20 and an adapter amplification primer comprising SEQ ID NO: 17.

5 A gene-specific fragment of 295 nucleotides (DST M51) was generated by the TOGA differential display system (U.S. Patent 5,459,037). In order to obtain sequence information adjacent to the known DNA fragment M51, the sequence was extended towards the 5' end of the gene in the antisense orientation using the aminated solid phase approach described in Example 6, 8, 10 and 11. An aminated M51 primer (SEQ. ID NO: 36) was synthesized, reverse complementary to 26
10 nucleotides of DST 51. The primer was directed in the antisense orientation and located 135 nucleotides downstream of the 5' end of the DST M51 fragment. The M51-specific sequence was extended using a biotinylated amplified adapted cDNA library (aacDNA_B) derived from mouse brain poly A+ selected mRNA using the method described in Example 6. The aminated gene-specific single strand (M51-
15 gspEXT_A) contained 135 nucleotides of the known DST M51 DNA fragment sequence at its 5' end and the adapter sequence (SEQ ID NO: 13) at its 3' end. The M51-gspEXT_A was isolated from the contaminating cDNA library (aacDNA_B) using the two consecutive capture protocols described in Example 8, resulting in the covalent attachment of M51-gspEXT_A to the wall of a N-oxysuccinimide surface
20 (NOS) PCR plate well (Costar DNA-bind Thermowell M, # 6573). The solid phase attached M51-gspEXT_A sequence was amplified using the M51-specific primer of SEQ ID NO: 37 (antisense orientation) and the adapter amplification primer of SEQ ID. NO: 18 (sense orientation) with the protocol described in Example 10. The resultant PCR product, M51-PCR1, was directly sequenced without subcloning using
25 the sequencing primer of SEQ ID NO: 38. The sequence analysis revealed that M51-PCR1 contained 111 basepairs of DST M51 (5' sense orientation) and was extended by 360 nucleotides of novel sequence. The protocols in Examples 6, 8, 10 and 11 were performed in a two day time frame, including the sequence analysis.

Based on the newly extended M51 sequence, a new sequencing primer (SEQ
30 ID NO: 40) located 120 nucleotides 5' of the DST M51 sequence was synthesized. A new direct sequencing reaction was performed using the M51-specific PCR product (M51-PCR1) with the sequencing primer of SEQ ID NO: 40. As a result of the second direct sequencing reaction, the sequence of M51 was further extended by 406

- 72 -

nucleotides, elongating the DST M51 sequence of 295 by 765 nucleotides within 4 working days. The M51-specific PCR product M51-PCR1 was used as template for 10 independent direct sequencing reactions using the protocol of Example 11.

In addition, new M51 gene-specific PCR products can be regenerated by using the solid phase attached gene-specific M51 extended sequence M51-gspEXT_A as
5 template. The NOS-coated PCR plate containing M51-gspEXT_A in one of the 96 wells can be stored in -20 degrees Celsius freezer. PCR plates containing covalently attached gene-specific first strands can be either stored at 4 degrees Celsius in PBS or lyophilized at -20 degrees Celsius for months without damage to the attached DNA.
10 For example, gene-specific DNA has been covalently attached and stored for 13 months without noticeable damage to the DNA. The PCR can be performed using the primers of SEQ ID NO: 37 (antisense orientation) and SEQ ID NO: 18 (sense orientation) as described above, creating the PCR product M51-PCR1. Alternatively, a new M51-specific PCR product can be synthesized located further upstream in the
15 novel extended sequence using the M51-specific primer of SEQ ID NO: 41 (antisense orientation) and the adapter primer of SEQ. ID. 18 (sense orientation). The newly synthesized PCR product, M51-PCR2, is located 78 nucleotides 5' of the DST M51 and contains the known adapter sequence on its 5' end (sense orientation). Either the PCR product M51-PCR1 or M51-PCR2 can be used for further extension of the M51
20 sequence by primer walking. M51-PCR2 was used as a template for direct sequencing using the primer of SEQ ID NO: 42, located 530 nucleotides upstream (sense) of DST M51. The sequence was further elongated by 405 nucleotides towards the 5' end of the gene, extending the DST M51 by 1170 of novel nucleotides. The sequence was further extended by direct sequencing (Example 11), using a sequencing
25 primer of SEQ ID NO: 43, located 695 nucleotides upstream of the DST M51 within the novel sequence generated using sequencing primer SEQ ID NO: 42. The direct sequencing generated 420 nucleotides of sequence overlapping and verifying the sequence generated with SEQ ID NO: 42 and extended the sequence of M51 by 174 nucleotides. The total time consumption for the primer walk using 4 sequencing
30 primers, including the synthesis of the new sequence primers and the sequence analysis, was one week. During this time, the M51 sequence was extended by 1341 nucleotides of novel sequence without the need of subcloning the extended sequences.

- 73 -

EXAMPLE 17

This example demonstrates the ability to use multiwell PCR plates in combination with the solid phase direct sequencing protocols for the simultaneous extension of multiple gene fragments.

5 The important parameters can be optimized conveniently in a single 96 well plate. Table 2, below is a diagrammatic representation of the use of a 96 well plate to optimize conditions for six templates (DST 1-6), two annealing temperatures in PCR, e.g. 55 degrees Celsius and 60 degrees Celsius, different primers, two different magnesium concentrations (M1, M2) and different template concentrations. Such a
10 configuration is readily used for automating the method by using one or more robots, such as a Biomek robot.

Each of the multiple gene fragments is defined by its specific sequence composition, its length and its expression profile. DNA fragments derived from cDNA libraries or differential cDNA display systems are only defined by their length
15 (partial gene sequence) and their characteristic sequence. In most cases no additional information on the full length gene size, gene sequence composition or its mRNA abundance is known. This limited information on the gene creates a challenge for the determination of the full-length sequence, especially in a genome that has not been completely sequenced. As experimentally determined the variables for the successful
20 sequence extension of a gene is dependent on the abundance of the gene in the adapted cDNA library, its length, the choice of gene specific primers for the generation of the gene specific 1st strand, the primers for the amplification of the gene specific product and the primers for the direct sequencing. In addition the successful amplification of the gene specific sequence by PCR is dependent on the annealing
25 temperature and the magnesium concentration used for the PCR amplification. All these variables can be easily modified for the optimization of the extended sequence generation simultaneously using a 96 well PCR plate for the experimental design (Table 2). Since no cloning steps are involved and the system can be complete automated and even without prior knowledge of the gene specific parameters optimal
30 sequence extension can be generated for many different genes simultaneously.

In a specific example, DST 1 is the cyclophilin gene fragment as described in Example 13, DST 2 is the M41 is the gene fragment described in Example 15, DST 3

- 74 -

is the M51 is the gene fragment described in Example 16, and DST4, DST 5 and DST 6 are other similar gene fragments.

Using the method described in Example 1, a capture modified cDNA_A library is constructed using mRNA isolated from a tissue source such as whole mouse brain and an aminated anchor primer comprising SEQ ID NO: 2. Using the method
5 described in Example 2, an adapted, aminated acDNA_A library is constructed by ligating a double-stranded adapter molecule comprising SEQ ID NO: 13 and 14 to the 5' end of the cDNA_A. The acDNA_A library is amplified according to the method described in Example 3 using an anchor amplification primer comprising SEQ ID NO:
10 20 and an adapter amplification primer comprising SEQ ID NO: 17. The cyclophilin-specific cDNA is synthesized in the antisense orientation according to the method described in Example 8 using an aminated cyclophilin primer comprising SEQ ID NO: 24. The resultant cyclophilin specific single-stranded cDNA in antisense orientation comprises a aminated cyclophilin-specific sequence at its 5' end and an
15 adapter sequence at its 3' end. The cyclophilin cDNA is purified using the method described in Example 8. The purified cyclophilin cDNA is amplified according to the method described in Example 10 using a cyclophilin-specific amplification primer comprising SEQ ID NO: 50 and an adapter amplification primer comprising SEQ ID NO: 17. Finally, the cyclophilin cDNA is directly sequenced according to the method
20 described in Example 11 using a cyclophilin sequencing primer comprising SEQ ID NO: 29.

Using the method described in Example 1, a capture modified cDNA_A library is constructed using mRNA isolated from a tissue source such as whole mouse brain and an aminated anchor primer comprising SEQ ID NO: 2. Using the method
25 described in Example 2, an adapted, aminated acDNA_A library is constructed by ligating a double-stranded adapter molecule comprising SEQ ID NO: 13 and 14 to the 5' end of the cDNA_A.

A biotinylated gene-specific primer corresponding to the M41 gene-specific fragment was synthesized (SEQ ID NO: 30). The M41 cDNA was generated in the
30 antisense orientation according to the method described in Example 6 using the biotinylated M41 primer comprising SEQ ID NO: 30. The resultant M41 cDNA comprised an biotinylated M41-specific sequence at its 5' end and an adapter sequence at its 3' end. The M41 cDNA was purified using the method described in Example 7.

- 75 -

The purified M41 cDNA was amplified according to the method described in Example 10 using a M41-specific amplification primer comprising SEQ ID NO: 32 and an adapter amplification primer comprising SEQ ID NO: 18. Finally, the M41 cDNA was directly sequenced according to the method described in Example 11 using
5 an M41sequencing primer comprising SEQ ID NO: 33.

Using the method described in Example 1, a capture modified cDNA_B library was constructed using mRNA isolated from whole brain tissue and a biotinylated anchor primer comprising SEQ ID NO: 1. Using the method described in Example 2, an adapted acDNA_B library was constructed by ligating an adapter molecule
10 comprising SEQ ID NO: 13 to the 5' end of the cDNA_B. The acDNA_B library was amplified according to the method described in Example 3 using an anchor amplification primer comprising SEQ ID NO: 20 and an adapter amplification primer comprising SEQ ID NO: 17.

A gene-specific fragment of 295 nucleotides (DST M51) was generated by the
15 TOGA differential display system (U.S. Patent 5,459,037). In order to obtain sequence information adjacent to the known DNA fragment M51, the sequence was extended towards the 5' end of the gene in the antisense orientation using the aminated solid phase approach described in Example 6, 8, 10 and 11. An aminated M51 primer (SEQ. ID NO: 36) was synthesized, reverse complementary to 26
20 nucleotides of DST 51. The primer was directed in the antisense orientation and located 135 nucleotides downstream of the 5' end of the DST M51 fragment. The M51-specific sequence was extended using a biotinylated amplified adapted cDNA library (aacDNA_B) derived from mouse brain poly A+ selected mRNA using the method described in Example 6. The aminated gene-specific single strand (M51-
25 gspEXT_A) contained 135 nucleotides of the known DST M51 DNA fragment sequence at its 5' end and the adapter sequence (SEQ ID NO: 13) at its 3' end. The M51-gspEXT_A was isolated from the contaminating cDNA library (aacDNA_B) using the two consecutive capture protocols described in Example 8, resulting in the covalent attachment of M51-gspEXT_A to the wall of a N-oxysuccinimide surface
30 (NOS) PCR plate well (Costar DNA-bind Thermowell M, # 6573). The solid phase attached M51-gspEXT_A sequence was amplified using the M51-specific primer of SEQ ID NO: 37 (antisense orientation) and the adapter amplification primer of SEQ ID. NO: 17 (sense orientation) with the protocol described in Example 10. The

- 76 -

resultant PCR product, M51-PCR1, was directly sequenced without subcloning using the sequencing primer of SEQ ID NO: 38. The sequence analysis revealed that M51-PCR1 contained 111 basepairs of DST M51 (5' sense orientation) and was extended by 360 nucleotides of novel sequence. The protocols in Examples 6, 8, 10 and 11

5 were performed in a two day time frame, including the sequence analysis.

Based on the newly extended M51 sequence, a new sequencing primer (SEQ ID NO: 40) located 120 nucleotides 5' of the DST M51 sequence was synthesized. A new direct sequencing reaction was performed using the M51-specific PCR product (M51-PCR1) with the sequencing primer of SEQ ID NO: 40. As a result of the
10 second direct sequencing reaction, the sequence of M51 was further extended by 406 nucleotides, elongating the DST M51 sequence of 295 by 765 nucleotides within 4 working days. The M51-specific PCR product M51-PCR1 can be used as template for 10 independent direct sequencing reactions using the protocol of Example 11.

In addition, new M51 gene-specific PCR products can be regenerated by using
15 the solid phase attached gene-specific M51 extended sequence M51-gspEXT_A as template. The NOS-coated PCR plate containing M51-gspEXT_A in one of the 96 wells can be stored in -20 degrees Celsius freezer. PCR plates containing covalently attached gene-specific first strands can be either stored at 4 degrees Celsius in PBS or lyophilized at -20 degrees Celsius for months without damage to the attached DNA.

20 For example, gene-specific DNA has been covalently attached and stored for 13 months without noticeable damage to the DNA. The PCR can be performed using the primers of SEQ ID NO: 37 (antisense orientation) and SEQ ID NO: 17 (sense orientation) as described above, creating the PCR product M51-PCR1. Alternatively, a new M51-specific PCR product can be synthesized located further upstream in the
25 novel extended sequence using the M51-specific primer of SEQ ID NO: 41 (anti sense

- 77 -

sequence was further extended by direct sequencing (Example 11), using a sequencing primer of SEQ ID NO: 43, located 695 nucleotides upstream of the DST M51 within the novel sequence generated using sequencing primer SEQ ID NO: 42. The direct sequencing generated 420 nucleotides of sequence overlapping and verifying the sequence generated with SEQ ID NO: 42 and extended the sequence of M51 by 174 nucleotides. The total time consumption for the primer walk using 4 sequencing primers, including the synthesis of the new sequence primers and the sequence analysis, was one week. During this time, the M51 sequence was extended by 1341 nucleotides of novel sequence without the need of subcloning the extended sequences.

10

Table 2													
		Annealing Temp 1						Annealing Temp 2					
		DST 1	DST 2	DST 3	DST 4	DST 5	DST 6	DST 1	DST 2	DST 3	DST 4	DST 5	DST 6
seq. primer 1	Template conc. 1	M1	M1	M1	M1	M1	M1	M1	M1	M1	M1	M1	M1
		M2	M2	M2	M2	M2	M2	M2	M2	M2	M2	M2	M2
seq. primer 2	Template conc. 2	M1	M1	M1	M1	M1	M1	M1	M1	M1	M1	M1	M1
		M2	M2	M2	M2	M2	M2	M2	M2	M2	M2	M2	M2
seq. primer 1	Template conc. 2	M1	M1	M1	M1	M1	M1	M1	M1	M1	M1	M1	M1
		M2	M2	M2	M2	M2	M2	M2	M2	M2	M2	M2	M2
seq. Primer 2	Template conc. 2	M1	M1	M1	M1	M1	M1	M1	M1	M1	M1	M1	M1
		M2	M2	M2	M2	M2	M2	M2	M2	M2	M2	M2	M2

All of the references cited herein, including patents, published patent applications and publications, are hereby incorporated in their entireties by reference.

While the invention has been described with an emphasis upon preferred aspects of the invention, it will be readily apparent to those of ordinary skill in the art that variations of the preferred embodiments can be used and that it is intended that the invention can be practiced otherwise than is specifically described herein.

Accordingly, the present invention includes all modifications encompassed within the spirit and scope of the invention as defined by the following claims.

- 78 -

We claim:

1. A method for amplifying a gene-specific polynucleotide comprising the steps of:
 - synthesizing a population of double stranded DNA molecules containing a
5 gene-specific polynucleotide and having a 3' end and a 5' end using an anchor primer having a first captureable moiety, wherein the anchor primer is at the 3' end of each double stranded DNA molecule;
 - ligating a double stranded adapter molecule to the 5' end of each double
10 stranded DNA molecule producing double stranded DNA molecules comprising both an anchor primer and an adapter molecule;
 - synthesizing a single stranded gene-specific polynucleotide using a gene-specific oligonucleotide primer having a second captureable moiety, wherein the captureable moiety of the gene-specific oligonucleotide primer is different from the captureable moiety of the anchor primer;
 - 15 isolating the single stranded gene-specific polynucleotide; and
amplifying the isolated gene-specific polynucleotide using both a gene-specific primer and primer that hybridizes to either a sequence located in the anchor primer or a sequence located in the adapter molecule.
2. The method of claim 1, comprising the additional step of sequencing the
20 amplified isolated gene-specific polynucleotide.
3. The method of claim 1, wherein the anchor primer comprises the sequence 5'-N'(N)_a(T)_b-3' where N' is a nucleotide to which a captureable moiety is attached, a is an integer of from about 15 to about 50, and b is an integer of from about 12 to about
25 25.
4. The method of claim 3, wherein N' is a biotinylated nucleotide.
5. The method of claim 3, wherein N' is an aminated nucleotide.
6. The method of claim 4, wherein the anchor primer comprises the sequence 5'-NAATTCAACTGGAAGCGGCCGCAGGAA(T)₁₈-3' where N is a biotinylated
guanidylate nucleotide (SEQ ID NO: 1).
- 30 7. The method of claim 5, wherein the anchor primer comprises the sequence 5'-NAATTCAACTGGAAGCGGCCGCAGGAA(T)₁₈-3' where N is an aminated
guanidylate nucleotide (SEQ ID NO: 2).

- 79 -

8. The method of claim 1, wherein the double-stranded adapter molecule comprises the sequence of at least one rare restriction enzyme site wherein the rare restriction enzyme site occurs in a mammalian genome less than of about one site per 15,000 nucleotides.
- 5 9. The method of claim 8, wherein the double-stranded adapter molecule comprises the sequence of two or more rare restriction enzyme sites which are sequentially arranged, selected from the group consisting of NotI, AscI, BaeI, FseI, SfiI, Sgfl, SgrAI, SrfI, PacI, PmeI, PpuMI, RsrII, SapI, SexAI, Sse8387I, SbfI, SpeI, Sall, RsrI, and NheI restriction enzyme sites.
- 10 10. The method of claim 9, wherein the double-stranded adapter molecule comprises the sequence of contiguous SbfI, SpeI, Sall, RsrI, and NheI restriction enzyme sites.
11. The method of claim 1, wherein the double-stranded adapter molecule comprises an RNA polymerase promoter sequence selected from the group consisting
- 15 of T3, SP6, and T7 RNA polymerase promoter sequences.
12. The method of claim 11, wherein the double-stranded adapter molecule comprises the T7 polymerase promoter sequence.
13. The method of claim 12, wherein the double-stranded adapter molecule comprises the sequence of 5' - ATAGCCTGCAGGTAATACGACTCACTA
- 20 TAGGGACTAGTCGACGGACCGCTAGCATCAGATC - 3' (SEQ ID NO: 13).
14. The method of claim 12, wherein the double stranded adapter molecule comprises SEQ ID NO: 15.
15. The method of claim 1, comprising the additional step of amplifying the double stranded DNA molecules comprising both an anchor primer and an adapter
- 25 molecule before the step of synthesizing the single stranded gene-specific polynucleotide, wherein the double stranded DNA molecules are amplified using a 3' end primer that hybridizes to a sequence located in the anchor primer and a 5' end primer that hybridizes to a sequence located in the adapter molecule.
16. The method of claim 15, wherein the 3' end amplification primer comprises
- 30 the sequence 5'- NAATTCAACTGGAAGCGGCCGCAGGA -3' (SEQ ID NO: 20 or 21) and the 5' end amplification primer comprises the sequence of 5' - CCTGCAGGTAATACGACTCACTATAGG - 3' (SEQ ID NO: 17).

- 80 -

17. The method of claim 15 wherein the 5' end amplification primer comprises the sequence of SEQ ID NO: 18.
18. The method of claim 1, comprising the additional steps of:
synthesizing a population of single-stranded RNA molecules from the
5 population of double stranded DNA molecules comprising an anchor primer and an adapter molecule, and
digesting the double stranded DNA template molecules, before the step of synthesizing the single stranded gene-specific polynucleotide.
19. The method of claim 1, wherein the gene specific primer comprises the
10 sequence 5'- N'(N)_c-3' where N' is a nucleotide to which a captureable moiety is attached and c is an integer of from about 17 to about 40.
20. The method of claim 19 wherein N is a biotinylated nucleotide.
21. The method of claim 19, wherein N is an aminated nucleotide.
22. The method of claim 1 further comprising affixing the first captureable moiety
15 of the anchor primer of the double-stranded DNA molecule following synthesis of the single-stranded gene-specific polynucleotide, wherein said first captureable moiety of the anchor primer is affixed to a substrate comprising a coating selected from the group consisting of streptavidin, avidin, neutravidin, N-oxysuccinimide ester, dimethyladipimide-2-HCl, dimethylpimelimide-HCl, dimethylsuberimide-2HCl,
20 dimethyl 3,3'-dithiobispropionimide-2HCl, disuccinimidyl glutarate, disuccinimidyl suberate, bis(sulfosuccinimidyl)suberate, dithiobis(succinimidyl propionate), ethylene glycobis(succinimidylsuccinate), ethylene glycobis(sulfosuccinimidylsuccinate), disuccinimidyl tartarate, disulfosuccinimidyl tartarate, bis[2-succinimidylloxycarbonyloxy)ethyl]sulfone, bis[2-(sulfosuccinimidylloxycarbonyloxy)
25 ethyl]sulfone, and N-hydroxysuccinimidyl 2,3-dibromopropionate, succinimidyl 4-(N-maleimidomethyl) cyclohexane-1-carboxylate, sulfosuccinimidyl 4-(N-maleimidomethyl) cyclohexane-1-carboxylate, m-maleimidobenzoyl-N-hydroxysuccinimide ester, m-maleimidobenzoyl-N-hydroxysulfosuccinimide ester, succinimidyl 4-(p-maleimidophenyl)-butyrate, sulfosuccinimidyl 4-(p-
30 maleimidophenyl)-butyrate, bismaleimido-hexane, N-(γ-maleimidobutyryloxy)succinimide ester, N-(γ-maleimidobutyryloxy)sulfosuccinimide ester, N-succinimidyl(4-iodoacetyl)aminobenzoate, sulfosuccinimidyl(4-iodoacetyl)aminobenzoate, 1,4-Di-[3'-2'-pyridyl]dithio(propionamido)butane], 4-

- 81 -

succinimidylloxycarbonyl- α -(2-pyridyl-dithio)toluene, sulfosuccinimidyl-6-[α -
 methyl- α -(2-pyridyldithio)-toluamido]hexane, N-succinimidyl-3-(2-pyridyldithio)-
 propionate, succinimidyl 6-[3-(2-pyridyldithio)propion-amido]hexanoate,
 sulfosuccinimidyl 6-[3-(2-pyridyldithio)propionamido]hexanoate, 1-ethyl-3-(3-
 5 dimethylaminopropyl)carbodiimide hydrochloride, N,N'-dicyclohexylcarbodiimide,
 1,5-difluoro-2,4-dinitrobenzene, N-5-azido-2-nitrobenzoyloxysuccinimide, N-
 hydroxysuccinimidyl-4-azidobenzoate, N-hydroxysulfosuccinimidyl-4-azidobenzoate,
 N-hydroxysuccinimidyl-4-azidosalicylic acid, N-hydroxysulfosuccinimidyl-4-
 azidosalicylic acid, sulfosuccinimidyl-(4-azidosalicylamido)-hexanoate, p-
 10 nitrophenyl-2-diazo-3,3,3-trifluoropropionate, 2-diazo-3,3,3-
 trifluoropropionylchloride, N-succinimidyl-(4-azidophenyl)1,3'-dithiopropionate,
 sulfosuccinimidyl-(4-azidophenyldithio)propionate, sulfosuccinimidyl 2-(7-azido-4-
 methylcoumarin-3-acetamide) ethyl-1,3'-dithiopropionate, sulfosuccinimidyl 7-azido-
 4-methylcoumarin-3-acetate, sulfosuccinimidyl 2(m-azido-o-nitrobenzamido)-ethyl-
 15 1,3'-dithiopropionate, N-succinimidyl-6-(4'-azido-2'-nitrophenylamino)hexanoate,
 sulfosuccinimidyl-6-(4'-azido-2'-nitrophenylamino)hexanoate, sulfosuccinimidyl 2-(p-
 azidosalicylamido)ethyl-1,3'-dithiopropionate, and sulfosuccinimidyl 4-(p-
 azidophenyl)butyrate and mixtures thereof.

23. The method of claim 22, wherein the first captureable moiety of the anchor
 20 primer of the double stranded DNA molecule is affixed to a substrate comprising a
 coating of streptavidin.

24. The method of claim 22, wherein the first captureable moiety of the anchor
 primer of the double stranded DNA molecule is affixed to a substrate comprising a
 coating of N-oxysuccinimide ester.

25. The method of claim 1, further comprising affixing the second captureable
 moiety of the gene-specific primer of the gene-specific single stranded polynucleotide
 following synthesis of the single-stranded gene-specific polynucleotide, wherein said
 second captureable moiety of the gene-specific primer is affixed to a substrate
 comprising a coating selected from the group consisting of streptavidin, avidin,
 30 neutravidin, N-oxysuccinimide ester, dimethyladipimide-2-HCl,
 dimethylpimelimide-2-HCl, dimethylsuberimide-2HCl, dimethyl 3,3'-
 dithiobispropionimide-2HCl, disuccinimidyl glutarate, disuccinimidyl suberate,
 bis(sulfosuccinimidyl)suberate, dithiobis(sulfosuccinimidyl propionate), ethylene

- 82 -

glycobis(succinimidylsuccinate), ethylene glycobis(sulfosuccinimidylsuccinate), disuccinimidyl tartarate, disulfosuccinimidyl tartarate, bis[2-succinimidylloxycarbonyloxy)ethyl]sulfone, bis[2-(sulfosuccinimidylloxycarbonyloxy)ethyl]sulfone, and N-hydroxysuccinimidyl 2,3-dibromopropionate, succinimidyl 4-

5 (N-maleimidomethyl) cyclohexane-1-carboxylate, sulfosuccinimidyl 4-(N-maleimidomethyl) cyclohexane-1-carboxylate, m-maleimidobenzoyl-N-hydroxysuccinimide ester, m-maleimidobenzoyl-N-hydroxysulfosuccinimide ester, succinimidyl 4-(p-maleimidophenyl)-butyrate, sulfosuccinimidyl 4-(p-maleimidophenyl)-butyrate, bismaleimidohehexane, N-(γ-

10 maleimidobutyryloxy)succinimide ester, N-(γ-maleimidobutyryloxy)sulfosuccinimide ester, N-succinimidyl(4-iodoacetyl)aminobenzoate, sulfosuccinimidyl(4-iodoacetyl)aminobenzoate, 1,4-Di-[3'-2'-pyridyldithio(propion-amido)butane], 4-succinimidylloxycarbonyl-α-(2-pyridyldithio)toluene, sulfosuccinimidyl-6-[α-methyl-α-(2-pyridyldithio)-toluamido]hexane, N-succinimidyl-3-(2-pyridyldithio)-

15 propionate, succinimidyl 6-[3-(2-pyridyldithio)propionamido]hexanoate, sulfosuccinimidyl 6-[3-(2-pyridyldithio)propionamido]hexanoate, 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride, N,N'-dicyclohexylcarbodiimide, 1,5-difluoro-2,4-dinitrobenzene, N-5-azido-2-nitrobenzoyloxysuccinimide, N-hydroxysuccinimidyl-4-azidobenzoate, N-hydroxy-sulfosuccinimidyl-4-

20 azidobenzoate, N-hydroxysuccinimidyl-4-azidosalicylic acid, N-hydroxysulfosuccinimidyl-4-azidosalicylic acid, sulfosuccinimidyl-(4-azidosalicylamido)-hexanoate, p-nitrophenyl-2-diazo-3,3,3-trifluoropropionate, 2-diazo-3,3,3-trifluoropropionyl-chloride, N-succinimidyl-(4-azidophenyl)1,3'-dithiopropionate, sulfosuccinimidyl-(4-azidophenyldithio)propionate,

25 sulfosuccinimidyl 2-(7-azido-4-methylcoumarin-3-acetamide) ethyl-1,3'-dithiopropionate, sulfosuccinimidyl 7-azido-4-methylcoumarin-3-acetate, sulfosuccinimidyl 2(m-azido-o-nitrobenzamido)-ethyl-1,3'-dithiopropionate, N-succinimidyl-6-(4'-azido-2'-nitrophenylamino)hexanoate, sulfosuccinimidyl-6-(4'-azido-2'-nitrophenylamino)hexanoate, sulfosuccinimidyl 2-(p-

30 azidosalicylamido)ethyl-1,3'-dithiopropionate, and sulfosuccinimidyl 4-(p-azidophenyl)butyrate and mixtures thereof.

- 83 -

26. The method of claim 25, wherein the second captureable moiety of the gene-specific primer of the gene-specific single stranded polynucleotide is affixed to a substrate comprising a coating of streptavidin.
27. The method of claim 25, wherein the second captureable moiety of the gene-specific primer of the gene-specific single stranded polynucleotide is affixed to a substrate comprising a coating of N-oxysuccinimide ester.
28. The method of claim 1, wherein the purified single stranded gene-specific polynucleotide is amplified using a 5' end gene-specific primer and a 3' end amplification primer, the 3' end amplification primer hybridizing to a portion of the sequence of the anchor primer.
29. The method of claim 28, wherein the 3' end amplification primer comprises SEQ ID NO: 23.
30. The method of claim 6, wherein the purified single stranded gene-specific polynucleotide is amplified using a 3' end gene-specific primer and a 3' end amplification primer, the 5' end amplification primer hybridizing to a portion of the sequence of the anchor primer.
31. The method of claim 30, wherein the 5' end amplification primer comprises SEQ ID NO: 17.
32. The method of claim 30, wherein the 5' end amplification primer comprises SEQ ID NO: 18.
33. A method for isolating a gene-specific polynucleotide comprising the steps of:
synthesizing a population of double stranded DNA molecules having a 3' end and a 5' end;
ligating a double stranded adapter molecule to the 5' end of each double stranded DNA molecule;
attaching at least one gene-specific oligonucleotide primer having a captureable moiety to a solid substrate, wherein the gene-specific primer is attached via the captureable moiety;
synthesizing at least one single stranded gene-specific polynucleotide using the attached gene-specific primer(s);
purifying the single-stranded gene-specific polynucleotide(s); and
amplifying the gene-specific polynucleotide(s) using a gene-specific primer and an primer that hybridizes to a sequence located in the adapter molecule.

- 84 -

34. The method of claim 33, wherein the double-stranded adapter molecule comprises the sequence of at least one rare restriction enzyme site wherein the rare restriction enzyme site occurs in a mammalian genome less than about one site per 15,000 nucleotides.
- 5 35. The method of claim 33, wherein the double-stranded adapter molecule comprises the sequence of SbfI, SpeI, SalI, RsrI, and NheI sequential contiguous restriction enzyme sites.
36. The method of claim 33, wherein the double-stranded adapter molecule comprises the T7 polymerase promoter sequence.
- 10 37. The method of claim 33, wherein the double stranded adapter molecule comprises the sequence of 5' –ATAGCCTGCAGGTAATACGACTCACTA TAGGGACTAGTCGACGGACCGCTAGCATCAGATC – 3' (SEQ ID NO: 13).
38. The method of claim 33, wherein the double-stranded adapter molecule comprises SEQ ID NO: 15.
- 15 39. The method of claim 33, wherein the gene specific primer comprises the sequence 5'- N'(N)_c -3' where N' is a nucleotide to which a captureable moiety is attached and c is an integer of from about 17 to about 40.
40. The method of claim 39 wherein N' is a biotinylated nucleotide.
41. The method of claim 39, wherein N' is an aminated nucleotide.
- 20 42. The method of claim 33, wherein the captureable moiety of the gene-specific primer is affixed to a substrate comprising a coating of streptavidin.
43. The method of claim 33, wherein the captureable moiety of the gene-specific primer is affixed to a substrate comprising a coating of N-oxysuccinimide ester.
44. The method of claim 33, wherein the purified single stranded gene-specific polynucleotide is amplified using a 5' end gene-specific primer and a 3' end amplification primer that hybridizes to a sequence located in the adapter molecule.
- 25 45. The method of claim 44, wherein the 3' end amplification primer comprises SEQ ID NO: 17.
46. The method of claim 44, wherein the 3' end amplification primer comprises
- 30 47. The method of claim 33, wherein the purified single stranded gene-specific polynucleotide is amplified using a 5' end gene-specific primer and a 3' end amplification primer that hybridizes to a sequence located in the adapter molecule.

- 85 -

48. The method of claim 47, wherein the 3' end amplification primer comprises SEQ ID NO:19.
49. The method of claim 1, comprising the additional step of cloning the isolated single-stranded gene-specific polynucleotide into a vector to form a recombinant
5 construct.
50. The recombinant construct produced by the method of claim 49.
51. The method of claim 49, wherein the recombinant construct is introduced into a host cell.
52. The host cell comprising the recombinant construct produced by the method of
10 claim 51.
53. An isolated polynucleotide having an attached captureable moiety chosen from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:45 and SEQ ID NO:46.
54. An isolated double stranded polynucleotide chosen from the group consisting
15 of
- SEQ ID NO: 13 and complementary sequence SEQ ID NO: 14, and
SEQ ID NO: 15 and complementary sequence SEQ ID NO: 16.
55. An isolated polynucleotide comprising the sequence chosen from the group consisting of SEQ ID NO: 17-19.
- 20 56. An anchor amplification primer oligonucleotide having an attached biotin moiety comprising the sequence consisting of SEQ ID NO: 20.
57. An anchor amplification primer oligonucleotide having an attached primary amine comprising the sequence consisting of SEQ ID NO: 21.
58. An anchor amplification primer oligonucleotide comprising the sequence
25 consisting of SEQ ID NO: 23.
59. A composition comprising the oligonucleotide of claim 58 and an aqueous carrier.
60. An article of manufacture comprising the oligonucleotide of claim 53 and a streptavidin-coated substrate.
- 30 61. An article of manufacture comprising the oligonucleotide of claim 54 and a NOS-coated substrate.
62. A kit suitable for isolating and sequencing a gene-specific polynucleotide from a population of DNA or RNA polynucleotides comprising at least one single stranded

- 86 -

anchor primer oligonucleotide having an attached captureable moiety and at least one double stranded adapter oligonucleotide in an amount sufficient for at least one assay and suitable containers.

63. The kit of claim 62 further comprising at least one single stranded
5 amplification anchor primer oligonucleotide and at least one single stranded amplification adapter primer oligonucleotide.

64. The kit of claim 62 further comprising at least one single stranded gene-specific primer oligonucleotide having a captureable moiety, wherein the captureable moiety of the anchor primer oligonucleotide is different from the captureable moiety
10 of the gene-specific primer oligonucleotide.

65. The kit of claim 62 wherein the gene-specific primer oligonucleotide is attached to a substrate comprising a coating.

66. The kit of claim 65 wherein the substrate comprising a coating selected from the group consisting of streptavidin, avidin, neutravidin, N-oxysuccinimide ester,
15 dimethyladipimide-2-HCl, dimethylpimelimide-HCl, dimethylsuberimide-2HCl, dimethyl 3,3'-dithiobispropionimide-2HCl, disuccinimidyl glutarate, disuccinimidyl suberate, bis(sulfosuccinimidyl)suberate, dithiobis(succinimidyl propionate), dithiobis(sulfosuccinimidyl propionate), ethylene glycobis(succinimidylsuccinate), ethylene glycobis(sulfosuccinimidylsuccinate),
20 disuccinimidyl tartarate, disulfosuccinimidyl tartarate, bis[2-succinimidylloxycarbonyloxy)ethyl]sulfone, bis[2-(sulfosuccinimidylloxycarbonyloxy)ethyl]sulfone, and N-hydroxysuccinimidyl 2,3-dibromopropionate, succinimidyl 4-(N-maleimidomethyl) cyclohexane-1-carboxylate, sulfosuccinimidyl 4-(N-maleimidomethyl) cyclohexane-1-carboxylate, m-maleimidobenzoyl-N-hydroxysuccinimide ester, m-maleimidobenzoyl-N-hydroxysulfosuccinimide ester,
25 succinimidyl 4-(p-maleimidophenyl)-butyrate, sulfosuccinimidyl 4-(p-maleimidophenyl)-butyrate, bismaleimido-hexane, N-(γ -maleimidobutyryloxy)succinimide ester, N-(γ -maleimidobutyryloxy)sulfosuccinimide ester, N-succinimidyl(4-iodoacetyl)aminobenzoate, sulfosuccinimidyl(4-iodoacetyl)aminobenzoate, 1,4-Di-[3'-2'-pyridyldithio(propionamido)butane], 4-succinimidylloxycarbonyl- α -(2-pyridyldithio)toluene, sulfosuccinimidyl-6-[α -methyl- α -(2-pyridyldithio)-toluamido]hexane, N-succinimidyl-3-(2-pyridyldithio)-propionate, succinimidyl 6-[3-(2-pyridyldithio)propionamido]hexanoate,

- 87 -

sulfosuccinimidyl 6-[3-(2-pyridyldithio)propionamido]hexanoate, 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride, N,N'-dicyclohexylcarbodiimide, 1,5-difluoro-2,4-dinitrobenzene, N-5-azido-2-nitrobenzoyloxysuccinimide, N-hydroxysuccinimidyl-4-azidobenzoate, N-hydroxysulfosuccinimidyl-4-azidobenzoate, N-hydroxysuccinimidyl-4-azidosalicylic acid, N-hydroxysulfosuccinimidyl-4-azidosalicylic acid, sulfosuccinimidyl-(4-azidosalicylamido)-hexanoate, p-nitrophenyl-2-diazo-3,3,3-trifluoropropionate, 2-diazo-3,3,3-trifluoropropionylchloride, N-succinimidyl-(4-azidophenyl)1,3'-dithiopropionate, sulfosuccinimidyl-(4-azidophenyldithio)propionate, sulfosuccinimidyl 2-(7-azido-4-methylcoumarin-3-acetamide) ethyl-1,3'-dithiopropionate, sulfosuccinimidyl 7-azido-4-methylcoumarin-3-acetate, sulfosuccinimidyl 2(m-azido-o-nitrobenzamido)-ethyl-1,3'-dithiopropionate, N-succinimidyl-6-(4'-azido-2'-nitrophenylamino)hexanoate, sulfosuccinimidyl-6-(4'-azido-2'-nitrophenylamino)hexanoate, sulfosuccinimidyl 2-(p-azidosalicylamido)ethyl-1,3'-dithiopropionate, and sulfosuccinimidyl 4-(p-azidophenyl)butyrate and mixtures thereof.

67. The kit of claim 62 wherein the substrate is a multiwell plate.

68. An isolated polynucleotide selected from the group consisting of:

NAATTCAACTGGAAGCGGCCGCAGGAATTTTTTTTTTTTTTTTTT

(SEQ ID NO: 1&2);

20 NAATTCAACTGGAACCTGCAGGAGGAATTTTTTTTTTTTTTTTTT

(SEQ ID NO: 3&4);

NAATTCAACTGGAACTAGTAGGAATTTTTTTTTTTTTTTTTT (SEQ ID NO: 5&6);

25 NAATTCAACTGGAAGTCGACAGGAATTTTTTTTTTTTTTTTTT (SEQ ID NO: 7&8);

NAATTCAACTGGAACGGATCCGAGGAATTTTTTTTTTTTTTTTTT (SEQ ID NO: 9&10); and

NAATTCAACTGGAAGCTAGCAGGAATTTTTTTTTTTTTTTTTT (SEQ ID NO: 11&12),

30 where N is chosen from A, G, C, and T, and the N is covalently bound to a captureable moiety chosen from the group consisting of biotin, primary amines, avidin, streptavidin, neutravidin, primary carboxylates, thiol alcohols, thiol carboxylates, carbonyls, sugars, lipids, and peptides.

- 88 -

69. A method for determining without cloning an extended sequence of a gene-specific polynucleotide that has a partial gene-specific sequence known, comprising the steps of:

5 synthesizing a population of double stranded DNA molecules having a 3' end and a 5' end using an anchor primer having a first captureable moiety, wherein the anchor primer is at the 3' end of each double stranded DNA molecule;

ligating a double stranded adapter molecule to the 5' end of each double stranded DNA molecule producing double stranded DNA molecules comprising both an anchor primer and an adapter molecule;

10 synthesizing a single stranded gene-specific polynucleotide using a gene-specific oligonucleotide primer having a second captureable moiety, wherein the captureable moiety of the gene-specific oligonucleotide primer is different from the captureable moiety of the anchor primer;

isolating the single stranded gene-specific polynucleotide; and

15 amplifying the isolated gene-specific polynucleotide using both a gene-specific primer that hybridizes to either a sequence located in the anchor primer or a sequence located in the adapter molecule; and

sequencing the amplified isolated gene-specific polynucleotide without cloning, thereby determining the sequence of a polynucleotide.

20

1/10

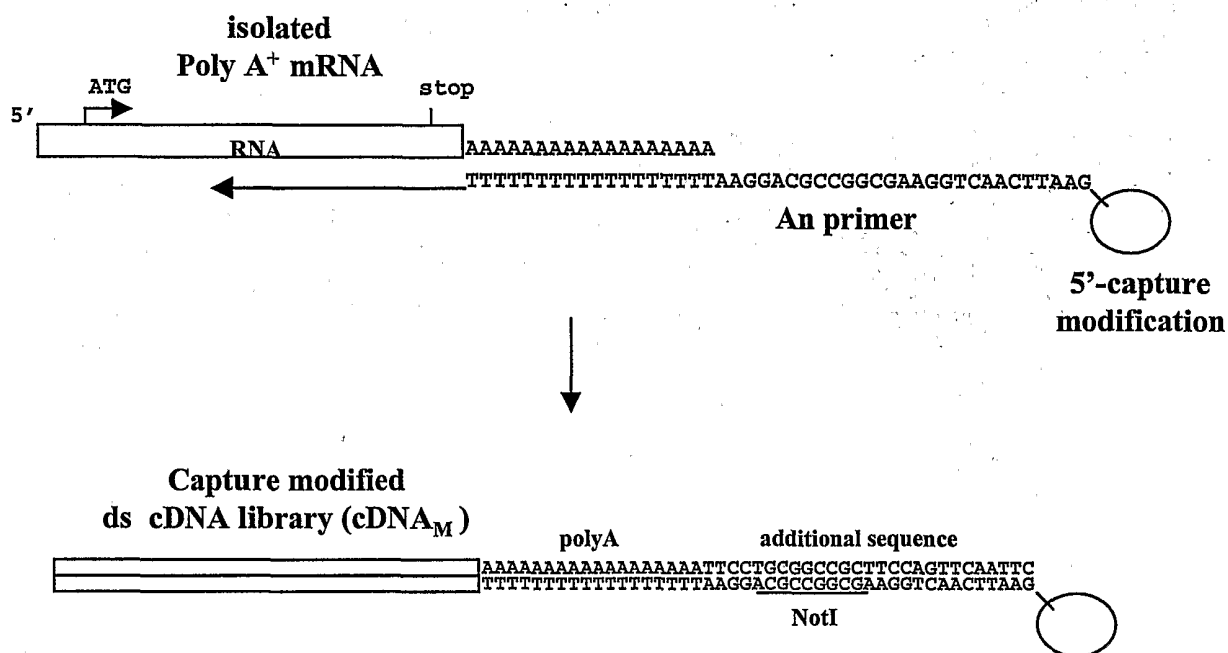


Fig. 1

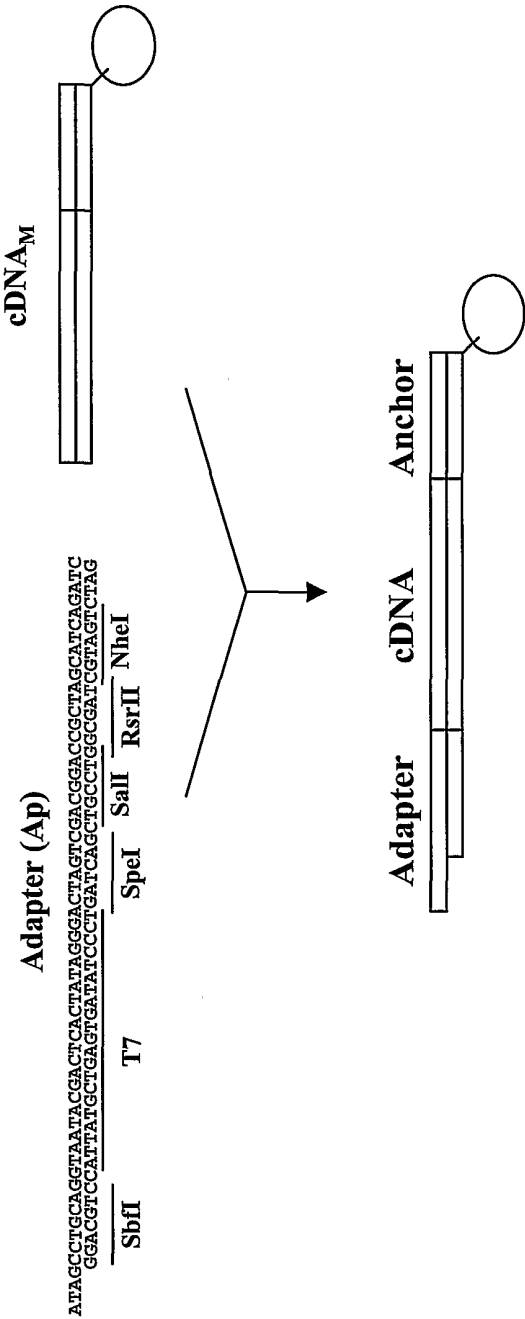


Fig. 2

3/10

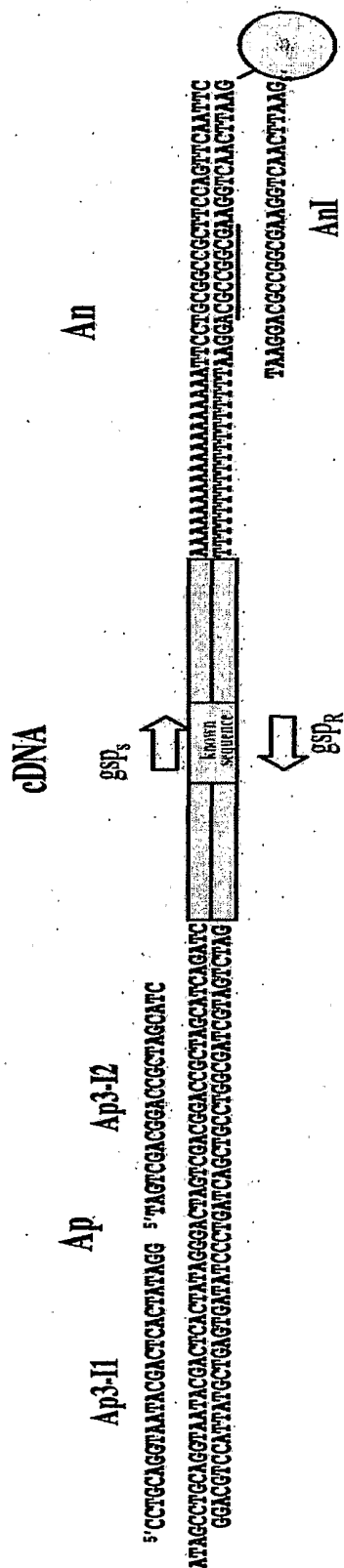


Fig. 3

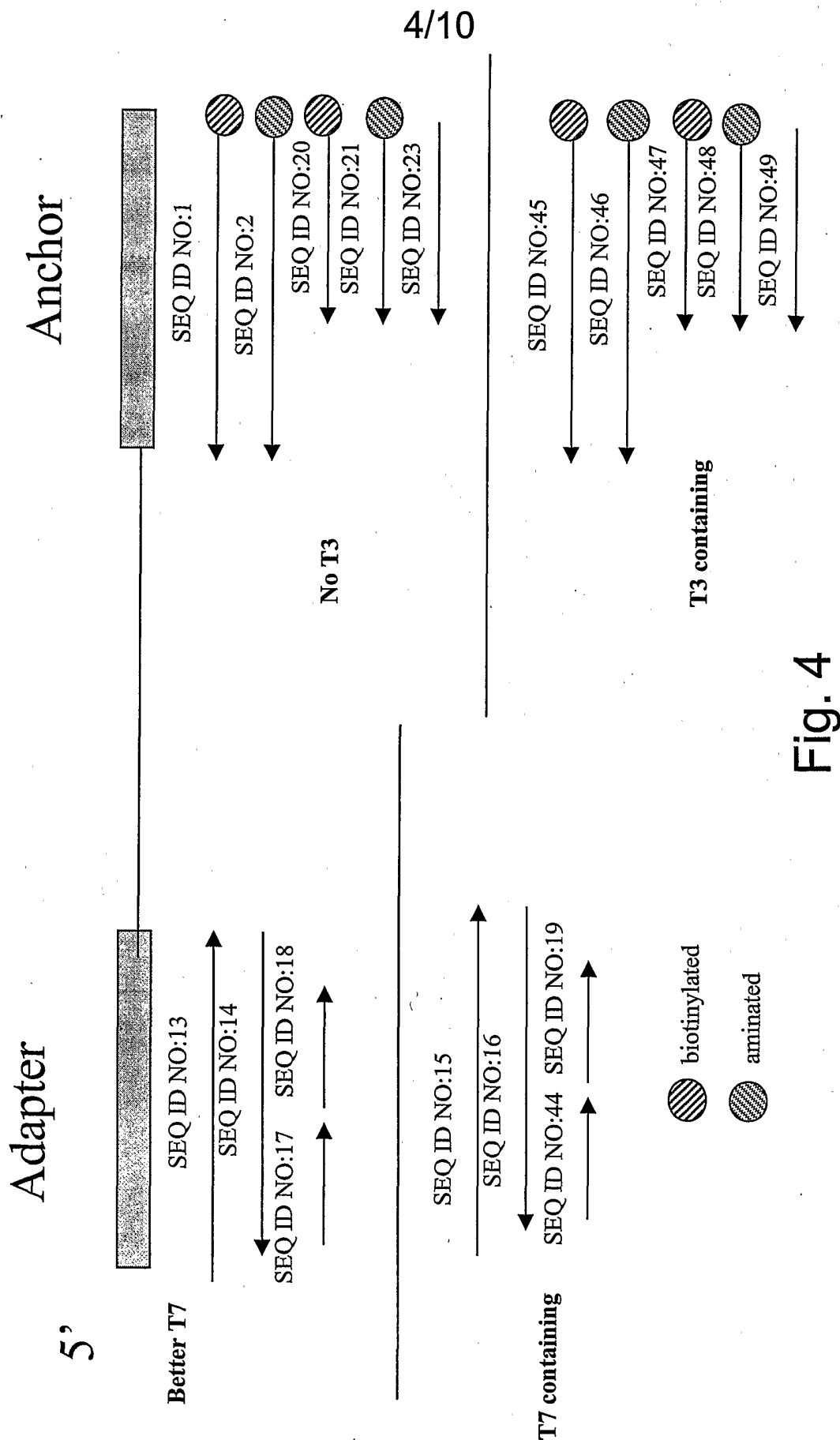


Fig. 4

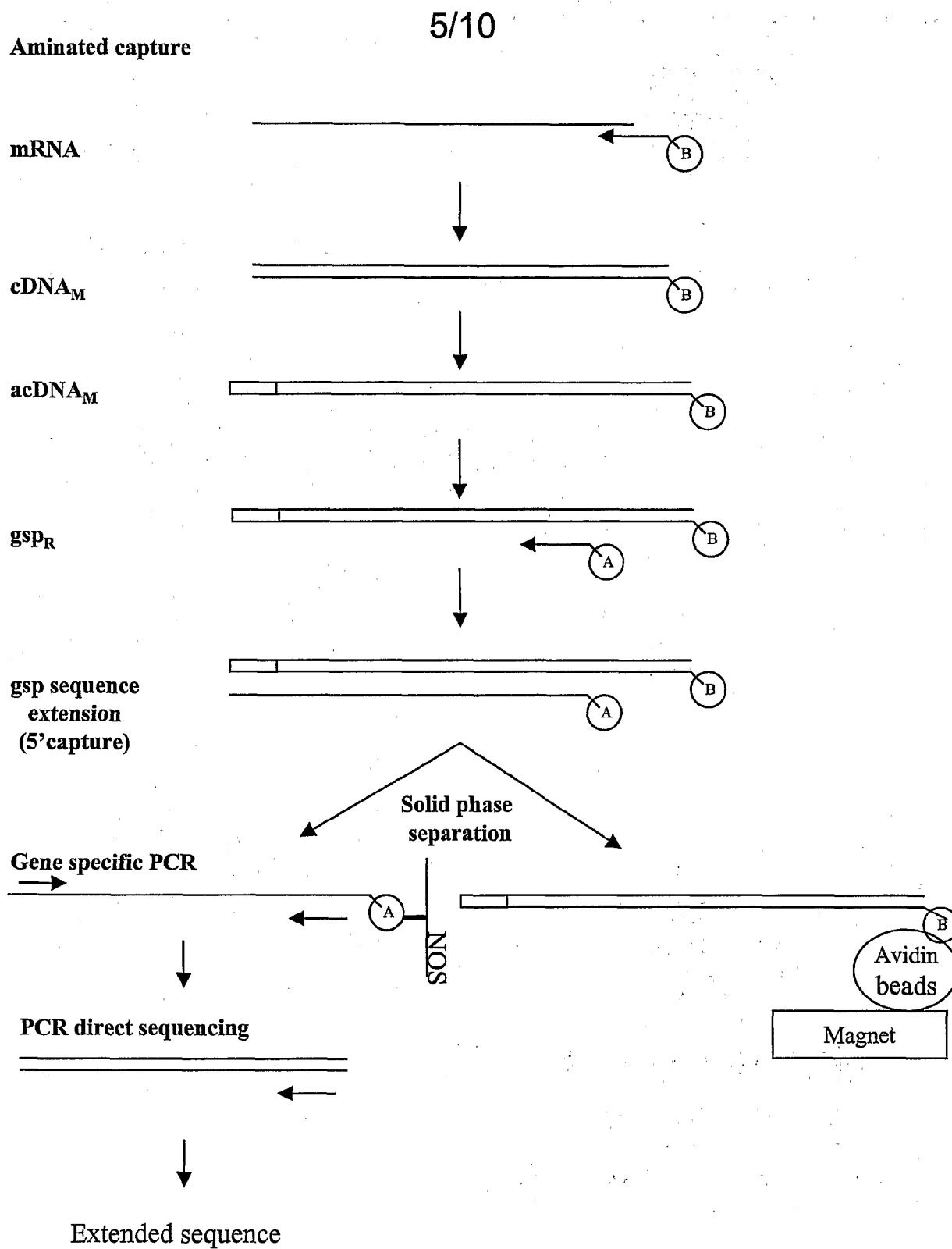


Fig. 5

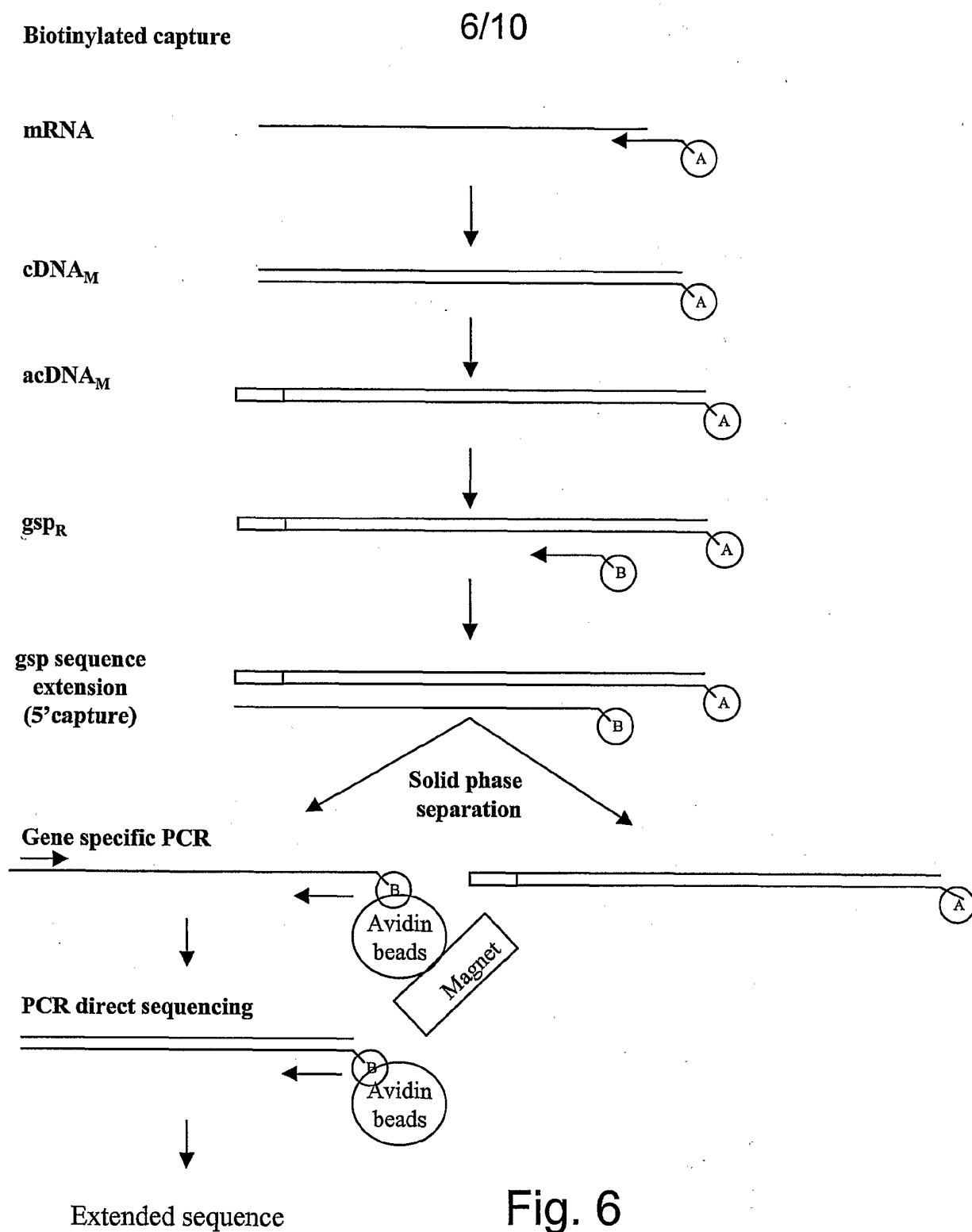


Fig. 6

7/10

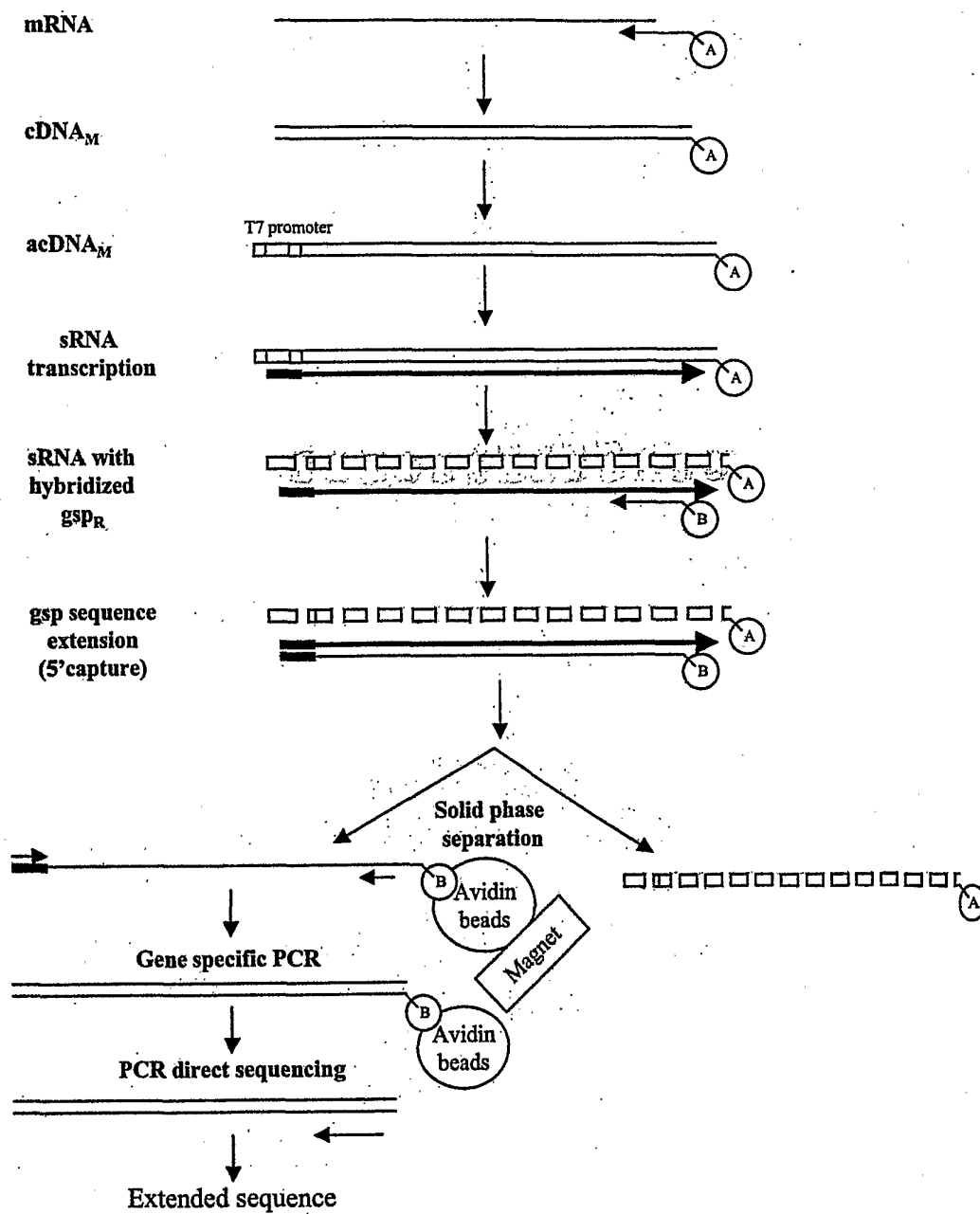


Fig. 7

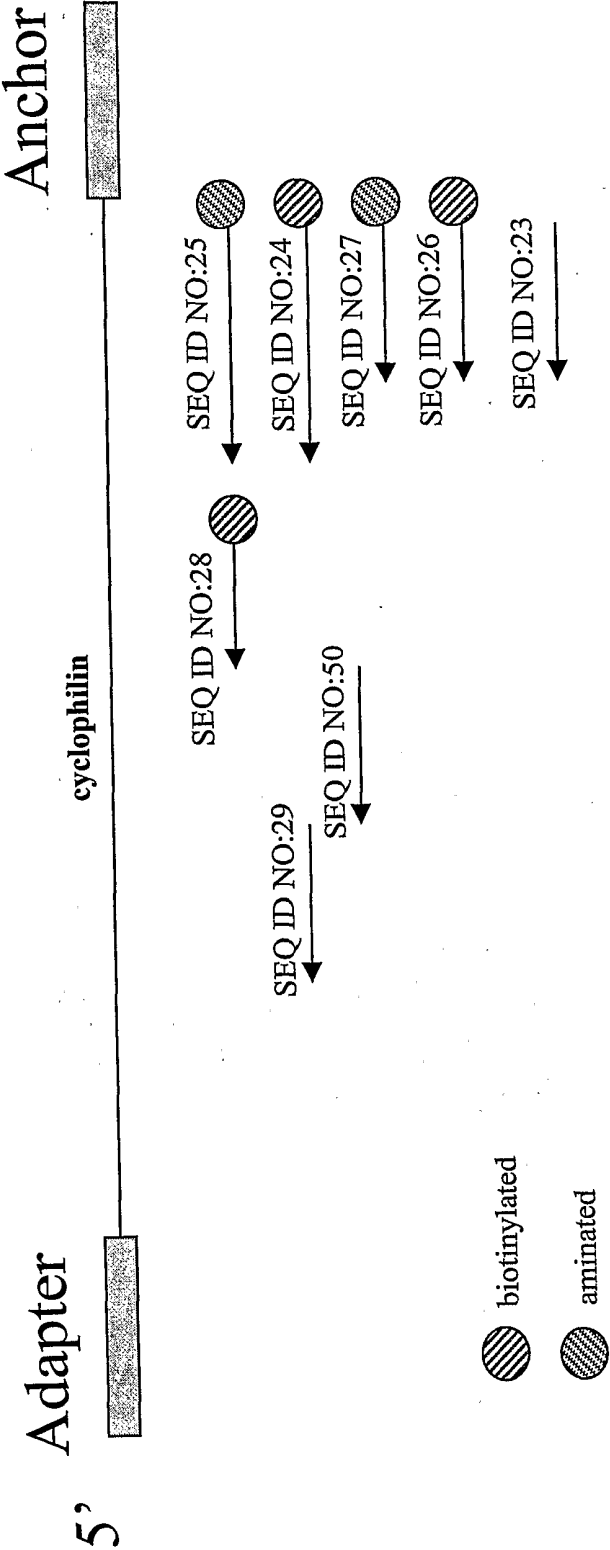


Fig. 8

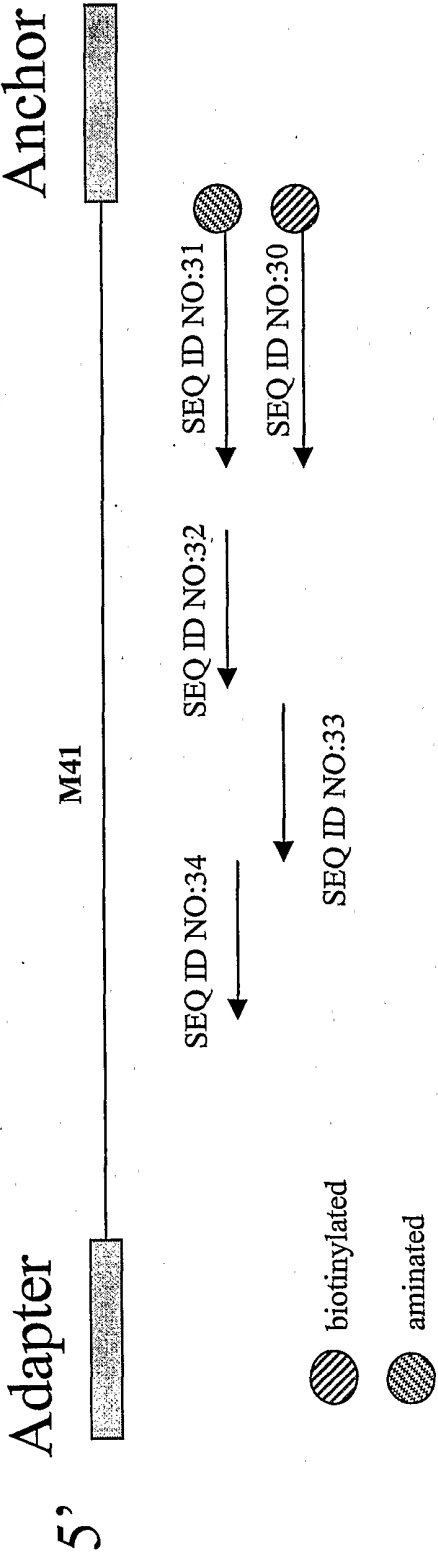


Fig. 9

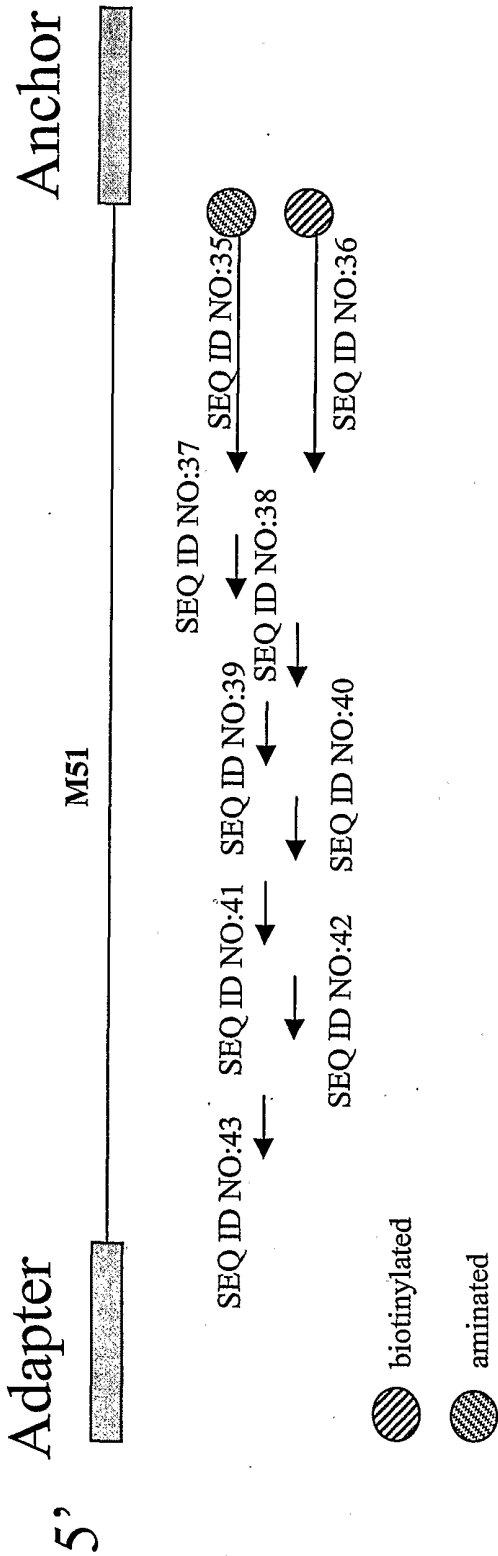


Fig. 10

- 1 -

SEQUENCE LISTING

<110> Müller, Rolf
Riddle, Gretchen
Glass, James

<120> Methods for Rapid Isolation and Sequence Determination of
Gene-Specific Sequences

<130> 99,023-A

<150> US 09/560,845

<151> 2000-04-28

<160> 50

<170> PatentIn version 3.0

<210> 1

<211> 45

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> ()..()

<223> Biotinylated anchor primer in which n is a biotinylated
guanine

<220>

<221> modified_base

<222> (1)..(1)

<223> n is a biotinylated guanine

<400> 1

naattcaact ggaagcggcc gcaggaattt tttttttttt ttttt
45

<210> 2

<211> 45

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> ()..()

<223> Aminated anchor primer in which n is an aminated guanine

<220>

<221> modified_base

<222> (1)..(1)

<223> n is an aminated guanine

<400> 2

naattcaact ggaagcggcc gcaggaattt tttttttttt ttttt

45

- 2 -

<210> 3
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Biotinylated anchor primer in which n is a biotinylated
guanine

<220>
<221> modified_base
<222> (1)..(1)
<223> n is a biotinylated guanine

<400> 3
naattcaact ggaacctgca ggaggaattt tttttttttt ttttt

45

<210> 4
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Aminated anchor primer in which n is an aminated guanine

<220>
<221> modified_base
<222> (1)..(1)
<223> n is an aminated guanine

<400> 4
naattcaact ggaacctgca ggaggaattt tttttttttt ttttt

45

<210> 5
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Biotinylated anchor primer in which n is a biotinylated
guanine

<220>
<221> modified_base
<222> (1)..(1)
<223> n is a biotinylated guanine

- 3 -

<400> 5
naattcaact ggaaactagt aggaattttt tttttttttt ttt 43

<210> 6
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Aminated anchor primer in which n is an aminated guanine

<220>
<221> modified_base
<222> (1)..(1)
<223> n is an aminated guanine

<400> 6
naattcaact ggaaactagt aggaattttt tttttttttt ttt 43

<210> 7
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Biotinylated anchor primer in which n is a biotinylated guanine

<220>
<221> modified_base
<222> (1)..(1)
<223> n is a biotinylated guanine

<400> 7
naattcaact ggaagtcgac aggaattttt tttttttttt ttt 43

<210> 8
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Aminated anchor primer in which n is an aminated guanine

<220>
<221> modified_base

- 4 -

<222> (1)..(1)
<223> n is an aminated guanine

<400> 8
naattcaact ggaagtcgac aggaattttt tttttttttt ttt 43

<210> 9
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Biotinylated anchor primer in which n is a biotinylated guanine

<220>
<221> modified_base
<222> (1)..(1)
<223> n is a biotinylated guanine

<400> 9
naattcaact ggaacggatc cgaggaattt tttttttttt ttttt 45

<210> 10
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Aminated anchor primer in which n is an aminated guanine

<220>
<221> modified_base
<222> (1)..(1)
<223> n is an aminated guanine

<400> 10
naattcaact ggaacggatc cgaggaattt tttttttttt ttttt 45

<210> 11
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Biotinylated anchor primer in which n is a biotinylated guanine

- 5 -

<220>
<221> modified_base
<222> (1)..(1)
<223> n is a biotinylated guanine

<400> 11
naattcaact ggaagctagc aggaattttt tttttttttt ttt 43

<210> 12
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Aminated anchor primer in which n is an aminated guanine

<220>
<221> modified_base
<222> (1)..(1)
<223> n is an aminated guanine

<400> 12
naattcaact ggaagctagc aggaattttt tttttttttt ttt 43

<210> 13
<211> 61
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Adapter molecule - sense strand

<400> 13
atagcctgca ggtaatacga ctactatag ggactagtcg acggaccgct agcatcagat 60
c 61

<210> 14
<211> 57
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Adapter molecule - antisense strand

<400> 14

- 6 -

gatctgatgc tagcgggtccg tcgactagtc cctatagtga gtcgtattac ctgcagg 57

<210> 15
<211> 60
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Adapter molecule - sense strand

<400> 15
acgagcggat aacaatttca cacagggcgg ccgctaatac gactcactat agggggtcgac 60

<210> 16
<211> 60
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Adapter molecule - antisense strand

<400> 16
gtcgaccctt atagtgagtc gtattagcgg ccgccctgtg tgaaattggt atccgctcgt 60

<210> 17
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Adapter molecule amplification primer

<400> 17
cctgcaggta atacgactca ctatagg 27

<210> 18
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Adapter molecule amplification primer

<400> 18
tagtcgacgg accgctagca tcagatc 27

- 7 -

<210> 19
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Adapter molecule amplification primer

<400> 19
ccgctaatac gactcactat aggggtcgac 30

<210> 20
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Anchor amplification primer in which n is a biotinylated
guanine

<220>
<221> modified_base
<222> (1)..(1)
<223> n is a biotinylated guanine

<400> 20
naattcaact ggaagcggcc gcagga 26

<210> 21
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Anchor amplification primer in which n is a aminated guanine

<220>
<221> modified_base
<222> (1)..(1)
<223> n is an aminated guanine

<400> 21
naattcaact ggaagcggcc gcagga 26

<210> 22

- 8 -

<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Anchor amplification primer

<400> 22
taatacgact cactataggg a 21

<210> 23
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Cyclophilin primer in which n is an aminated guanine

<400> 23
gaattcaact ggaagcggcc gcaggaat 28

<210> 24
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Cyclophilin primer in which n is an aminated guanine

<220>
<221> modified_base
<222> (1)..(1)
<223> n is an aminated guanine

<400> 24
natcataaac ttaactttgc aatccagcta gacttgaagg g 41

<210> 25
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Cyclophilin primer in which n is a biotinylated guanine

- 9 -

<220>
<221> modified_base
<222> (1)..(1)
<223> n is a biotinylated guanine

<400> 25
natcataaac ttaactttgc aatccagcta gacttgaagg g 41

<210> 26
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Cyclophilin primer in which n is an aminated guanine

<220>
<221> modified_base
<222> (1)..(1)
<223> n is an aminated guanine

<400> 26
ntcataaaact taactttgca atcc 24

<210> 27
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Cyclophilin primer in which n is a biotinylated guanine

<220>
<221> modified_base
<222> (1)..(1)
<223> n is a biotinylated guanine

<400> 27
ntcataaaact taactttgca atcc 24

<210> 28
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Cyclophilin primer in which n is an aminated guanine

- 10 -

<220>
<221> modified_base
<222> (1)..(1)
<223> n is an aminated guanine

<400> 28
ntcctgagct acagaaggaa tggtttgatg ggtaaaatgc c 41

<210> 29
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Cyclophilin sequencing primer

<400> 29
ttagagctgt ccacagtcgg 20

<210> 30
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Primer for novel gene M41 in which n is an aminated adenine

<220>
<221> modified_base
<222> (1)..(1)
<223> n is an aminated adenine

<400> 30
ntcacaagtc cagaatgctg ggagc 25

<210> 31
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Primer for novel gene M41 in which n is a biotinylated adenine

<220>
<221> modified_base

- 11 -

<222> (1)..(1)

<223> n is a biotinylated adenine

<400> 31

ntcacaagtc cagaatgctg ggagc

25

<210> 32

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> ()..()

<223> Gene specific primer for novel gene M41

<400> 32

ctccaggagc ttcagggtcca tgttgc

26

<210> 33

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> ()..()

<223> Gene specific primer for novel gene M41

<400> 33

ctgaggagca gaaacaggat tctcc

25

<210> 34

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> ()..()

<223> Gene specific primer for novel gene M41

<400> 34

ctaacctcag caggaagtgt ccaag

25

<210> 35

<211> 32

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> ()..()

- 12 -

<223> Primer for novel gene M51 in which n is a biotinylated thymine

<220>

<221> modified_base

<222> (1)..(1)

<223> n is a biotinylated thymine

<400> 35

nctagattca ggcaagtgac acattttgac cg

32

<210> 36

<211> 32

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> ()..()

<223> Primer for novel gene M51 in which n is an aminated thymine

<220>

<221> modified_base

<222> (1)..(1)

<223> n is an aminated thymine

<400> 36

nctagattca ggcaagtgac acattttgac cg

32

<210> 37

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> ()..()

<223> Sequencing primer for novel gene M51

<400> 37

accgactcat ctgaagaaca actcc

25

<210> 38

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> ()..()

<223> Sequencing primer for novel gene M51

<400> 38

- 13 -

gccacagcac agcctgagtc tgaac

25

<210> 39

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> ()..()

<223> Sequencing primer for novel gene M51

<400> 39

ggacggtaga ccatgaatt

19

<210> 40

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> ()..()

<223> Sequencing primer for novel gene M51

<400> 40

acttagtcta caaaaaccta g

21

<210> 41

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> ()..()

<223> Sequencing primer for novel gene M51

<400> 41

ctgttttcct aacagttcga atg

23

<210> 42

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> ()..()

<223> Sequencing primer for novel gene M51

<400> 42

ctacctaaca atgaaaattt tcaacc

26

- 14 -

<210> 43
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Sequencing primer for novel gene M51

<400> 43
caactgaata tgtcaagtca aattcc 26

<210> 44
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Adapter molecule amplification primer

<400> 44
acgagcggat aacaatttca cacagg 26

<210> 45
<211> 65
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Biotinylated anchor primer containing a T3 promoter site

<220>
<221> modified_base
<222> (1)..(1)
<223> n is a biotinylated guanine

<400> 45
naattattaa ccctcactaa agggacaact ggaagcggcc gcaggaattt tttttttttt 60
ttttt 65

<210> 46
<211> 65
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature

- 15 -

<222> ()..()

<223> Aminated anchor primer containing a T3 promoter site

<220>

<221> modified_base

<222> (1)..(1)

<223> n is an aminated guanine

<400> 46

naattattaa ccctcactaa agggacaact ggaagcggcc gcaggaattt tttttttttt 60
ttttt 65

<210> 47

<211> 40

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> ()..()

<223> Biotinylated anchor primer containing a T3 promoter site

<220>

<221> modified_base

<222> (1)..(1)

<223> n is a biotinylated guanine

<400> 47

naattattaa ccctcactaa agggacaact ggaagcggcc 40

<210> 48

<211> 40

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> ()..()

<223> Aminated anchor primer containing a T3 promoter site

<220>

<221> modified_base

<222> (1)..(1)

<223> n is an aminated guanine

<400> 48

naattattaa ccctcactaa agggacaact ggaagcggcc 40

<210> 49

<211> 39

<212> DNA

<213> Artificial Sequence

- 16 -

<220>
<221> misc_feature
<222> ()..()
<223> Anchor amplification primer

<400> 49
aattattaac cctcactaaa gggacaactg gaagcggcc 39

<210> 50
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> ()..()
<223> Cylophilin amplification primer

<400> 50
gtcataaact taactttgca atcc 24