

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
17 November 2005 (17.11.2005)

PCT

(10) International Publication Number
WO 2005/108608 A1

(51) International Patent Classification⁷: **C12Q 1/68**

(74) Agents: **OKUYAMA, Shoichi** et al.; 8th Floor, Akasaka Noa Bldg., 2-12, Akasaka 3-chome, Minato-ku, Tokyo 1070052 (JP).

(21) International Application Number:
PCT/JP2004/010174

(22) International Filing Date: 9 July 2004 (09.07.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
2004-139869 10 May 2004 (10.05.2004) JP

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, 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, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(71) Applicants (*for all designated States except US*): **KABUSHIKI KAISHA DNAFORM** [JP/JP]; 3-35, Mita 1-chome, Minato-ku, Tokyo 1080073 (JP). **RIKEN** [JP/JP]; 2-1, Hirosawa, Wako-shi, Saitama 3510198 (JP).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **HAYASHIZAKI, Yoshihide** [JP/JP]; 22-8, Inarimae, Tsukuba-shi, Ibaraki 3050061 (JP). **HARBERS, Matthias** [DE/JP]; c/o Kabushiki Kaisha Dnaform, 3-35, Mita 1-chome, Minato-ku, Tokyo 1080073 (JP). **LEZHAVA, Alexander** [GE/JP]; c/o Kabushiki Kaisha Dnaform, 3-35, Mita 1-chome, Minato-ku, Tokyo 1080073 (JP). **CARNINCI, Piero** [IT/JP]; c/o Riken, 2-1, Hirosawa, Wako-shi, Saitama 3510198 (JP).

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report

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

(54) Title: METHOD FOR ISOLATING NUCLEIC ACID ISOFORMS

(57) Abstract: A method is disclosed for isolating nucleic acids that comprise sequences corresponding to a portion of a gene that is differentially spliced, comprising the steps of hybridizing single-stranded linear nucleic acids with a single-stranded circular nucleic acid so as to form a circular hybrid, and recovering a circular hybrid that has a single-stranded loop structure resulting from an unpaired region in either of a hybridized linear nucleic acid or the hybridized circular nucleic acid hybridized or both.

WO 2005/108608 A1

DESCRIPTION

METHOD FOR ISOLATING NUCLEIC ACID ISOFORMS

Field of the Invention

5 The present invention relates to the isolation, identification, analysis, preparation and cloning of nucleic acid isoforms for which at least a part of their sequences is complementary to each other and at least another part of their sequences are mutually unrelated or non-complementary. More specifically, the present invention relates to a method for obtaining information and materials of value to understand biological
10 phenomena such as RNA processing and use such information and materials in commercial applications.

Background Art

15 Recent results from genome sequencing projects in combination with our knowledge about transcribed genes have unveiled the complexity of transcriptome and its mechanisms associated with the utilization of genomic information. In particular, it has become ever clearer that the complexity of higher organisms can only be encoded and maintained by making alternative use of genomic information encoded in DNAs. It is
20 now clear that the information is not simply transcribed and copied into messages used in protein synthesis, but that additional mechanisms for combinatorial rearrangements and processing of the information are essential for the diversification and expansion of the genetic pool (Zavolan M, et al., Genome Res. 13 (2003) 1290-1300, hereby incorporated herein by reference). This is for example demonstrated by the whole
25 genome sequences from such different organisms as humans, nematode *C. elegans*, fly *Drosophila melanogaster*, and complex bacteria *Streptomyces*. Such genome sequences have shown unexpectedly small differences in the total numbers of genes encoded in each of those genomes while biological differences are enormous. The human genome would encode only about 1.5 times as many genes as that of the relatively simple
30 nematode *C. elegans*. This uncanonical phenomenon may be explained by mechanisms of alternative splicing which are found increasingly important during the development eukaryotes.

The about 25-year-old discovery that eukaryotic genes consist of introns and exons initiated a new field in life science focusing on mRNA processing (Black DL, Annu. Rev. Biochem., 72 (2003) 291-336, hereby incorporated herein by reference). The
5 initial transcript obtained as a direct copy of the coding DNA is often processed further before its transport from the nucleus to the place of protein synthesis. During this processing, parts of the message – so-called introns – are excised, while other parts of the message – so-called exons – are recombined into a mature transcript which becomes subjected to translation into proteins. As introns and exons are both transcribed into a
10 pre-mRNA, additional steps are required, in which non-coding introns have been removed and coding exons have been linked together in the correct order. This so-called splicing process is essential for the correct processing of mRNA molecules and for a correct translation of genetic information into proteins. Due to the differential splicing of its pre-mRNA, a single gene can encode multiple isoforms on the protein
15 level. Each isoform is distinct by alternative exon usage. Recent studies expand on this general concept, as non-coding messages have also been found to be spliced and their messages are processed in a similar way as with coding messages so as to allow them to fulfill their thus-far unknown functions (Numata K et al., Genome Res. 13 (2003) 1301-1306, hereby incorporated herein by reference).

20 An understanding of such alternative splicing mechanisms and the distinct proteins resulting therefrom have become more important as an increasing number of reports point to human diseases and aging aberrations related to miss-splicing or a lack of alternatively spliced isoforms (Caceres JF, and Kornblihtt AR, Trends Genet. 18 (2002) 186-193; and Faustino NA, and Cooper TA, Genes&Dev. 17 (2003) 419-437, both
25 hereby incorporated herein by reference). Therefore, there is a clear demand for the detection and characterization of alternatively spliced mRNA molecules so as to help develop novel methods for the development of assays and for the identification of targets toward drug discovery as well as diagnostics (Bracco L, and Kearsey J, Trends
30 Biotech. 21 (2003) 346-353, hereby incorporated herein by reference).

According to the present state of the art, the identification of sequence variations is thus far a complex and tedious task. For the identification of differently spliced variants, in particular, it is necessary to clone related sequences out of a plurality of nucleic acids such as those presented by cDNA libraries and forward individual clones derived thereof to further analysis. In such experiments, a given library comprising a plurality of nucleic acids can be screened for the presence of related nucleic acid molecules by having a representative nucleic acid molecule which can be used as a probe or the like to allow for the identification of other nucleic acid molecules sharing common partial stretches of sequences. Different approaches are generally used as they are well known to a person skilled in the art, or as described, for example, by Sambrook J and Russell DW in "Molecular Cloning – A Laboratory Manual", Cold Spring Harbor Laboratory Press, New York 2001, which is hereby incorporated herein by reference. Techniques that can be used for the present invention include, but are not limited to, such classical methods as colony hybridization and commercially available systems like the RecA-based Clone CaptureTM cDNA Selection Kit available from Clontech (Catalog #: K1056-1, the handbook [Protocol #: PT3246-1, Version #: PR97278], which is hereby incorporated herein by reference) and the GeneTrapper® cDNA Positive Selection System available from Invitrogen, former Gibco BRL Life Technologies (Cat. No. 10356-020, the handbook of which is hereby incorporated herein by reference). Likewise, various approaches have been described based on PCR amplification, making use of the end sequences common to the related nucleic acid molecules or targeting at the stepwise assembly of entire transcripts from individually amplified exons (Horton RM, Molecular Biotech. 3 (1995) 93-99, hereby incorporated herein by reference). However, any PCR-based approach is dependent on at least partial sequence information for designing primers. Thus, such approaches are limited in their potential for the identification of new spliced variants, and lack the potential to allow for highly parallel cloning of target genes desirable for genome wide studies.

Here, it should be noted that the primer design based on computer predictions can easily lead to the creation of artificial molecules which do not reflect the reality of naturally occurring species. Similarly, approaches that make use of partial sequences derived from transcribed regions including, but not limited to, the so-called Expressed Sequence

Tags or ESTs are not effective in providing comprehensive data on the variability of alternatively spliced transcripts (Xu Q and Lee C, Nuc. Acids Res. 31 (2003) 5635-5643, hereby incorporated herein by reference). In such a case, only approaches that focus on the use of intact transcripts or their corresponding full-length cDNAs can be considered
5 appropriate.

Although an initial analysis of individual clones obtained using any of the aforementioned approaches or any other means can be performed by restriction-enzyme digestion followed by electrophoretical separation of the resulting fragments, only full-
10 length sequencing and further computational analysis can unveil the entire genetic information of such clones. This process is, however, quite time consuming and expensive, and it furthermore does not allow for an easy up-scaling for high throughput analysis. The lack of effective measures for the parallel analysis of sequence variations poses a clear limitation on studies of differentially spliced mRNA molecules in present
15 genomic research and development projects. More effective methods for parallel sequencing or alternative approaches for the cloning of alternatively spliced genes in a functional and non-redundant manner have to be found. The need for cloning such related nucleic acid molecules will also remain necessary in the future when more powerful microarray platforms will become available for genome-wide exon usage
20 mapping, as described, for example, by Shoemaker DD et al. in Nature 409 (2001) 922-927, which is hereby incorporated herein by reference.

Independent from the progresses made in highly-parallel high-throughput sequencing approaches, it is in any case preferable to have a method for the direct cloning of
25 alternatively spliced messages to allow for a targeted analysis of such clones in a more cost effective way. This emphasizes the need for improved methods focusing on recovery of related nucleic acid molecules sharing partial stretches having common sequences along with distinct parts as found in differentially spliced transcripts.

30 US patent 6,251,590, which is hereby incorporated herein by reference, discloses such a method for identification and/or cloning of differentially spliced nucleic acids from a standard biological sample and a test biological sample. The method includes preparing

a plurality of RNAs from one sample and a plurality of DNAs from another sample followed by hybridization and formation of RNA/DNA hybrids. The RNA molecule comprising an unpaired region corresponding to a portion of the DNA which is differentially spliced between the samples is then identified. The method disclosed in

5 US 6,251,590 is limited to the preparation of RNA/DNA hybrids since the strategy for identification of the unpaired region is carried out essentially by means of the use of RNase H enzyme. This enzyme cuts RNA bound to DNA, but does not cut single-stranded RNA (as found in the unpaired region). The single-stranded RNA can then be recovered and cloned by standard methodology known to a person skilled in the art.

10 This method, however, shows several drawbacks and lacks efficiency due to the fact that the RNase H cuts the RNA that is hybridized to DNA into fragments of 3 to 10 nucleotides that could interfere with the later cloning step, and that RNA fragments of generally 10 to 50 nucleotides which are only partially hybridized to DNA can be released into the mixture of RNA fragments after digestion with RNase H. This would

15 make it difficult to distinguish the unpaired regions from byproducts of the enzymatic reaction, in which both RNA fragments can have a length of some 10 to 50 nucleotides. US 6,251,590 further proposes a method for recovering RNA molecules that correspond to the unpaired region by carrying out a reverse transcription reaction initiated by random priming. The problem is, however, that random primers can hybridize at any

20 position within the RNA molecule corresponding to the unpaired region, as well as at such positions in RNA fragments unspecifically released during the RNase H treatment. Thus, this strategy is not very reliable as the unpaired region cannot be recovered in its full-length, and it is unclear whether artifacts that are not associated with unpaired regions have occurred. Furthermore, small portions of sequences or fragments of the

25 unpaired region are highly likely to be recovered. Therefore, this method is not efficient as it results in a high background of false positives and can give rise to undesirable artifacts. Moreover, as the sequence information obtained by this approach only partially describes the unpaired regions, the method does not teach how corresponding transcripts have been assembled from different exons. Here, additional information on

30 the flanking regions is essential to identify splicing sites at which different exons have been re-associated.

An alternative approach for the cloning and analysis of unpaired regions in nucleic acids has been disclosed in WO2004/053163, which is hereby incorporated herein by reference. This approach in continuation of the aforementioned technology allows the selective cloning of unpaired regions along with their flanking regions. Thus, this approach overcomes one of the critical drawbacks of US patent 6,251,590 as the resulting fragments cover the entire sequence of the unpaired regions. In combination with alignments of the sequence information derived from the unpaired regions along with the sequence information derived from the flanking regions, individual exons and their splicing sites can be identified. However, this approach is based on the hybridization of DNA fragments derived from one or more samples. During hybridization, DNA molecules having stretches of matching sequences form dimeric molecules, and stretches of unrelated sequences, e.g., those derived from alternative exon usage give rise to loop structures comprised of single-stranded DNA. Such hybrids are then subjected to partial digestion with frequently-cutting restriction endonucleases to separate sections with unpaired regions. This step is essential to perform this approach so as to clone unpaired regions into libraries. Such libraries would most likely cover all unpaired regions present in the samples. For the effective cloning of the unpaired regions, further enrichment is required by binding DNAs to a single-stranded DNA binding substance, such as protein, antibody or randomized oligonucleotide, before cloning into the library. However, as this selection step is dependent on the presence of unpaired regions separated by stretches of double-stranded DNAs, overhangs of single-stranded DNA at the ends of the molecules have to be removed before the enrichment step using an exonuclease such as Exonuclease VII. Since this approach again depends on this exonuclease treatment, alterations at the ends of the related nucleic acid molecules cannot be detected, and full-length molecules cannot be cloned with their entire integrity intact. Although this approach offers a potent and robust method for obtaining unpaired regions with their flanking regions for the identification and characterization of alternative exon usage, and is of great value for the annotation of individual exons, this approach still does not allow for the cloning of the corresponding full-length messages and does not make it possible to analyze alternative splicing events. Furthermore, this technology does not teach how to select

full-length clones comprising unmatched regions for their selective cloning, but rather focuses on the selective cloning of internal exons subject to alternative splicing.

Therefore, an improved and efficient method is clearly needed for the identification,
5 selection and preparation of nucleic acids which result from the same or related regions
in the genome. The methods according to the aforementioned disclosures focus only on
the individual exon as a subject of an alternative splicing event. However, in order to
fully understand the biological relevance of related splicing events, and in order to use
alternatively spliced transcripts in assays or commercial applications, the cloning of
10 full-length cDNA clones comprising the entire message is definitely required. Such
cloning should lead to a progress in this field and development of commercial products
or services.

Summary of the Invention

15

It is an object of the present invention to overcome at least some of these problems that
exist in the state of the art and provides a reliable, fast, and effective technology for the
identification, analysis and/or cloning of full-length cDNAs derived from samples
comprising pluralities of nucleic acids with partially unrelated sequences. The present
20 invention provides a method necessary for the selective cloning of alternatively spliced
messages which are important for capturing the true value of a transcriptome and
transcripts it contains.

The present invention provides a method for preparing nucleic acid molecules having
25 unpaired regions that can be performed in a genome wide scale or otherwise. The
method of the present invention can be used to detect full-length nucleic acid molecules
having unpaired regions. This method does not require the fragmentation of genetic
information. The present invention makes it possible to manipulate full-length mRNAs
or cDNAs derived from such mRNAs so as to specifically identify and enrich such
30 nucleic acid molecules among a plurality of nucleic acids which are distinct from one
another by having an unrelated partial sequence as well as regions of common
sequences.

More specifically, the present invention provides a method for isolating nucleic acid molecules whose sequences include a portion complimentary to each other and a mutually unrelated portion, comprising the steps of: hybridizing a single-stranded linear nucleic acid molecule with a single-stranded circular nucleic acid molecule so as to form a hybrid molecule; filling in a single-stranded region of the hybrid circular nucleic acid which region is not hybridized with the single-stranded linear nucleic acid molecule with nucleic acids so as to form a double-stranded hybrid which is double stranded except at any single-stranded loop portion; and recovering a hybrid molecule that has a single-stranded loop portion resulting from an unpaired region in either of the hybridized linear nucleic molecule acid or the hybridized circular nucleic acid molecule or both. Here, "mutually unrelated portion" means portions of two nucleic acid molecules that are insufficiently complementary or non-complementary to each other so that at those portions hybridization does not occur under standard experimental conditions.

In one embodiment, the invention relates to the use of full-length cDNAs. Full-length cDNAs are specifically cloned and enriched to comprise entire transcripts that naturally occur in a biological sample. Such full-length cDNAs that make up a library or that are individually isolated are manipulated such that nucleic acid molecules partially having stretches of common sequences and unrelated sequences are separated from nucleic acid molecules which are entirely distinct from one another or are identical to each other over their entire sequences.

In a different embodiment, the invention relates to the use ribonucleic acids which are derived from a biological sample or synthesized *in vitro* and which comprise the full-length of naturally occurring transcripts. Thus, the invention provides a method for manipulating ribonucleic and deoxyribonucleic acids such that nucleic acid molecules having stretches of common sequences are separated from nucleic acid molecules that are entirely distinct from each other or identical to each other over their entire sequences.

Although it is desirable to perform the invention with nucleic acid molecules having the entire information of naturally occurring transcripts, the invention is not limited to the use of such nucleic acids found in or related to biological samples. Thus, the invention provides a method that allows for manipulation of any nucleic acids such that nucleic acid molecules having common sequences in at least a part of their entire sequences and unrelated sequences can be separated from those that are entirely distinct from each other or are identical to each other over their entire sequences. Thus, the invention is unrelated to the nature of the nucleic acid to which it is applied, but allows use of any nucleic acid so as to select only those nucleic acids that have parts of their sequences in common along with unrelated sequences.

In order to carry out the invention, a set of one or more nucleic acid molecules (called "driver") is put into contact with a plurality of nucleic acids comprising one or more distinct nucleic acid molecules (called "tester"). Normally, the "tester" and the "driver" comprise complex mixtures of nucleic acids out of which related nucleic acid molecules partially having stretches of common sequences can be isolated by the method of the present invention. Thus, the invention relates to a procedure for selecting specific nucleic acids by bringing two samples, the "tester" and the "driver", into contact.

In order to select individual nucleic acid molecules present in the "tester" and/or the "driver", the nucleic acid molecules have to be presented as single-stranded nucleic acids, and they can be comprised of RNAs and/or DNAs. Thus, the invention relates to the hybridization of such single-stranded molecules, and the nucleic acids derived from the "tester" and the "driver" should, to a limited extent, have sense and antisense orientations with respect to one another.

For the purpose of the current invention, hybrid molecules formed by related nucleic acid molecules which share stretches of complementary sequences while having stretches of unrelated sequences can be selectively separated from single-stranded nucleic acid molecules and dimeric nucleic acid molecules that have identical sequences. Thus, the invention provides a method for the fractionation of nucleic acid molecules by

bringing together single-stranded nucleic acid molecules presented as a “driver” and a “tester.”

Dimeric nucleic acid molecules in which monomer molecules have stretches of
5 common sequences and at least one region of unrelated sequence relative to each other,
and that are selectively enriched from a plurality of nucleic acids by means of an
embodiment of the present invention can be manipulated so as to isolate one of the two
strands within the dimeric nucleic acid molecule and use it in further manipulation.
Such manipulations include, but are not limited to, the cloning of one of the two strands
10 of such molecule by standard methods known to a person skilled in the art. Thus, the
invention provides a method for isolating nucleic acid molecules which partially share
homologous sequences and otherwise have regions of differing sequences.

In one embodiment of the invention the “driver” is composed of RNAs whereas the
15 “tester” is composed of DNAs. Thus, the invention allows for the formation of hybrid
nucleic acid molecules in which one strand derived from the “driver” is made up of
RNA, and the other strand is made up of DNA derived from the “tester”.

In a preferable embodiment of the invention, the “tester” is derived from a plasmid.
20 Such a plasmid is used as a template for the preparation of a circular single-stranded
DNA. Thus, the “tester” may comprise elements required for the replication of the
“tester” molecule *in vitro* or *in vivo*.

In a different embodiment of the invention, the “driver” is prepared *in vitro* by using a
25 RNA polymerase and a DNA template. Such a template can be a linear template
prepared by any method known to a person skilled in the art including, but not limited
to, the use of cDNAs.

In a preferable embodiment of the invention, the “driver” is prepared from naturally
30 occurring RNA prepared from a biological sample. Such RNA includes, but is not
limited to, total RNA preparations as well as enriched RNA fractions comprised of
mRNA.

The invention further relates to the cloning of related nucleic acid molecules which may comprise naturally occurring sequences. As such nucleic acid molecules relate to naturally occurring species, the invention provides a method for the analysis and
5 cloning of naturally occurring nucleic acid molecules derived from the same gene but are distinct in terms of their exon usage. Thus, the invention provides an approach for the isolation and cloning of alternatively or differentially spliced gene products.

In a different embodiment the invention provides a method for the selective cloning of
10 alternatively or differentially spliced transcripts. Such gene products are often called isoforms and may be cloned according to the invention without the need of fragmentation. These nucleic acid isoforms originate from the same gene but differ in its exon and intron usage. Thus, the invention may relate to preparation of isoforms and the handling and isolation of molecules comprising the entire length of transcribed
15 molecules.

Furthermore, the invention provides a method for the analysis of the nucleic acid molecules prepared by means of the invention by subjecting such molecules to partial or full-length sequencing. Such sequence information can be obtained by any approach
20 known to a person skilled in the art, including, but not limited to, end-sequencing or full-length sequencing by primer walking, shotgun approaches or random transposon integration. Thus, the invention also provides a method for the computational analysis of sequence information related to nucleic acid molecules prepared according to an embodiment of the present invention.

25

In a different embodiment of the invention nucleic acid molecules prepared according to an embodiment of the invention are analyzed by means of hybridization to reference samples. Such reference samples comprise nucleic acid molecules having sequences at least partially homologous to the nucleic acid molecules prepared according to an
30 embodiment of the invention. Thus, the knowledge about the nucleic acid molecules within a sample can be used for the annotation of such nucleic acid molecules which hybridize to molecules in the sample.

Furthermore, the invention provides a method for the analysis of nucleic acid molecules prepared according to an embodiment of the invention by the use of endonucleases having specific recognition sites. The fragmentation of a nucleic acid molecule by such an endonuclease can produce fragments of specific length characteristic for any given nucleic acid molecule. Thus, the present invention provides a different method for the analysis of nucleic acid molecules which may reflect alternative exon usage.

The present invention provides a method for the cloning of alternatively spliced transcripts of nucleic acid molecules or clones derived thereof, which may comprise full-length molecules, allowing for functional analysis of differently spliced transcripts, which may be called nucleic acid isoforms, and the products encoded by them. Thus, the present invention relates to a method for functional assays on proteins derived from alternatively spliced transcripts.

The present invention provides a new approach for the isolation, cloning and analysis of related nucleic acid molecules having common stretches of sequences in parts, so as to overcome fundamental restrictions existing in the state of the art for the analysis and use of genes and transcript thereof which are subject to regulation on the level of alternative splicing. Thus, the present invention allows addressing entirely new questions on how alternatively or differentially spliced transcripts contribute to biological phenomena including, but not limited to, diseases and public health problems.

Furthermore, the present invention provides a new method for the isolation of transcripts which are differentially spliced and encode for proteins of different activities. Therefore, the present invention opens up new areas for commercial applications including, but not limited to, the target identification for diagnostics and drug discovery, direct use as medications, and the preparation of proteins having distinct enzymatic functions. Similarly, the present invention can be applied to design a kit comprising the necessary reagents to perform the invention for research and development. Thus, the present invention can be applied in a wide area of commercial interest and beyond, and will contribute to future developments in the relevant fields.

Brief Description of the Drawings

Figure 1 is a schematic diagram showing the principle for the preparation of starting materials, "Tester" and "Driver".

Figure 2 is a schematic diagram showing the preparation of the "Tester" by way of enzymatic digestion.

Figure 3 is a schematic diagram showing the preparation of "Tester" by way of ligation-mediated or asymmetric PCR.

Figure 4 is a schematic diagram showing the tagging of "Tester" DNAs.

Figure 5 is a schematic diagram showing the preparation of "Driver" molecules for hybridization (optional).

Figure 6 is a schematic diagrams showing possible combinations formed as a result of DNA/RNA hybridization.

Figure 7 is a diagram showing the Exo VII digestion of a franking single-stranded portion and the fill-in process for second strand synthesis.

Figure 8 is a diagram showing isolation or identification of species with loop structures using a substance that binds to single-stranded nucleic acids.

Figure 9 is a diagram showing the removal of RNAs by RNase treatment.

Figure 10 shows the results of insert size check on input and output libraries by *Pvu* II digestion: Lanes: 1 and 6 – *1/Sty*I, 2. G2-1/*Pvu*II (single clone from the initial library G2), 3. G3-2/*Pvu*II (single clone from the initial library G3), 4. G2A-3/*Pvu*II (single

clone from the new library G2), 5. G3A-4/*Pvu*II (single clone from the initial library G3).

Figure 11 shows the results of sequence analysis by BLAST search in NCBI database.

5 Related sequence information is given in Example 17.

Detailed Description of the Invention

10 The present invention can be employed in a wide range of applications in gene discovery, genomic research, and manufacturing or services for producing recombinant DNAs. The invention is also generally applicable to life science and medical research. The methods disclosed herein are of high commercial value and contribute to many applications in the field of biotechnology. In particular, the approach of the present invention will greatly contribute to academic and commercial research and development
15 in any field in which questions and products related to alternative exon usage need to be addressed.

The invention encompasses methods for handling single-stranded as well as double-stranded nucleic acids in the form of linear and circular nucleic acid molecules.
20 Double-stranded DNA means a nucleic acid molecule which is composed of two polymers formed by deoxyribonucleotides and in which the two polymers have sequences that are complimentary to each other in a manner sufficient for allowing their association to form a dimeric molecule. The two polymers are bound to one another by specific hydrogen bonds formed between matching base pairs within the deoxyribonucleotides. Any DNA molecule composed only of one polymer chain
25 formed by two or more deoxyribonucleotides having no matching complementary DNA molecule to associate with is considered to be a single-stranded DNA molecule for the purpose of the present invention, even if such a molecule may form secondary structures comprising double-stranded DNA portions. As used interchangeably herein, the terms
30 "nucleic acid molecule(s)" and "polynucleotide(s)" include RNA or DNA regardless of single or double-stranded, coding or non-coding, complementary or not, and sense or antisense, and also include hybrid sequences thereof. In particular, it encompasses

genomic DNA and complementary DNA which are transcribed or non-transcribed, spliced or not spliced, incompletely spliced or processed, independent from its origin, cloned from a biological material, or obtained by means of synthesis. RNA for the purpose of the present invention is considered a single-stranded nucleic acid molecule even if such a molecule may form secondary structures comprising double-stranded portions. In particular, RNA encompasses for the purpose of the invention any form of nucleic acid molecule comprised of ribonucleotides, and does not related to a particular sequence or origin. Thus, RNA can be transcribed *in vivo* or *in vitro* by artificial systems or it may be non-transcribed, spliced or not spliced, incompletely spliced or otherwise processed. It may be independent from its natural origin or derived from artificially designed templates, mRNA, tRNA, rRNA, obtained by means of synthesis, or any mixture thereof. More precisely, the expressions "DNA", "RNA", "nucleic acid", and "sequence" encompass nucleic acid materials themselves and are thus not restricted to particular sequence information, vector, phagemid or any other specific nucleic acid molecules. The term "nucleic acid" is also used herein to encompass naturally occurring nucleic acids, artificially synthesized or prepared nucleic acids, any modified nucleic acids into which at least one or more modifications have been introduced by naturally occurring events or through approaches known to a person skilled in the art. The terms "purity", "enriched", "purification", "enrichment", or "selection" are used interchangeably herein and do not require absolute purity or enrichment of a product, and these terms are intended to provide relative definitions. The terms "specific", "preferable", or "preferential" are also used interchangeably herein and do not require absolute specificity of a DNA or RNA hybridization probe, or an enzyme for its substrate or an activity, but rather they are intended to have relative definitions which include the possibility that an enzyme may have low or lower affinity to other compounds related or unrelated to its substrate. Similarly, the terms used to name an enzyme, an enzymatic activity, or a single-stranded-nucleic acid binding substance are used herein to describe the function or activity of such a component, but do not require the absolute purity of such components. Thus, any mixture containing such an enzyme, enzymatic activity, single-stranded-DNA, RNA or nucleic acid binding substance or mixtures thereof with other components of the same, related or unrelated function are within the scope of the invention. Similarly, DNA or RNA molecules may function in a

specific manner as hybridization probes, and as such, are related to as “complementary sequences” for the purpose of the invention. When such probes are applied for the detection of a related nucleic acid molecule, such a probe and the target molecule may be distinct by naturally occurring or artificially introduced mutations in individual
5 positions, but still they may be considered to involve complementary sequences. The term “biological sample” includes any kind of material obtained from living organisms including microorganisms, animals, and plants, as well as any kind of infectious particles including viruses and prions which depend on a host organism for their replication. As such “biological sample” include any kind material obtained from a
10 patient, animal, plant or infectious particle for the purpose of research, development, diagnostics or therapy. Thus, the invention is not limited to the use of any particular nucleic acid molecules or their origin, but the invention provides general means to be applied to and used for the work on and the manipulation of any given nucleic acid. Any such nucleic acid molecules as applied to perform the invention can be obtained or
15 prepared by any method known to a person skilled in the art including, but not limited to, those described by Sambrook J and Russell DW, *ibid*, which is hereby incorporated herein by reference.

The present invention relates further to a method for the preparation of nucleic acid
20 molecules having unpaired regions that can be performed on a genome wide or otherwise. In particular, the invention relates to such a method for the preparation, preferably in their entire length, of nucleic acid molecules having unpaired regions. The method of the present invention may not require the fragmentation of information during the process of performing it. Thus, the invention enables the manipulation of
25 full-length mRNAs or cDNAs derived from such mRNAs with the goal to specifically identify and enrich such nucleic acid molecules which are distinct to one another by having regions of common sequences along with at least one region of unrelated sequence. Thus, the invention relates to a method for capturing the entire genetic information including, but not limited to, that presented by naturally occurring
30 transcripts to make such genetic information available for analysis and use for protein synthesis.

The method according to an embodiment of the present invention may comprise several steps. In the following, we will discuss such steps with reference to specific examples, particularly in terms of choice of RNA and DNA for what we call "tester" and "driver", so as to make discussions easy to understand, but the present invention is by no means
5 limited to such examples.

Preparation of starting materials:

In one embodiment, the invention makes use of a plurality of nucleic acids presented in, for example, a cDNA library or alike, the so-called "tester". Preferably, such a
10 library comprises full-length cDNA clones. Even further preferably, the individual molecules within the "tester" carry all genetic elements required for replication in a host organism. For performing a method of the invention, the "tester" is brought into contact with the so-called "driver" which may comprise a plurality of single-stranded RNA molecules.

15

Preparation of single-stranded circular "tester" DNA:

As standard DNA preparations are commonly obtained in the form of double-stranded DNA, the present invention may utilize the strand-specific conversion of double-stranded DNAs into single-stranded circular DNAs.

20

Hybridization of "tester" DNA with "driver" RNA:

The single-stranded circular DNA presented as the "tester" is in the next step brought into contact with the "driver," which may, for example, be RNA. To allow for a specific hybridization between DNA molecules presented as the "tester" and RNA
25 molecules presented as the "driver", the DNA and RNA molecules must have a sense and antisense orientation to one another.

Processing of DNA/RNA hybrids:

End-sequences in RNA molecules not matching with complementary sequences within
30 the associated circular DNA can be removed by means of an RNase or exonuclease, which at the same time will destroy free regions of RNA having no matching DNA. The single-stranded portions of the resulting DNA/RNA hybrid molecules are then

subjected to fill-in process. Some regions of single-stranded circular DNA which are not masked by RNA associated with this DNA are filled in by nucleic acids. This step can be performed by a DNA polymerase, and is primed by the RNA within the hybrid molecule or by oligonucleotide primers. Such oligonucleotide priming leads to the
5 filling-in of remaining single-stranded DNA portions having no matching RNA attached to them.

Selection of DNA/RNA hybrids comprising unmatched regions:

After performing the aforementioned reactions on the DNA/RNA hybrids, only those
10 DNA/RNA hybrids having regions which are distinct between the DNA and RNA will have partial single-stranded regions. Only hybrids formed by related molecules from the "tester" and "driver" having common parts in their sequences along with some unrelated sequences will give rise to hybrid molecules having loop structures of single-stranded RNA or DNA connected to double-stranded regions. Thus, such DNA/RNA
15 hybrids having loop structures of single-stranded nucleic acids can be selectively enriched using a substance that binds to single-stranded nucleic acid. Preferably, such a single-stranded nucleic acid binding substance may bind equally well to RNA and DNA as long as they are single stranded.

20 Removal of RNA from DNA/RNA hybrids:

The RNA portion of DNA/RNA hybrids obtained by the aforementioned selection step may be removed from such hybrids by treatment with an RNase.

Cloning of selected molecules:

25 The partially double-stranded and partially single-stranded DNA molecules obtained by the aforementioned RNase treatment are transferred into a host for their cloning into libraries. Preferably, single-stranded regions within the molecules can be filled-in with DNA by means of a DNA polymerase. Even more preferably, the partly single-stranded molecules are directly transferred into a host in which its replication machinery will
30 perform the reconstitution of the double-stranded DNA. If the DNA molecules in the "tester" have all necessary genetic elements for the replication of the DNA molecules

within the host, no further manipulation or cloning steps are required to perform the invention.

Clone isolation:

- 5 Individual clones derived from the aforementioned library can be isolated by any method as known to a person skilled in the art.

Clone analysis:

- 10 Individual clones derived for the aforementioned library can be analyzed by various methods including, but not limited to, DNA sequencing, hybridization-based approaches, or approaches based on digestion with sequence specific endonucleases.

In the following, each step that may be used to perform certain aspects of the present invention is described in somewhat more detail.

15

An embodiment of the present invention involves the processing of DNA and RNA molecules presented as the "tester" and the "driver," respectively. This is shown in Fig. 1. The first step may be the selection and preparation of the necessary materials. Any kind of DNA or RNA molecules can be used regardless their origin or sequence content.

- 20 In a preferable embodiment, full-length cDNAs are used, and such cDNAs or cDNA clones are specifically selected and enriched to constitute entire transcripts naturally occurring in a biological sample.

- 25 Different approaches are known to a person skilled in the art for the preparation of cDNA libraries as described in more detail by Sambrook J and Russell DW, *ibid.* In brief, cDNA library preparations may involve the principle steps required to prepare a complementary DNA from a given mRNA using a reverse transcriptase, the formation of the second DNA strand, and the cloning of the double-stranded cDNA derived from an RNA transcript into a permissible cloning vector for further propagation. In the preparation of the cDNA libraries, it is desirable to ensure the cloning of intact full-length cDNA clones including, but not limited to, the aspects described by Das M et al., *Physiol. Genomics* 6 (2001) 57-80, which is hereby incorporated herein by reference.
- 30

Approaches for the selection of full-length cDNAs most commonly make use of the Cap structure attached to the 5'-end of Pol II transcripts. Any approach, as known to a person skilled in the art, for the recognition and/or modification of the Cap structure for the purpose of cloning or enriching full-length messages or cDNAs derived thereof can be applied to perform the invention. One such approach is the so-called Cap trapper approach as has been disclosed in US patent 6,143,528, which is hereby incorporated herein by reference, for which the Cap structure is selectively modified to allow for the introduction of a biotin moiety (Carninci P et al., *Genomics* 37 (1996) 327-336, hereby incorporated herein by reference). To perform the Cap trapper approach, a given mRNA is transcribed into cDNA by means of a reverse transcriptase to form RNA/DNA hybrids. The Cap structure within such a hybrid and any other free mRNA is then biotinylated in a chemical reaction. Remaining single-stranded RNA which is not masked by hybridization to complementary DNA is then destroyed by treatment with single-stranded RNA specific RNase I. RNase I treatment will also remove Cap structures from those RNA/DNA hybrids for which the reverse transcription reaction does not extend the complementary DNA to the 5'-end of the RNA. Thus, only full-length hybrids comprising the Capped full-length mRNA and the full-length complementary DNA can be selectively retained on a streptavidin coated matrix binding to the biotin moiety, and become subject to full-length selection. Details of a protocol for performing the Cap trapper method are available in the public domain (Carninci P and Hayashizaki Y., *Methods in Enzymology* 3003 (1999) 19-44, hereby incorporated herein by reference).

However, alternative approaches for the Cap-selection have been reported including the so-called Oligo-Capping, in which the 5' ends of mRNAs are dephosphorylated with a phosphatase, such as BAP (bacterial alkaline phosphatase), followed by treatment with the de-capping enzyme TAP (tobacco acid pyrophosphatase). Subsequently a ribonucleotide or a deoxyribonucleotide can be attached to the 5'-end of the mRNA with a RNA ligase replacing the original Cap-structure with an oligonucleotide (Maruyama K, and Sugano S, *Gene* 138 (1994) 171-174, hereby incorporated herein by reference). This oligonucleotide can be used at a later stage of the cDNA library

preparation to specifically prime the second strand synthesis. Only single-stranded DNA can be primed for second strand synthesis which comprises the complementary sequence information derived from the oligonucleotide attached to the 5'-end of the RNA. Thus, only cDNAs derived from full-length mRNA will allow for the second strand synthesis and as such become clonable into a cDNA library. Alternatively, some methods are in use to attach oligonucleotides directly to the Cap structure in a chemical reaction as disclosed in US patent 6,022,715, which is hereby incorporated herein by reference.

10 Similarly the so-called Cap-switch method as disclosed in US patent 5,962,272, which is hereby incorporated herein by reference, allows preparing the first-strand cDNA in presence of a Cap-switch oligonucleotide. The Cap switching mechanism then lets the first strand synthesis be continued on the Cap-switched oligonucleotides so as to enable a second strand cDNA synthesis.

15

Even furthermore, approaches have been described which make use of specific Cap-binding substances including, but not limited to, a Cap-binding protein (Edery I et al., Mol. Cell Biol. 15 (1995) 3363-3371, hereby incorporated herein by reference) or an antibody that specifically binds to the Cap structure (Theissen H et al. EMBO J. 12 (1986) 3209-3017, hereby incorporated herein by reference). In another embodiment, depending on the quality of RNA, particular enzyme and reaction conditions can allow for reaching the Cap-structure of RNA with high efficiency, thus being applicable for full-length cDNA cloning (Carninci P et al., Biotechniques 32 (2002) 984-985, hereby incorporated herein be reference).

25

Thus, the present invention may utilize many approaches for the preparation of full-length cDNA libraries, and any such approach is suitable. Therefore, the present invention is not dependent or limited to any particular approach or method for cDNA synthesis and for manipulating or cloning into a related cDNA library of cDNA molecules.

30

Although it is desirable to use nucleic acid molecules comprising the entire information related to naturally occurring transcripts, any nucleic acids found in or related to biological samples may also be used. Thus, the invention relates to the use of any nucleic acid fragments presented as the "tester", and such fragments can have an incomplete or random sequence that may include such sequences unrelated to naturally occurring genetic information. Such nucleic acid fragments can be prepared by any method known to a person skilled in the art including, but not limited to, fragmentation by enzymatic activities like those of nucleases, sheering with the use of a mechanic force, sonification, or random priming as well as synthesis.

As described above the invention can make use of a plurality of nucleic acids presented in a cDNA library which may be called "tester." Preferably, such a library comprises full-length cDNA clones. More preferably, individual molecules within the "tester" may carry all genetic elements required for replication in a host organism. In even more preferably, the "tester" may be derived from a plasmid, and such a plasmid may be used as a template for the preparation of a single-stranded circular DNA. Thus, the "tester" may comprise elements required for the replication of the "tester" molecule *in vitro* or *in vivo*. DNA subjected to library preparation is commonly cloned into plasmids, which include double-stranded DNA molecules covalently closed to form circular DNA and replicated as extrachromosomal molecules. Plasmids behave as accessory genetic units harboring regulatory elements in the so-called replicon and use the replication machinery of their host bacteria to maintain and control their copy numbers within the host cell. Often plasmids further contain genes encoding for enzymatic activities which can be used as selection markers. For the purpose of using plasmids as vectors for handling, amplifying and manipulating or propagation of cloned DNA, those selection markers commonly encode genes conferring resistance to specific antibiotics, and thus allow for their selection by bacterial phenotypes. Many different types of cloning vectors have been developed in the field as known to a person skilled in the art. The invention is not limited to any particular vector or plasmid to be used in library preparation, but any vector, plasmid, phagemid, phage, yeast artificial chromosome, or bacterial artificial chromosome can be used. In a preferable application, cDNA libraries provided in vector pFLC as disclosed in patent application WO 02/070720 A1 may be

used, as this vector system has proven advantageous in the preparation of full-length cDNA libraries as described by Carninci P and Hayashizaki Y, *ibid*.

According to one embodiment of the invention, the “driver” is composed of RNA
5 whereas the “tester” is composed of DNA. Thus, the invention allows for the formation
of hybrid nucleic acid molecules in which one strand is derived from the “driver” and
composed of RNA, and the other strand is composed of DNA derived from the “tester”.
In one embodiment of the invention, the “driver” is prepared *in vitro* by means of a
RNA polymerase and by using a DNA template. Such a template can be a linear DNA
10 prepared by any method known to a person skilled in the art including, but not limited
to, the use of cDNA. Methods for the preparation of such cDNA samples or any
plurality thereof have been disclosed in the aforementioned description for the
preparation of the “tester”. Thus, any such library, any amplification product, or any
individual clone derived from such a library can be used for the preparation of the
15 “driver” regardless whether or not such a clone or library comprises full-length cDNA.
RNA transcripts can be prepared from DNA templates by any method known to a
person skilled in the art including, but not limited to, those described by Sambrook J
and Russell DW, *ibid*. Commonly circular DNA templates are initially cleaved by
means of an endonuclease which has a recognition site adjacent to the termination site
20 of the transcript, whereas this step can be omitted for the use of linear amplification
products as templates in RNA synthesis. From such a liner DNA template, RNA
transcripts can be obtained by means of a RNA polymerase including, but not limited to,
T4-, T7-, or SP6 RNA polymerase, and the transcription reaction may be terminated at
the end of the linearized template.

25

In a different embodiment, the invention relates to the use ribonucleic acids derived
from a biological sample in which such ribonucleic acids comprise full-length naturally-
occurring transcripts or parts thereof. Such RNA includes, but is not limited to, total
RNA preparations as well as enriched RNA fractions made up of mRNA. Samples of
30 mRNA or total RNA can be prepared by standard methods known to a person skilled in
the art including, but not limited to, those described in more detail in Sambrook J and
Russel DW, *ibid*, or Carninci P et al., *Biotechniques* 33 (2002) 306-309, both of which

are hereby incorporated herein by reference. It can be preferable to use a cytoplasmic RNA preparation such as the one described by Carninci P et al., *ibid*, to reduce the number of unprocessed introns remaining in total RNA preparations from cells or tissues. However, any other approach may be used for the preparation of mRNA,
5 cytoplasmic RNA, or total RNA.

The preparation of mRNA from total RNA or cytoplasmic RNA is preferable, but not essential, as the use of total RNA can provide satisfying results. Generally speaking, mRNA represents about 1-3 % of total RNA preparations, and it can be subsequently
10 prepared by using commercial kits based on oligo dT-cellulose matrices. Such commercial kits include, but are not limited to, the MACS mRNA isolation kit from Milteny, the handbook of which is hereby incorporated herein by reference.

In a different embodiment of the invention, the "driver" is derived from synthetic RNA
15 prepared by any method known to a person skilled in the art including, but not limited to, the methods disclosed in US patent US 6,649,751, which is hereby incorporated herein by reference. Thus, any ribonucleic acid derived from a natural resource or created by artificial design can be used.

20 In the following step, the "tester" is converted into single-stranded circular DNA. The preparation of strand-specific single-stranded DNA is preferable as it enables the hybridization to the "driver". Many approaches have been disclosed for the preparation of single-stranded DNAs as known to a person skilled in the art. Standard approaches for the preparation of single-stranded DNA most frequently make use of so-called
25 phagemids, which are plasmid-phage hybrids obtained by cloning the *cis*-acting regulatory sequences for the initiation and termination of DNA synthesis from the single-stranded genomic DNA of the bacteriophage M13 genome into cloning vectors. Such phagemids allow for the *in vivo* preparation of single-stranded DNA when the host bacteria are infected by a helper wild-type or mutant filamentous bacteriophage carrying
30 replication-defective intergenic regions. After infection, the gene II product encoded by the helper phage introduces a strand-specific nick into the intergenic region of the phagemids initiating a rolling-circle like replication of one strand. Thereafter, single-

stranded copies of the phagemid DNA are packed into the progeny bacteriophage particles and extruded into the medium, from which the single-stranded DNA can be isolated.

5 As an alternative to the *in vivo* preparation of single-stranded DNA, *in vitro* approaches have been developed making use of combinations of two different enzymatic activities. This is shown in Fig. 2. Most commonly, a combination of the replication initiator protein Gene II of the bacteriophage f1 and the exonuclease III from *E. coli* (Exo III) is used in such systems. Here, Gene II acts as a site-specific endonuclease that recognizes
10 the f1 ori in a phagemid vectors, and cleaves the viral strand. Exo III attacks the free 3'-end of the nicked strand and digests it until the other strand is released as single-stranded circular DNA. Such a system can be commercially obtained, e.g., as part of the so-called GeneTrapper® cDNA Positive Selection System from Invitrogen, formerly Gibco BRL/Life Technologies (CAT. NO. 10356-020, the instruction manual
15 thereof is hereby incorporated herein by reference). However, as the efficiency of Gene II enzymatic activity in these reactions tends to be low, other strand-specific nicking enzymes have been developed. These include artificially-engineered nicking endonucleases which cleave only one DNA strand within their recognition sequence on a double-stranded DNA substrate. Such enzymes include, but are not limited to, the
20 commercially available nucleases N.*Bpu* 10I (FERMENTAS UAB, Vilnius, Lithuania), N.*Bbv* C IA, N.*Bst* NB I and N. *Alw* I (New England Biolabs® Inc, Beverly, USA). A detailed protocol for the application of N.*Bpu* 10I for the preparation of single-stranded DNA from supercoiled double-stranded plasmids containing an appropriate recognition site can be found on the website of Fermentas UAB under <http://www.fermentas.com/>,
25 and this protocol is hereby incorporated herein by reference.

For the purpose of this invention, single-stranded circular DNA may preferably be prepared by a novel approach based on the denaturation of a double-stranded circular DNA followed by the strand-specific annealing of a primer to one of the single-stranded
30 DNA molecule. This approach is shown in Fig. 3. Such a primer allows the synthesis of the second strand using such single-stranded DNA as a template to which the primer can bind. Any DNA polymerase known to a person skilled in the art can perform the

synthesis of the second strand, and depending on the objectives of the experiment, a DNA polymerase with or without strand displacement activity can be applied including, but not limited to, the Klenow fragment of DNA polymerase I, Vent, or Deep Vent DNA polymerase, T4 and T7 DNA polymerases, DNA polymerase I, *Taq* polymerase, 5 *Tfl* DNA polymerase, *Tth* DNA polymerase, *Tli* DNA polymerase, or any other DNA polymerase known in the field. In a preferable application, a DNA polymerase without strand displacement activity is applied such as T7 DNA polymerase, Vent, or Deep Vent DNA polymerase, and entirely filled-in and thus double-stranded DNA molecules can be subjected to DNA ligation of the newly synthesized strand to the priming 10 oligonucleotide. In such an application, the ligation step can reconstitute the initial circular double-stranded starting material, whereas one copy of one single-stranded circular DNA has been created to which the priming oligonucleotide does not bind. Any DNA ligase known to a person skilled in the art including, but not limited to, T4 DNA ligase, *E. coli* ligase, or *Taq* DNA ligase can be used to perform the ligation 15 reaction under standard conditions. Thus, this method allows the selective preparation of one strand of single-stranded DNA from a double-stranded DNA template by the combined use of a DNA polymerase and a DNA ligase. Similar to the amplification of linear DNA by means of the PCR method mentioned later on, this approach can be performed as a cycle to prepare and enrich single-stranded circular DNA using 20 thermostable DNA polymerase and DNA ligase.

Additional protocols for the preparation of linear single-stranded DNA are known, which may be used as well. It is within the scope of the present invention to prepare a linear single-stranded DNA which in an additional step is circularized by a ligation step. 25 Preferably, such a ligation step is mediated by an oligonucleotide comprising sequence regions complementary to the end sequences of the linear single-stranded DNA. Thus, such an oligonucleotide can connect the open ends of the linear single-stranded DNA molecule and enable a more effective ligation of the ends. Any DNA ligase known to a person skilled in the art including, but not limited to, T4 DNA ligase, *E. coli* ligase, or 30 *Taq* DNA ligase can be used to perform the ligation reaction under standard conditions. The remaining oligonucleotide can be removed from the circular single-stranded DNA by heating the template followed by a size fractionation step. However, it is within the

scope of the invention to leave such an oligonucleotide as attached to the circular single-stranded DNA, and use it at a later stage for priming the fill-in reaction of the circular single-stranded DNA at portions where such circular single-stranded DNA should be partly or entirely converted into double-stranded DNA. Thus, linear single-stranded
5 DNA can be used, and such liner DNA may have to be modified into circular single-stranded DNA.

For the preparation of linear single-stranded DNA, various technologies have been developed as familiar to a person skilled in the art. In many cases these approaches use
10 a DNA-polymerase-based synthesis of single-stranded DNA from a DNA or RNA template. Any amplification method from linear template DNA or RNA yielding an excess of single-stranded DNA over the template can be applied. Such approaches include the use of primed reactions driven by a DNA polymerase performed as an individual reaction or as a cyclic reaction to allow for a linear amplification of the
15 product. For the preparation of high quality single-stranded DNA, it may be advisable to transcribe a DNA template first into RNA by means of a RNA polymerase. The template DNA can then be destroyed by means of a deoxyribonuclease before the RNA transcript is used as a template for the synthesis of single-stranded DNA by means of a reverse transcriptase. By the use of two different forms of nucleic acids in the two
20 independent reactions, the approach offers a method for the removal of the templates by a deoxyribonuclease and a ribonuclease, respectively.

In a particular case, the synthesis of single-stranded DNA can be achieved by the so-called asymmetric PCR reaction, in which the two primers are used at different
25 concentrations. After the rate-limiting primer is exhausted, the reaction switches from the exponential amplification of double-stranded DNA to the linear amplification of the one strand primed by the primer used in excess over the rate-limiting primer. In an alternative approach lambda exonuclease is used to digest the one strand of double-stranded DNA having a 5'-phosphorylated end. Such a template can be prepared in
30 PCR reactions in which only one out of two primers is phosphorylated at the 5'-end. The lambda exonuclease, also denoted as "StrandaseTM", is commercially available from Novagen, Madison, USA, and the documentation on its "StrandaseTM ssDNA

Preparation Kit", Cat. No. 69202, is hereby incorporated herein by reference. Similarly, the enzyme can also be obtained as lambda exonuclease from Epicentre, Madison, USA (Cat. Nos. LE035H and LE032K, the documentation on which is hereby incorporated herein by reference).

5

Alternative approaches for the preparation and purification of single-stranded DNA have been disclosed in Japanese patent application No. 2004-030686, which is hereby incorporated herein by reference. Here, the use of a double-stranded DNA specific endonuclease is described, allowing for the purification of single-stranded DNA in which such single-stranded DNA is derived from double-stranded templates. The most preferable double-strand-specific endonuclease is the Duplex-Specific Nuclease (DSN) from crab hepatopancreas, as described by Shagin DA et al., Genome Res. 12 (2002) 1935-1942, which is hereby incorporated herein by reference. DSN is characterized for its double-strand specificity, which allows the authors to use the enzyme for the detection of SNPs in double-stranded DNA (Shagin DA et al., *ibid*), and as further described by the provider Evrogen (Moscow, Russia), whose product information on DSN is hereby incorporated herein by reference (<http://www.evrogen.com/index.shtml>). As single-stranded DNA can fold into secondary structures with in part double-stranded DNA, it is more preferably to applied DSN together with a substance having single-stranded DNA binding affinity including, but not limited to, the use of SSB from *E. coli*, products of phage T4 Gene 32, adenovirus DBP, an antibody directed against single-stranded DNA, calf thymus UP1, or any mixture thereof. Thus, the combination of DSN with a single-stranded DNA binding substance can provide an effective means to help the preparation of linear or circular single-stranded DNA. The aforementioned approach using a double-strand specific endonuclease would be preferable for the preparation of single-stranded circular DNA to be applied to perform the invention as prepared by any of the aforementioned methods for single-stranded DNA preparation.

25

A "tester" prepared according to any of the aforementioned approaches can be used for the purpose of the present invention. As shown in Fig. 4, in a different embodiment, it may be preferable to hybridize short oligonucleotides to the single-stranded "tester" so that such oligonucleotides hybridize to regions within the tester that are unrelated to the

30

DNA inserts. Such tagging oligonucleotides hybridize with the tester backbone, and allow for the priming for the second-strand synthesis. Preferably, such tagging oligonucleotides are 15 to 100 nucleotides long. More preferably, such tagging oligonucleotides are 20 to 50 nucleotides long. Even more preferably, such tagging
5 oligonucleotides are 20 to 30 nucleotides long.

In still another embodiment, a tagging oligonucleotide is used at the same time to enable the circularization of linear single-stranded DNA in a ligation reaction, and such an oligonucleotide remains attached to the "tester" at later manipulation steps.

10

In yet another embodiment, tagging oligonucleotides bind to regions in the tester backbone, which are flanking the inserted DNA fragment. The tagging oligonucleotides can be used to mask regions within the tester, and the "tester", which could otherwise hybridize with related sequences within the "driver", may become partially masked.

15 This embodiment of the invention is in particular of value when the "driver" comprises of RNA synthesized *in vitro* using a template. Therefore, the tagging oligonucleotides can serve a similar function as the removal of related sequences within the "driver" using an RNase.

20 To perform the invention, the "tester" is brought into contact with a "driver" which comprises single-stranded nucleic acids. Preferably, the "driver" includes single-stranded RNAs, and such RNA molecules can be prepared by any method known to a person skilled in the art. Such RNA can be prepared from a biological sample by a standard method including, but not limited to, those described by Sambrook J and
25 Russell DW, *ibid*, or they may be chemically synthesized *in vitro* by a standard method including, but not limited to, those disclosed in US patent US 6,649,751, which is hereby incorporated herein by reference, or they can be prepared *in vitro* in an enzymatic reaction using an RNA polymerase including, but not limited to, the use of T3, T7, or SP6 RNA polymerase. As such enzymes are commercially available,
30 conditions are described in the public domain including, but not limited to, those given by Sambrook J and Russell DW, *ibid*.

Preferably, the “driver” is derived from naturally occurring RNA prepared from a biological sample, an organism, or an infectious particle. Even more preferably, the “driver” is prepared from a given cDNA library or cDNA clone derived from such a library, and the vector hosting the library may comprise binding sites to initiate
5 transcription by an RNA polymerase.

RNA transcripts derived from a vector and prepared *in vitro* most commonly include within their sequences some regions which are derived from the vector and unrelated to the cDNA insert. In a more preferable application of the invention, such regions within
10 RNA transcripts reflecting vector sequences are removed from the “driver”. The removal of regions within RNA molecules can be performed by any method known to a person skilled in the art, including, but not limited to, the use of RNase H. This is shown in Fig. 5. In this embodiment of the invention, oligonucleotides having complementary sequence to the regions which are to be removed from the RNA
15 molecules are hybridized to the RNA molecules to form double-stranded regions comprising a stretch of RNA and the oligonucleotide. Then the RNA part within such double-stranded regions can be digested using RNase H, which cleaves RNA within RNA/DNA hybrids. Conditions for this enzymatic reaction are known to a person skilled in the art, including, but not limited to, those described by Sambrook J and
20 Russell DW, *ibid*.

The aforementioned single-stranded “tester” and the single-stranded “driver” can interact with one another in a hybridization reaction. Single-stranded DNA presented as the “tester” is brought into contact with the “driver” RNAs. To allow for a specific
25 hybridization between DNA molecules presented as the “tester” and RNA molecules presented as the “driver”, the DNA and the RNA molecules should preferably have sense and antisense orientation to one another. The orientation of naturally occurring RNA is often known and as such available in the public domain. Such information can be found for example in databases including, but not limited to, the DNA Data Bank of
30 Japan or DDBJ (<http://www.ddbj.nig.ac.jp/>), the National Center for Biotechnology Information or NCBI (<http://www.ncbi.nlm.nih.gov/>), or the European Bioinformatics Institute or EMBL-EBI (<http://www.ebi.ac.uk/index.html>). Otherwise the orientation of

RNA can be determined by experimental methods known to a person skilled in the art. Most often the orientation of RNA is derived from the related cDNA which has been cloned into a library. Similarly, the sequence derived from a cDNA clone can allow for the determination of the orientation as related to a naturally occurring species. Thus, the
5 necessary information for the design of the single-stranded "tester" and the "driver" can be obtained from partial sequence information or knowledge about the manipulations performed to make such materials. However, the invention is not limited to the use of defined sense and antisense pairs, as the invention can also be performed with mixtures in which the "tester" and/or "driver" comprise the sense and antisense strands for some
10 or all of the nucleic acid molecules.

A set of nucleic acid molecules comprising at least one molecule (the "driver") is put into contact with a plurality of nucleic acids comprising one or more distinct nucleic acid molecules (the "tester"). Preferably, the "tester" and the "driver" comprise
15 complex mixtures of nucleic acids out of which related nucleic acid molecules having common stretches of sequences in parts can be isolated in accordance with the present invention. Thus, the invention relates to a procedure for selecting specific nucleic acids by bringing two samples, the "tester" and the "driver", into contact. Depending on the experiment, the samples presented as the "tester" and the "driver" can have the same
20 origin or can be derived from distinct samples. Thus, the invention allows for the selection of nucleic acids from the same sample that contains related nucleic acid molecules having common stretches of sequences in parts or such nucleic acid molecules derived for distinct samples. Here, the invention offers new possibilities to obtain and analyze related nucleic acids derived from different biological materials or
25 conditions. Furthermore, such biological samples can be applied as isolated samples or in pools which comprise materials from different biological samples. According to the present invention, any sample may be used either as "tester" or as "driver", and terms "tester" and "driver" may be used interchangeably.

30 The hybridization of single-stranded DNA and RNA can be performed under any condition known to a person skilled in the art including, but not limited to, conditions described by Sambrook J and Russell DW, *ibid.* In a preferable application of the

invention the hybridization is performed using an excess of the "driver" over the "tester". Different conditions for the stringency of the hybridization can be calculated or designed by a person skilled in the art including, but not limited to, the use of CoT values.

5

A variety of hybrids that may be formed by the above-mentioned hybridization of the driver and tester are shown in Fig. 6.

10 The aforementioned DNA/RNA hybrids formed by nucleic acid molecules originating from the "tester" and "driver" are further processed in accordance with the invention. In the initial treatment, end-sequences in RNA molecules not matching to complementary sequences within the associated DNA are removed by an RNase or an exonuclease which at the same time will destroy free RNA having no matching DNA. Here, any RNase or exonuclease known to a person skilled in the art having specificity
15 for ribonucleic acid can be used, and examples include exonuclease VII. Such an exonuclease would preferentially digest RNA rather than DNA, and furthermore such an enzyme would digest ribonucleic acid only from the free ends and would not digest ribonucleic acids within the molecule. Such an exonuclease also would not digest RNA portions masked by hybridization with DNA. Further, it can be preferable if such an
20 exonuclease can digest linear single-stranded DNA and free RNA present in the plurality of nucleic acids at this stage.

In the following reaction step, DNA/RNA hybrid molecules from which RNA overhangs with free ends have been removed by a treatment with a RNase or
25 exonuclease are then subjected to filling-in of single-stranded DNA regions which are not masked by RNA associated to the DNA as shown in Fig. 7. This step can be performed by a DNA polymerase and can be primed by the RNA or by oligonucleotide primers (tagging oligonucleotides). Such oligonucleotides priming will lead to the filling-in of remaining single-stranded DNA portions having no matching RNA attached
30 to them. Any DNA polymerase known to a person skilled in the art can be applied for this fill-in reaction including, but not limited to, the Klenow fragment of DNA polymerase I, T4 and T7 DNA polymerases, DNA polymerase I, Vent, or Deep Vent

DNA polymerase, *Taq* polymerase, *Tfl* DNA polymerase, *Tth* DNA polymerase, *Tli* DNA polymerase. Preferably, this step is performed using the T7 DNA polymerase, Vent, or Deep Vent DNA polymerase, as those DNA polymerases have no strand-displacement activity.

5

After having performed the aforementioned reaction steps on the DNA/RNA hybrids, the resulting hybridization mixture should include only double-stranded nucleic acid molecules in which regions distinct in their sequences would give rise to the formation of loop structures containing stretches of single-stranded nucleic acids. For hybrid molecules formed by individual molecules having an entirely complementary sequence to each other, the resulting hybrid molecule would consist over its entire length of double-stranded DNA after the filling-in of single stranded regions. Similarly, nucleic acid molecules belonging to the "tester" to which no matching RNA could hybridize during the hybridization reaction would be subjected to the fill-in reaction as primed by the oligonucleotides added for the fill-in reaction and result in completely double-double stranded DNAs. Finally, single-stranded RNA molecules belonging to the "driver" having no matching sequence to any DNA molecules within the "tester" have been removed during the RNase treatment. Therefore, single-stranded nucleic acid regions are found only in those DNA/RNA hybrids which are composed of nucleic acid molecules having common stretches of sequences in parts along with unrelated sequences. Thus, such DNA/RNA hybrids comprising unpaired and thus single-stranded regions can be used for the selective enrichment of nucleic acid molecules having both related and unrelated partial sequences.

25 The aforementioned DNA/RNA hybrids comprising unmatched regions are then forwarded to a selection step, where nucleic acid molecules having unpaired regions are enriched over nucleic acid molecules comprising only double-stranded DNA or RNA and DNA. This step is shown in Fig. 8. After performing the aforementioned reactions for the removal of RNA overhangs and the filling-in of single-stranded regions on the DNA/RNA hybrids, only those DNA/RNA hybrids having regions, which are distinct between the DNA and the RNA, will contain in part single-stranded regions. Thus, only hybrids formed by related molecules from the "tester" and the "driver" having common

30

stretches of sequences in parts as well as unrelated sequences will give rise to hybrid molecules having loop structures of single-stranded RNA or DNA flanked by double-stranded regions. Thus, such DNA/RNA hybrids having loop structures of single-stranded nucleic acids can be selectively enriched using a substance that binds to a single-stranded nucleic acid. Such a single-stranded nucleic acid binding substance has a higher or even much higher affinity for single-stranded nucleic acids compared to that for double-stranded nucleic acids. Preferably, such a single-stranded nucleic acid binding substance can bind equally well to RNA and DNA. However, in a different embodiment of the invention, the single-stranded nucleic acid binding substance can have a clear preference for single-stranded DNA. The single-stranded nucleic acid binding substance may well have a higher or even much higher affinity for single-stranded DNAs compared to that for the single-stranded RNAs. In just another different embodiment of the invention, the single-stranded nucleic acid binding substance may well have a clear preference for single-stranded RNA. The single-stranded nucleic acid binding substance can have a higher or even much higher affinity for single-stranded RNAs compared to that for the single-stranded DNAs.

The present invention provides a method for distinguishing between unrelated sequences belonging to the "tester" (DNA) or the "driver" (RNA).

20

Any single-stranded nucleic acid binding substance known to a person skilled in the art can be used. In accordance to the aforementioned options in the experimental design on how to apply the invention, such single-stranded nucleic acid binding substances can have preference for the binding to either single-stranded DNA or RNA, or can bind to both of them with similar affinity. Many proteins having affinity to single-stranded DNA are known, including, but not limited to, SSB from *E. coli*, the product of the phage T4 Gene 32, the adenovirus DBP, an antibody directed against single-stranded DNA, calf thymus UP1, or any mixture thereof. SSB from *E. coli* is commercially available from various providers including, but not limited to, Stratagene, La Jolla, USA (Cat. No. 600201), Promega, Madison, USA (Cat. No. M3011), Amersham Biosciences, Cardiff, United Kingdom (Cat. No. E70032Y), and Epicentre, Madison, USA (Cat. No. SSB02200). Other single-stranded-DNA binding proteins can be found

30

in the public domain, including, but not limited to,, the product of the phage T4 Gene 32. The product of the phage T4 Gene 32 is commercially available from various providers including, but not limited to,, Nippon Gene, Tokyo, Japan (Cat. No. 312-03251), USB, Cleveland, USA (Cat. No. 74029Y) and Amersham Biosciences, Cardiff, United
5 Kingdom (Cat. No 25003911). In addition, autoantibodies against single-stranded DNA are found frequently in patients with non-rheumatic diseases including chronic active hepatitis and infectious mononucleosis. Such autoantibody can be purified by affinity-purification on a DNA matrix or obtained by immunization of an animal. Such antibodies can further be obtained in the public domain for diagnostic purposes, e.g., in
10 enzyme immunoassays. A human antibody against single-stranded DNA is commercially available from various providers including, but not limited to,, Immunovision, Springdale, USA (Code HSS-0100). Single strand nucleic-binding proteins have further been disclosed for example in EP 1041160 A1, which is hereby incorporated herein by reference, or other single-stranded nucleic acid binding
15 substances are disclosed in EP 0622457 A1, which is hereby incorporated herein by reference. Thus the invention relates to the selection of specific DNA/RNA hybrids having stretches of single-stranded DNA.

Similarly, many proteins are known to a person skilled in the art that bind specifically to
20 single-stranded RNAs. For the correct processing of transcripts in cells, mRNAs are associated with RNA binding proteins that influence or control the pre-mRNA processing including steps like splicing, localization, transport, and stabilization as described by Dreyfuss G et al, Nat. Rev. Mol. Cell Biol. 3 (2002) 195-205, which is hereby incorporated herein by reference. Although presently not commercially
25 available, any such protein having a high affinity for single-stranded RNA over double-stranded nucleic acids can be applied to perform the invention. Thus, the invention relates to the selection of specific DNA/RNA hybrids having stretches of single-stranded RNA.

30 In a preferable embodiment of the invention, the single-stranded nucleic acid binding substance can bind to RNA and DNA. In an even more preferable embodiment, such a single-stranded nucleic acid binding substance can bind to RNA and DNA with similar

affinity. Thus, such a single-stranded nucleic acid binding substance can be applied to enrich hybrid molecules comprising unpaired DNA or RNA portions regardless of whether such unpaired regions are RNA or DNA. A number of proteins are known to a person skilled in the art, which can bind to single-stranded DNA and RNA, including, but not limited to, nucleocapsid proteins or NC proteins which are small and highly basic proteins found to be associated with genomic RNA in retroviral particle or the related viral DNA transcripts (Lapadat-Tapolsky M et al., Nuc. Acid Res. 21 (1993) 831-839, hereby incorporated herein by reference). NC proteins have the ability to catalyze the folding of nucleic acids into preferable configurations for which they have to interact with single-stranded nucleic acids. Such proteins have been used in different reactions to eliminate secondary structures in RNA. Although those proteins are of value to perform the invention, it has to be noted that often they can also bind to double-stranded DNAs and may not be suitable for the purpose of the present invention, as they may cause a high background in the resulting libraries. Alternatively, mixtures of the aforementioned single-stranded DNA and the single-stranded RNA binding proteins can be used. Thus, any such mixture of individual proteins or groups of proteins could be applied for the enrichment of molecules having loop structures made up of either RNA or DNA.

Alternatively, the single-stranded nucleic acid binding substance may be a mixture of oligonucleotides, and such oligonucleotides have random sequences, preferably a random oligonucleotide of 15 to 30 nucleotides, or even more preferably of 25 nucleotides (it may be indicated as "25N", where N stands for any possible base within a RNA or DNA molecule). Thus, such a randomized oligonucleotide mixture comprising RNA or DNA can hybridize to any given sequence in an unpaired region independently whether or not such unpaired regions are DNA or RNA. Therefore, a mixture of randomized oligonucleotides is a preferable means to perform the invention.

Single-stranded nucleic acid-binding substances are preferably bound to a tag molecule or a matrix to be applied to perform the invention. A tag molecule may be selected from biotin, digoxigenin, antibody, antigen, a protein and nucleic acid binding molecule. The single-stranded nucleic acid binding molecule associated with a tag molecule may

be recovered by using a matrix. For the purpose of the present invention, a matrix may be selected from any immobilized form of avidin, streptavidin, digoxigenin-binding molecule, an antibody and its ligand and/or chemical matrix. If the applied tag is biotin, then the related matrix may be avidin or streptavidin; similarly, when the tag is digoxigenin, the matrix may be a digoxigenin-binding molecule (see Roche Diagnostics GmbH Catalog, the documentation therein is hereby incorporated herein by reference), or when the tag is an antigen, the matrix may be a different antibody or an antibody-binding protein such as protein I or protein G. The above lists of tags and matrices are, however, not limiting nor exhaustive, since a person skilled in the art knows other combinations of tags and associated matrices. Furthermore, a single-stranded nucleic acid binding substance can be directly bound to a matrix by a chemical reaction, thus making the use of a tag unnecessary. In one such embodiment, randomized oligonucleotides can be directly synthesized on the surface of a matrix such as glass, and then directly applied to the selection step. Similarly, oligonucleotides can be synthesized with an amino group at their 5'-ends, which can be used for covalent binding to a matrix in a chemical reaction.

The recovery of the desired DNA/RNA hybrids partially having regions of unrelated sequences is preferably carried out when the single-stranded nucleic acid binding substance is conveniently associated to a solid matrix. Such a solid matrix may include, but is not limited to, the commonly used modified surfaces such as metal beads, magnetic beads, inorganic polymer beads, organic polymer beads, glass beads, and agarose beads. Inorganic polymers include silica, ceramics, and the like. Organic polymers include polystyrene, polypropylene, polyvinyl alcohol, and the like. Metals include iron, copper, alloys such as stainless steel and the like. Further examples for such tags, matrices and surfaces are disclosed in EP 0622457 A1, which is hereby incorporated herein by reference.

To perform the invention, the selection step may preferably be performed using randomized oligonucleotides immobilized by binding to a matrix. Hybrids of DNA/RNA isoforms comprising unpaired regions are isolated in this way from other hybrids not having unpaired regions and are recovered by being released from the single

strand nucleic acid-binding molecule according to standard methodologies, such as heating, for example, at 40 to 60°C, preferably 50°C. If random oligonucleotides are used as the single-stranded nucleic acid binding substance, a heating up to 50°C is usually sufficient for releasing the hybrids having unpaired regions from the single-
5 stranded nucleic acid binding substance.

After the selective enrichment of DNA/RNA hybrids having unpaired regions, the remaining RNA in such hybrid molecules can be removed using an RNase or a chemical treatment. This is shown in Fig. 9. The RNA portion of DNA/RNA hybrids
10 obtained by the aforementioned selection step is preferably removed from such hybrids by treatment with an RNase, although a person skilled in the art knows other methods which are well within the scope of the invention. Any RNase known to a person skilled in the art can be used, and such an RNase should have a much higher enzymatic activity toward RNA as compared to DNA. Thus, such an activity would specifically digest the
15 region made of RNA whereas regions made of DNA would remain. In a preferred embodiment of the invention, the RNase could be RNase I or RNase H. In an even more preferable embodiment a mixture of RNase I and RNase H would be applied at this step. In just a different embodiment, treatment by alkali removes regions comprising RNA from nucleic acid molecules made of RNA and DNA.

20

Nucleic acid molecules obtained after the aforementioned RNase treatment may be made up of DNA, and such molecules have regions composed of double-stranded DNA as well as those composed of single-stranded DNA. Such molecules can be used for the direct cloning of the selected molecules by standard approaches known to a person skilled in the art or further described by Sambrook J and Russel DW, *ibid.* In one
25 embodiment of the invention the partially double-stranded and partially single-stranded DNA molecules obtained by the aforementioned RNase treatment are transferred into *E. coli* to allow for their cloning into libraries. More preferable single-stranded regions within such molecules can be filled-in by means of a DNA polymerase before the
30 transformation step. Any DNA polymerase known to a person skilled in the art can be applied for the fill-in reaction including, but not limited to, the Klenow fragment of DNA polymerase I, Vent, or Deep Vent DNA polymerase, T4 and T7 DNA

polymerases, DNA polymerase I, Taq polymerase, Tfl DNA polymerase, Tth DNA polymerase, Tli DNA polymerase. Preferably, this step is performed using the T7 DNA polymerase, Vent, or Deep Vent DNA polymerase, as those DNA polymerases displays no strand displacement activity. Even more preferably, the partly single-stranded molecules may be directly transferred into *E. coli*, and the replication machinery of the host performs the reconstitution of the double-stranded DNA. If the DNA molecules provided as the "tester" include all necessary genetic elements for the replication of the DNA molecules within the host, no further cloning steps are required to perform the invention. Thus, the invention provides a method for the cloning of dimeric nucleic acid molecules comprising stretches of common sequences and at least one region having unrelated sequence, and such molecules are selectively enriched from a plurality of nucleic acids and can be further converted so that one of the two strands within the dimeric nucleic acid molecule can be isolated. Thus, the invention provides a method for the isolation of nucleic acid molecules which partially share homologous sequences and which have at least one region of unmatched sequence.

As the invention can be performed with a plurality of nucleic acid molecules, the resulting libraries may include a plurality of nucleic acids. Individual clones derived from such libraries can be isolated by any method known to a person skilled in the art including, but not limited to, the use of robotic systems. Individual clones obtained from the aforementioned library can be further analyzed by various means including, but not limited to, DNA sequencing, hybridization-based approaches, approaches based on digestion with sequence specific endonucleases or other standard approaches as known to a person skilled in the art.

Furthermore, the "tester" and the "driver" may comprise single-stranded DNAs. In such an embodiment of the invention, the "driver" can be prepared by any of the aforementioned approaches for the preparation of single-stranded linear DNAs. After hybridization of the "driver" to the "tester", overhangs of single-stranded DNA derived from the "driver" are removed by means of an exonuclease, including, but not limited to, exonuclease VII. The fill-in reaction of the vector backbone can be performed by any of the DNA polymerases named above, and such a DNA polymerase should lack any

strand-displacement activity. The selection of molecules comprising loop structures can be performed by any of the aforementioned methods. After the selection step, the regions made up of single-stranded DNA are removed using T4 endonuclease VII, which specifically cuts at junction points. This enzyme can be obtained from USB, catalog number 78333 or 78300, and the product description thereof is hereby
5 incorporated herein by reference. In a different embodiment of the invention, the loop structures are not removed from the hybrid molecules, but the two strands are separated by denaturation, filled-in by the use of specific primers and a DNA polymerase, and transferred into *E. coli* for amplification. Thus, the invention relates to a method for
10 using DNA/DNA hybrids in the selection of nucleic acid molecules partially having unpaired regions.

In just a different embodiment of the invention, a selective binding substance may be used to label the "driver". Such selective binding substance includes, but is not limited
15 to, biotin and digoxigenin. Thus, the invention relates to a stepwise selection of nucleic acid molecules belonging to the "tester". In one selection step, the selective binding substance attached to the "driver" is used to select all nucleic acid molecules within the "tester" having complementary sequences within the "driver". In the other selection
20 step, nucleic acid molecules partially having regions of complementary sequence along with at least one region of unrelated sequence are selected using a single-stranded nucleic acid binding substance. Such a single-stranded nucleic acid binding substance can be labeled by a selective binding substance different from those belonging to the "driver". Examples for the parallel use of two different labels could include, but are not
25 limited to, the use of biotin and digoxigenin. Thus the invention relates to a method for the subtraction of a plurality of nucleic acids and/or the enrichment of related nucleic acid molecules, and such steps can be performed using different selective binding substances in the two selection steps.

In a different embodiment of the invention, the remaining molecules within the "tester"
30 which have not been selected during the selection step using a single-stranded nucleic acid binding substance are to be cloned. In such an embodiment, free DNA/RNA hybrid molecules are subjected to RNase treatment to remove the RNA portion

followed by transformation of *E. coli*. Thus, the invention allows for the selection of nucleic acid molecules which lack certain regions but otherwise share a related sequence with each other.

5 Even further, the present invention relates to the selection of nucleic acid molecules within a plurality of nucleic acid molecules lacking certain sequence elements with respect to each other. In such an embodiment of the invention, probes may be designed for "driver" preparation, and each molecule in the driver has a region which should not be present in the target molecules as well as flanking regions related to regions which
10 are present in the target molecules. Such nucleic acids belonging to the "driver" can form hybrids with target molecules belonging to the "tester" having loop structures. Such loop structures would give rise to the removal or enrichment of such structures by means of binding to a single-stranded nucleic acid binding substance or specific binding molecules. Thus, the invention relates to a method for the selection or removal
15 of nucleic acid molecules having unrelated or undesired sequence elements. In one such embodiment, the invention can be used to remove intron-containing clones from a clone collection. In a more preferable embodiment, the invention may be used to select clones which do not have certain exons. Thus, the invention provides a method for driving the selection for new splice variants, and the "driver" can be designed to drive the
20 enrichment or removal of specified nucleic acid molecules.

In just a different embodiment, the invention provides a method for analyzing two or more samples for their content of related nucleic acids which have related and unrelated regions in parts within their sequences. For each sample, a "tester" and a "driver" are
25 prepared in accordance with the invention, and in the following steps the "driver" from the first sample is brought into contact with the "tester" from the second sample, whereas the "driver" from the second sample is brought into contact with the "tester" of the first sample. Thus, the invention can be applied in both directions, where a sample allows for the preparation of a "tester" and a "driver" at the same time. Similarly, the
30 "tester" and the "driver" from the same sample can be brought into contact to search for related nucleic acids which partially have related and unrelated regions within their sequences and within the same sample. Even furthermore, different "testers" and

different “drivers” of the same or different origin can be mixed before performing the invention to cover a larger variety or complexity within the same experiment.

5 In another embodiment, the invention provides a method for the isolation of individual clones from a plurality of nucleic acids prepared in accordance with the invention, and such clones are forwarded to isolation of plasmid DNA. The isolated plasmid DNA can then be digested with one or more restriction endonuclease(s), and in a preferable embodiment such an endonuclease would cut in the flanking regions of the inserted DNA fragment. Thus, the invention provides a method for the determination of the
10 insert size of the nucleic acids isolated according to the invention.

Furthermore, the invention provides a method for the analysis of the nucleic acid molecules prepared according to the invention, in which such molecules are subjected to partial or full-length sequencing. Such sequence information can be obtained by any
15 approach known to a person skilled in the art, including, but not limited to, end-sequencing or full-length sequencing by primer walking, shotgun approaches or random transposon integration. Thus, the invention comprises a method for the computational analysis of sequence information related to nucleic acid molecules prepared according to the invention.

20

In a different embodiment of the invention, nucleic acid molecules prepared in accordance with the invention are analyzed by means of hybridization to reference samples that comprise nucleic acid molecules having sequences homologous to the nucleic acid molecules prepared according to the invention. Thus, the knowledge about
25 the nucleic acid molecules in the samples can be used for the annotation of nucleic acid molecules which hybridize to molecules in the sample.

In such an embodiment, the invention provides a method for isolating DNA inserts from clones prepared according to the invention, comprising the steps of fragmenting such
30 inserts isolated from the clones to form short nucleic acid molecules, and subjecting such short nucleic acid fragments to hybridization or sequencing experiments.

In still a different embodiment of the invention, the reference samples used in a hybridization experiment can be immobilized on a solid matrix. Nucleic acids prepared according to the invention can be analyzed by hybridization to a micro- or macroarray. In one embodiment, such an array may comprise oligonucleotides or nucleic acid
5 fragments. In a different embodiment, the oligonucleotides or nucleic acid fragments have sequences related to exons. In yet another embodiment, the array may comprise oligonucleotides comprising genomic sequences. In a preferable embodiment, the oligonucleotides comprise genomic sequences and are arranged to form a tiled array covering larger regions within genomic sequences, such as described by Cawley S et al.,
10 Cell 116 (2004) 499-509, which is hereby incorporated herein by reference.

Furthermore, in the hybridization reaction, cDNA inserts obtained according to the invention are transcribed into RNA, in which such RNA transcripts may be subjected to labeling. Such labeled RNA is then applied in the hybridization experiments, and the
15 labeled RNA is brought into contact with the DNA fragments or oligonucleotides on the array. As the RNA portion in the resulting RNA/DNA hybrids could be removed after signal detection by an RNase, the DNA probes on the microarray may be used in repetitive hybridization experiments. Thus, the invention provides a method for analyzing many inserts prepared according to the invention for their entire sequence
20 content by hybridization to an array in a high-throughput reaction, and in a cycling process individual inserts are hybridized to the array, signals are obtained by a reader, and RNA is removed using an RNase before the cycle can be repeated. In yet another embodiment, the nucleic acid molecules presented on the array are chemically modified to protect them against degradation. Many forms of chemical modifications including,
25 but not limited to, modification to phosphothioates are known to a person skilled in the art for the preparation of modified oligonucleotides which can be applied to arrays by standard approaches known in the field. Thus, the invention provides a method for the design of an apparatus for the analysis of related nucleic acid molecules which in part comprise unrelated regions by means of hybridization to reference samples.

30

Even furthermore, the invention provides a method for the analysis of nucleic acid molecules prepared according to the invention using endonucleases having specific

recognition sites. Thus, the fragmentation of a nucleic acid molecule by such an endonuclease can produce fragments of specific length characteristic for any given nucleic acid molecule. Thus, the invention provides a different method for analyzing nucleic acid molecules that may reflect alternative exon usage.

5

The invention relates to the cloning of related nucleic acid molecules in which such nucleic acid molecules can comprise naturally occurring sequences. As such nucleic acid molecules relate to naturally occurring species, the invention provides a method for analyzing and cloning naturally occurring nucleic acid molecules that derived from the same gene but are distinct by their exon usage. Thus, the invention provides approaches for isolating and cloning alternatively spliced gene products.

10

In a different embodiment, the invention provides a method for selectively cloning alternatively-spliced transcripts that are cloned according to the invention without the need of fragmentation. Thus, the invention relates to handling and isolating molecules comprising the entire length of transcribed molecules.

15

Because the invention provides a method for cloning alternatively spliced transcripts or clones derived thereof which comprise full-length molecules, the invention provides a method for the functional analysis of differentially or alternatively spliced transcripts and the products encoded by them. Thus, the invention relates to a method for functional assays on proteins derived from alternatively spliced transcripts. In one embodiment, such a protein is used as a drug. In a different embodiment, such a protein is used as an antigen to raise antibodies against such a protein. Such antibodies can be used as a drug or for diagnostic purposes. In yet another embodiment, such a protein has an enzymatic activity and can be used to perform a chemical reaction. In still another embodiment, the sequence encoding such a protein is used to prepare a DNA or RNA molecule to be used in therapy or diagnostics.

20

25

As the present invention provides a new approach for the isolating, cloning and analyzing related nucleic acid molecules partially having common sequence information, fundamental restrictions in the state of the art for the analysis and use of genes and

30

transcripts can be overcome, in which such genes and transcripts are subjected to regulation on the level of alternative splicing. Thus, the invention allows addressing entirely new questions on how alternatively or differentially spliced transcripts contribute to biological phenomena including, but not limited to, diseases and public health. Furthermore, the invention provides new methods for isolating such transcripts, in which differentially spliced transcripts would encode for proteins of different activity. Therefore, the invention opens up new areas for commercial applications in the field including, but not limited to, the target identification for diagnostics, targets for drug discovery or to be used directly as medications, and the preparation of proteins having distinct enzymatic functions. Similarly, the invention can be applied to design a kit comprising necessary reagents to perform the methods of present invention in research and development. Thus, the invention can be applied in a wide area of commercial interest and beyond, and will surely contribute to the future development of the field.

Example 1 – Isolation of RNA

Samples of mRNA or total RNA can be prepared by standard methods known to a person skilled in the art of molecular biology as for example given in more detail in Sambrook J and Russel DW, *ibid.* Furthermore, Carninci P et al. (*Biotechniques* 33 (2002) 306-309, hereby incorporated herein by reference) describe a method to obtain cytoplasmic mRNA fractions. Although the use of cytoplasmic RNA is preferable for the analysis of alternative splicing and related processes, any other approach for the preparation of mRNA or total RNA can be used to obtain similar results.

The preparation of mRNA from total RNA or cytoplasmic RNA is preferable but not essential in performing a method of the invention, as the use of total RNA can provide satisfying results in combination with the Cap-selection step performed during full-length cDNA library preparation. Here, we have commonly used the Cap-trapper approach, which effectively removes ribosomal RNA from library preparations. Generally speaking, mRNA represents about 1-3 % of total RNA preparations, and it can be subsequently prepared by using commercial kits based on oligo dT-cellulose matrixes. Such commercial kits including, but not limited to, the MACS mRNA

isolation kit (Miltény) provided satisfactory mRNA yields under the recommended conditions when applied for the preparation of mRNA fractions. One cycle of oligo-dT mRNA selection is sufficient as extensive mRNA purification can cause a loss of long mRNAs.

5

All mRNA samples used to perform the invention were analyzed for their ratios of the OD readings at 230, 260 and 280 nm to monitor the RNA purity. Removal of polysaccharides was considered successful when the 230/260 ratio was lower than 0.5 and an effective removal of proteins was obtained when the 260/280 ratio was higher than 1.8 or around 2.0. The RNA samples were further analyzed by electrophoresis in an agarose gel to prove a good ratio between the 28S and 18S rRNA in total RNA preparations, and to show the integrity of the RNA fractions.

Example 2 - cDNA library preparation

15

For the purpose of this example, full-length cDNA libraries were constructed as described by Carninci P and Hayashizaki Y, *ibid*. This approach makes use of the Cap-trapper approach for full-length cDNA cloning. DNA fragments were cloned into the phage/vector system pFLC, as disclosed in patent application WO 02/070720 A1, which is hereby incorporated herein by reference. Phage solutions prepared to perform the invention were stored in medium containing 7% DMSO and kept at -80°C . However, the invention is not limited to the aforementioned procedure for library preparation, as a person skilled in the art knows other methods for the preparation of full-length selected libraries.

25

For the purpose of this example, full-length cDNA libraries from melanocyte (melan-c), denoted as "G2", and melanoma (B16-F10Y), denoted as "G3", cell lines were prepared by Cap-trapping described above in the following Examples.

30 Example 3 - Plasmid DNA excision from amplified phage

The Cre/lox P recombination system was used for the excision of plasmid DNA from the phage library *in vivo*. For *in vivo* excision of the plasmid, *E. coli* BNN132 stocks can be obtained from CLONTECH (nowadays BD Biosciences) as part of their Lambda TriplEx™ Phagemid Cloning Vector (cat. # 6160-1 and 6161-1). Alternatively, the strain BM25.8 can be obtained from Novagen as part of their Lambda BlueSTAR™ Cloning System, where both strains express an active Cre recombinase.

E. coli BNN132 cells as taken from a glycerol stock were spread out on an agar plate with LB medium and Kanamycin (20 µg/mL) and incubated over night at 37°C. An individual colony was taken from the plate and used for inoculation of a 100 mL LBMM medium (LB medium enriched with 10 mM MgSO₄, 0.2% Maltose) culture. The culture was grown under vigorous shaking at 37°C until an OD₆₀₀ of 0.5 was achieved. After transfer of the culture into two 50 mL plastic tubes, bacteria were harvested by centrifugation at 4,000 rpm for 4 min at 4°C. The supernatant was discarded after the centrifugation step, and the pellet was dried by keeping the tube upside-down on a paper towel for about 1 min. After addition of 5 mL of 20 mM MgSO₄ to each tube, and mild vortexing of the tubes to re-suspend the pellets, both suspensions were united in one tube (total volume about 10 ml), and again mixed until no clumps remained in the suspension. For preparation of the phage infection, the cell suspension was pre-incubated at 37°C for 5 min in a water bath.

The phage DMSO stocks were thawed on ice, and phages were collected by centrifugation with Tomy MX-200 at 4°C for a few seconds. About 1x10¹⁰ pfu of a phage stock was generally applied per experiment and transferred into a new 1.5mL tube, to which the aforementioned BNN 132 cells were added. After incubation of the cells at 37°C under gently shaking (at 60 rpm) for 20 min for infection, the infected BNN 132 cells were transferred into a 500 ml Erlenmeyer flask containing 90 mL LB medium with 125 µg/ml Ampicillin and incubated under shaking (about 150 rpm) at 30°C for 1 hour for amplification of the plasmid DNA.

Alternatively, the plasmid DNA can be excised *in vitro* by the use of a commercially available Cre recombinase. To perform the plasmid excision *in vitro*, phage DNA was

prepared by a phage DNA extraction kit from Promega ("Wizard[®] Lambda Preps DNA Purification System", cat# A7290) according to the manufacture's instruction, which is hereby incorporated herein by reference. For the conversion, the phage DNA should be excised applying a Cre-recombinase, such as the Cre-recombinase from New England
5 BioLabs (cat# M0298S) or alternatively the Cre recombinase from Novagen (cat#69247-3) or CLONTECH (cat# 631614). The enzymatic reaction can be performed according to the manufacture's instructions, and the resulting plasmid DNA can be amplified by retransformation into a host according to standard procedures such as those described by Sambrook J and Russell DW, *ibid*.

10

Example 4 - Plasmid DNA purification

Plasmid DNA purification from aforementioned cultures was performed by using a "Wizard[®] Plus Midipreps DNA Purification System" from Promega (cat #A7640).

15 DNA as isolated by the means of the kit was further purified using a S400 column (Amersham-Pharmacia) by centrifugation at 3000 rpm for 1 min at 4°C in a buffer containing TE. The final volume of the plasmid DNA preparation was adjusted to 50 µl with H₂O.

20 The plurality of plasmid DNA obtained was characterized by digestion with the restriction endonuclease PvuII to measure the size of the cDNA inserts by gel electrophoresis. Similarly, the insert sizes of the libraries were determined in the same way.

25 Example 5 – PCR amplification of cDNA inserts

The aforementioned libraries G2 and G3 were prepared using RIKEN vector pFLCII, and primers were designed for PCR amplification of the cDNA inserts in a way that they comprise binding sites for T3 or T7 RNA polymerase as indicated by their name.

30 PCR reactions were performed under the following conditions: 2.5 µl of each 10 µM of primer T3GW1 (SEQ ID NO: 1),

GAGAGAGAGAATTAACCCCTCACTAAAGGGACAAGTTTGTACAAAAAAGC

and T7GW2 (SEQ ID NO: 2),

GAGAGAGAGAATTAACCTCACTAAGGGACCACTTTGTACAAGAAAGC, or for
opposite direction: T3GW2 (SEQ ID NO: 3),

GAGAGAGAGAATTAACCTCACTAAGGGACCACTTTGTACAAGAAAGC and

5 T7GW1 (SEQ ID NO:4),

GAGAGAGAGTAATACGACTCACTATGGGACAAGTTTGTACAAAAAAGC,

template 4 µl (40ng), 2XGC buffer 50 µl, 2.5 mM dNTPs 16 µl, and H₂O 25 µl were

mixed for reaction setup. After a hot start at 95°C for 1 min, 1 µl LA Taq (5u/µl) were

added to initiate the reaction. The reaction was run with some 10-20 cycles of 95°C for

10 1 min, 55°C for 30 sec, and 72°C for 8 min. After completion of the reaction,

remaining polymerase was destroyed by Proteinase K (Qiagen) digestion, followed by

extraction with phenol/chloroform and chloroform, and ethanol precipitation. Purified

PCR products were dissolved in 100 µl of H₂O. Each of those steps was conducted

under standard conditions as for example described by Sambrook J and Russell DW,

15 *ibid.*

Example 6 – RNA synthesis

RNA was synthesized using either T3 or T7 RNA polymerase (Life Technologies),

20 depending on the orientation of the DNA inserts in the vector used for the driver
preparation. All experiments were designed to have sense run-off RNA within the

driver. Starting from the aforementioned PCR product (10 µl, equal to some 3 µg of

DNA), reactions were performed using following condition: 3 µl of T7 or T3 RNA

polymerase (50 u/µl) were added 40 µl of 5xT7/T3 buffer (Life Technologies), 20 µl of

25 0.1 M DTT, 1.6 µl of 10 mg/ml BSA, 10 µl of 10 mM rNTP (Nippongene), and 115.4 µl

of H₂O in a total volume of 200 µl. After incubation at 37°C for 5 hour, synthesized

RNA was purified by DNaseI (RQ1, RNase-free, Promega) treatment for about 15 min

by addition of 20 µl of 10 mM CaCl₂ and 1 µl of DNase (10 u/µl Units), and incubated

at 37°C. Samples were further diluted with 100 µl of H₂O and purified using a

30 QIAGEN RNeasy RNA purification Kit (Qiagen) according to manufacturer's

instructions, which is hereby incorporated herein by reference.

Similarly, *in vitro* synthesis of RNA can be used for the preparation of digoxigenin labeled RNA such as needed for subtraction experiments. Using a DIG RNA Labeling Kit available from Roche Diagnostics GmbH (Cat. No. 1 175 025), such reactions can be performed with the enzymes SP6 and T7 RNA polymerase. The handbook for the
5 above-mentioned kit is hereby incorporated herein by reference.

Example 7 - Removal end-sequences within RNA molecules

Regions common to the vector within the tester and the *in vitro* synthesized RNA were
10 removed by hybridization of short DNA oligonucleotides comprising such regions within the RNA. For RNA as transcribed from libraries cloned into the vector pFLCII, such oligonucleotides had the following sequences: Primer 5'-
AGGAGAGGTCTAGACCACTTTGTACAAGAAAGCTGGGTGGATCCGGACTGT
TTTT (SEQ ID NO: 5), Primer RNAII 5'-
15 CTCGACCTCGAGAGCCTGCTTTTTTTGTACAACTTGTGGCCCGGTACCC (SEQ
ID NO:6). The hybridization was performed in a buffer containing 80% formamide (from a deionized stock), 250 mM NaCl, 25 mM HEPES, pH 7.5, 5 mM EDTA at 42°C over 4hours. RNA/DNA hybrids were precipitated by ethanol under standard
conditions, and were then dissolved in 44 µl of water, 5 µl of 10X RNase H buffer and 1
20 µl of RNase H (New England Biolabs). The reaction was performed at 42°C for 4
hours. Remaining oligonucleotides and free nucleotides were removed by
chromatography on an S-400 column (Amersham-Pharmacia) according to the
manufacture's instructions, which are hereby incorporated herein by reference. RNA
was further purified under standard conditions for Proteinase K digestion,
25 chloroform/phenol extraction, and ethanol precipitation as described by Sambrook J and
Russell DW, *ibid*. Purified RNA was dissolved in 50 µl of H₂O, out of which 3 µl of
RNA were used for concentration measurement and quality control by agarose gel
electrophoresis.

30 Example 8 - Generation of single-stranded DNA by the means of Gene II and
Exonuclease III

For each tester preparation, mix in a 1.5 ml micro-centrifugation tube at room temperature 2 μ l of 10X Gene II Buffer (Invitrogen), 5 μ g double-stranded phagemid DNA (1 μ g/ μ l) and H₂O to a final volume of 19 μ l. The nicking reaction is initiated by addition of 1 μ l of Gene II (Invitrogen). The reaction mixture was mixed gently and
5 incubated at 30°C for 45 min. After completion of the reaction, the remaining enzymatic activity was heat deactivated at 65°C for 5 min, and immediately chilled on ice for 1 min. A 1 μ l aliquot of the mixture was transferred to a new micro-centrifugation tube containing 9 μ l of TE buffer and 2 μ l of gel loading dye was transferred to be used for analysis by agarose gel electrophoresis. To the remaining 19
10 μ l of the reaction mixture, 2 μ l of Exo III (Invitrogen) was added, mixed gently, and incubated at 37°C for 60 min. Single-stranded DNA was further purified under standard conditions for Proteinase K digestion, chloroform/phenol extraction, and ethanol precipitation as described by Sambrook J and Russell DW, *ibid*. Again, an aliquot was taken out for analysis by agarose gel electrophoresis as described above.
15 Agarose gel electrophoresis was performed under standard conditions as described by Sambrook J and Russell DW, *ibid*, using a 0.8% agarose gel in 1XTAE or TBE buffer.

Example 9 – Tagging of tester by hybridization to specific primers

20 Oligonucleotides were designed having matching sequence to Ori (Ori-F 5'-AATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGAC (SEQ ID NO: 7)), Ampicillin resistance marker (Amp-F 5'-CCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGC (SEQ ID NO: 8)), and the 5'- and 3'- flanking
25 regions of the inserts (FLCII-1 5'-GGGTACCGGGCCACAAGTTTGTACAAAAAAGCAGGCTCTCGAGGTCGAG (SEQ ID NO: 9), FLCII-2 5'-AAAAACAGTCCGGATCCACCCAGCTTTCTTGTACAAAGTGGTCTAGACCTCTCCT (SEQ ID NO: 10)) within vector pFLCII. The aforementioned single-stranded vector DNA was mixed with above-mentioned primers in a buffer containing 40%
30 formamide (from a deionized stock), 375 mM NaCl, 25 mM HEPES, pH 7.5, 50 mM EDTA, and precipitated by ethanol under standard conditions. Hybridization of the oligonucleotides to the single-stranded vector carried out at Cot values of 1 to 20 in a

buffer containing 40 percent formamide (from a deionized stock), 0.375 M NaCl, 25 mM HEPES (pH 7.5), and 2.5 mM EDTA at 42°C for 8 hours. After the hybridization reaction, the sample was precipitated by ethanol under standard conditions, washed twice with 70% ethanol, and finally the hybrids were re-suspended in 90 µl of H₂O.

- 5 Remaining primers were removed by chromatography on an S-400 column according to the manufacturer's instructions, and after concentration by ethanol precipitation, the probe was dissolved in 20 µl of H₂O.

Example 10 - DNA/RNA hybridization

10

For the hybridization of single-stranded DNA against RNA a DNA/RNA ratio as 1:40 was applied, where about 1 µg of DNA was hybridized with 40 µg of RNA. Hybridizations were carried out at Cot values of 1 to 20 in a buffer containing 40 percent formamide (from a deionized stock), 0.375 M NaCl, 25 mM HEPES (pH 7.5),
15 and 2.5 mM EDTA at 42°C for 14 hours. After the hybridization reaction, the sample was precipitated by ethanol under standard conditions, washed twice with 70% ethanol, and finally the hybrids were re-suspended in 90 µl of H₂O.

Example 11 – Treatment by Exonuclease VII

20

Exonuclease VII was used for the degradation of remaining single stranded nucleic acids and RNA overhangs in DNA/RNA hybrids. Digestion was performed by addition of 10X Exo VII buffer and 1 µl of Exonuclease VII (USB, 10 u/µl), and the reaction was incubation at 37°C for 60 min. DNA/RNA hybrids were further purified under standard
25 conditions for Proteinase K digestion, chloroform/phenol extraction, and ethanol precipitation as described by Sambrook J and Russell DW, *ibid.* The sample was finally re-suspended in 50 µl of 0.1X TE.

Example 12 – Second-strand DNA synthesis by the means of T7 DNA polymerase.

30

The second DNA strand of DNA/RNA hybrids was synthesized by the means of T7 DNA polymerase (USB). The aforementioned DNA/RNA hybrids 1 µg, and 1 µl (2 µg) of primers:

Ori-F 5'- AATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGAC (SEQ

5 ID NO: 11)

Amp-F 5'- CCTTCGGTCCTCCGATCGTTGTCAGAAAGTAAGTTGGCCGC (SEQ

ID NO: 12)

FLCII-1 5'-

GGGTACCGGGCCACAAGTTTGTACAAAAAAGCAGGCTCTCGAGGTCGAG

10 (SEQ ID NO: 13)

FLCII-2 5'-

AAAAACAGTCCGGATCCACCCAGCTTTCTTGTACAAAGTGGTCTAGACCTCT
CCT (SEQ ID NO: 14), and the reaction mixture was heated to 65°C for 5 min.

Simultaneously, 10X polymerase buffer, 8 µl of 2.5 mM dNTPs, 1 µl of T7 DNA
15 polymerase (10 Units) were combined into a final volume of 50 µl. For monitoring the
progress of the reaction, 1 µl of ³²P-labeled dGTP (Amersham-Pharmacia) was added to
the reaction before incubation at 37°C for 30 min. DNA/RNA hybrids were further
purified under standard conditions for Proteinase K digestion, chloroform/phenol
extraction, and ethanol precipitation as described by Sambrook J and Russell DW, *ibid*.

20

Example 13 – Selection of DNA/RNA hybrids having in part unpaired regions

DNA/RNA hybrids comprising loop-structures derived from unpaired regions were
further selectively purified by binding to randomized N25mer oligonucleotides obtained
25 as biotinylated oligonucleotides from Invitrogen. For the pretreatment of the magnetic
beads, 500 µl of magnetic beads (CPG Inc.) and 5 µl of 20 µg/µl tRNA were incubated
on ice for 30 min under occasional mixing. After three-times washing with 1X CTAB
buffer, the magnetic beads were re-suspended in 500 µl of 1X CTAB buffer and 5 µl of
20 µg/µl tRNA. The N25mer oligonucleotides 5 µl (5 µg) were pre-heated to 94°C for
30 30 sec, before adding 5 µg DNA/RNA hybrids in a buffer containing 0.1x TE. After
incubation at 37°C for 3 min, followed by 3 min at room temperature, the same volume
of 2X CTAB buffer was added, and the reaction mixture was further incubated at 45°C

for 20 min. After the hybridization of the N25 randomized oligonucleotides to the DNA/RNA hybrids, the reaction mixture was applied to the magnetic beads, incubated at room temperature for 30 min under ongoing agitation, and washed up to five times with 500 μ l of 3 M TMA Buffer (3 M TMA, 20 mM EDTA, 50 mM Tris-HCl pH7.5).

- 5 Captured DNA/RNA hybrids were released by adding 50 μ l of 0.25X solution D containing 4 M guanidium thiocyanate, 0.5% n-lauryl sarcosine, 25 mM sodium citrate pH 7.0, 100 mM beta-mercaptoethanol with 0.5% biotin and incubation at 37°C for 10 min. Recovery rates were estimated based on measurements of the radioactive label. The recovery of bound hybrid molecules was repeated until about 80% of the bound
- 10 materials could be recovered. Released DNA/RNA hybrids were pooled, concentrated by isopropanol precipitation, and remaining short oligonucleotides were removed by gel filtration on a G50 column (Takara) according to the manufacture's instructions. The capture/release step can be repeated at least once more time to improve the enrichment of nucleic acids having in part unpaired regions.

15

Example 14 – RNA removal

- Enriched fractions containing DNA/RNA hybrids were dissolved in 100 μ l of H₂O plus the same volume of a solution containing 150 mM NaOH/15 mM EDTA. After
- 20 incubation at 45°C for 10 min, 100 μ l of 1M Tris-HCl pH7.0, 2 μ l RNase I (20 Units, Promega), and 2 μ l RNase H (120 Units/ μ l, Takara) were added, and the reaction mixture was incubated at 37°C for 15 min. Double-stranded DNA was purified under standard conditions for Proteinase K digestion, chloroform/phenol extraction, and ethanol precipitation as described by Sambrook J and Russell DW, *ibid*. The resulting
- 25 pellet was dissolved in 100 μ l of H₂O, and further purified by chromatography on a S400 column (Amersham-Pharmacia) according to the manufacture's instructions, which are hereby incorporated herein by reference. In a final step, the DNA samples were concentrated by ethanol precipitation and dissolved in 2 μ l of water.

30 Example 15 - Transformation of E. coli

Using 1µl of the aforementioned DNA solution, ElectroMAXTM DH10BTM Cells (Invitrogen) were transformed by electroporation using a Cell-Porator (Biometrer) according to the transformation procedures described in the manufacturer's manual. Transformed bacteria were selected on LB medium containing 125 µg/ml Ampicillin, and positive clones thereof were isolated and further characterized.

Example 16 – Determination of end-sequences

After the titer check, bacterial clones were collected by commercially available picking machines (Q-bot and Q-pix; Genetics) and transferred to 384-microwell plates. Transformed *E. coli* clones holding vector DNA were divided from 384-microwell plates and grown in four 96-deepwell plates. After overnight growth, plasmids were extracted either manually (Itoh M et al., Nucleic Acids Res. 25 (1997) 1315-1316, hereby incorporated herein by reference) or automatically (Itoh M et al., Genome Res. 9 (1999) 463-470, hereby incorporated herein by reference). Sequences were typically run on a RISA sequencing unit (Shimadzu) or a Perkin Elmer-Applied Biosystems ABI 377 in accordance with standard sequencing methodologies such as described by Shibata K et al., Genome Res. 10 (2000) 1757-1571, hereby incorporated herein by reference. Sequencing was alternatively performed using primers nested in the flanking regions of the cloning vector and a BigDye Terminator Cycle Sequencing Ready Reaction Kit v1.1 (Applied Biosystems, Cat. No. 4337449) and an ABI3700 (Applied Biosystems) sequencer according to the manufacture's product descriptions, which are hereby incorporated herein by reference.

Standard primers as used for vectors of the pFLC family included:

M13 Reverse primer: 5'-CAGGAAACAGCTATGAC (SEQ ID NO: 15)

M13 (-20) Forward primer: 5'-GTAAAACGACGGCCAG (SEQ ID NO: 16)

Example 17 – Sequence analysis by BLAST search in public databases

Clones isolated as above were sequenced from their 5' ends using aforementioned primers. 5'-end-related sequence information can be analyzed for its identity by

- standard software solutions to perform sequence alignments like NCBI BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>), FASTA, available in the Genetics Computer Group (GCG) package from Accelrys Inc. (<http://www.accelrys.com/>) or alike. Such software solutions allow for an alignment of 5'-end-related sequence to any sequence information within a database, including, but not limited to, the DNA Data Bank of Japan or DDBJ (<http://www.ddbj.nig.ac.jp/>), the National Center for Biotechnology Information or NCBI (<http://www.ncbi.nlm.nih.gov/>), or the European Bioinformatics Institute or EMBL-EBI (<http://www.ebi.ac.uk/index.html>).
- 10 An example of a BLAST search in GenBank using a 5'-end sequence is given, where the following sequence was obtained under the aforementioned conditions:

Clone ID: G2-5 (SEQ ID NO:17)

5' ACCAGGGGGGATCCTGGGATTCTGCCTTTCCGGAAGGTCGGGCTGGCTCT
15 CGGTGCTAGGCCGCCGTTCTAGGCCCAAGTCGGGATTCCAGGCCCTGCTCCA
AGCCAAGTCCGTAGCCGCGACGACACAGCGGGGATCGCAGAGGACGCGGC
CGCAGCTCGTTGGGGTGGCTGAGCCTTGAGAGGGCGGCAGCCGGGGTTCGCA
GGCCGTGACGGCCCCTCCCCGCCCTGAGCGCAGCGGCGAAGGACGACAGC
CCGGGACCGGAGGAGGTGAAGCGGTCACGTAAGTCCCCGGGTGATGACTC
20 ACCCGCTCTCCTGGCAGATTTGAGGTGCGTGCTGCTGTTGCTGCAGGGCGGA
CGTCGGTGTGCCAGAGCGGTGCTGCGGCGACCGCTTGCTTGCCCTTGGTTC
GGCGGCCCCGGGGTTTCCCGAGAGCGTTCTCTTTGTAACGTCTGCGGCCTGC
AGCTCGCAGCGAGCTCCGGCTTCTCCCATGAAGGTGTCTCTGGGTAACGGC
GACATGGGCGTCTCTGCCCACCTGCAGCCTTGCAAGTCTGGAACCTACGCGGT
25 TTTTACCAGCAACACGCACAGTTCGGTTGTATTGCAAGGCTTTGATCAGCT
CAGAATAGAGGGCCTACTTTGTGACGTGACGCTGGTACCTGGCGATGGAGA
GGAAATCTTNCCGGNTCATAGAGCCATGATGGCGTCTGCCAGTGATTATTTT
AAGGGCATGTTCACTGGAGGAATGAAAGAAAAAGATTTGATGTGCATCAAG
CTTCATGGGGTGAACAAAGTTGGTCTGAA3'

The aforementioned sequence data were used to search for related sequences within GenBank using BLAST under default settings. The graphical representation of the search result is shown in Figure 11.

5 Sequences giving alignments of high significant are listed below:

- [gi|26083294|dbj|AK033231.1](#) Mus musculus 15 days embryo mal... [1370](#) 0.0 [L][U]
[gi|27370321|ref|NM_172871.1](#) Mus musculus RIKEN cDNA C53005... [1356](#) 0.0 [L][U]
- 10 [gi|26340527|dbj|AK049803.1](#) Mus musculus 12 days embryo spi... [1356](#) 0.0 [L][U]
[gi|26333040|dbj|AK039099.1](#) Mus musculus adult male hypotha... [1170](#) 0.0 [L][U]
[gi|34869493|ref|XM_233157.2](#) Rattus norvegicus similar to H... [484](#) e-133 [L][U]
[gi|34783554|gb|BC039133.2](#) Homo sapiens KIAA1354 protein, m... [313](#) 3e-82 [L][U]
- 15 [gi|21750152|dbj|AK091715.1](#) Homo sapiens cDNA FLJ34396 fis,... [313](#) 3e-82 [L][U]
[gi|16214560|emb|AL162420.13](#) Human DNA sequence from clone ... [313](#) 3e-82
[gi|19584360|emb|AL713669.1|HSM803007](#) Homo sapiens mRNA; cDN... [313](#) 3e-82 [L][U]
- 20 [gi|7243088|dbj|AB037775.1](#) Homo sapiens mRNA for KIAA1354 p... [313](#) 3e-82 [L][U][G]
[gi|24308180|ref|NM_018847.1](#) Homo sapiens KIAA1354 protein ... [313](#) 3e-82 [L][U][G]
[gi|31874543|emb|BX538121.1|HSM806316](#) Homo sapiens mRNA; cDN... [145](#) 1e-31
- 25 [L][U]
[gi|34527916|dbj|AK122724.1](#) Homo sapiens cDNA FLJ16227 fis,... [98](#) 3e-17 [L][U]
[gi|14348719|emb|AL591986.1|HS169K131](#) Novel human gene mappi... [98](#) 3e-17 [L][U]
- 30 [gi|34531431|dbj|AK125356.1](#) Homo sapiens cDNA FLJ43366 fis,... [98](#) 3e-17 [L][U]
[gi|40353041|gb|BC064576.1](#) Homo sapiens kelch-like 13 (Dros... [98](#) 3e-17 [L][U]

- gi|7242972|dbj|AB037730.1| Homo sapiens mRNA for KIAA1309 p... 98 3e-17
[L][G]
gi|28972713|dbj|AK122491.1| Mus musculus mRNA for mKIAA1309... 68 3e-08
[L][U]
5 gi|13385673|ref|NM_026167.1| Mus musculus kelch-like 13 (Dr... 68 3e-08
[L][U][G]
gi|12836021|dbj|AK004677.1| Mus musculus adult male lung cD... 68 3e-08
[L][U][G]
gi|4753290|gb|AC006963.4|AC006963 Homo sapiens PAC clone RP... 62 2e-06 [L]

10

Example 18 – Biotinylation of RNA

About 10 µg of the RNA were labeled with 10 µl of Label IT reagent (Panvera) and 10 µl of labeling buffer A in a final volume of 100 µl according to the manufacture's instructions, which are hereby incorporated herein by reference. After incubation at 37°C for 1 hour, adding 1/20-volume of 5M NaCl and two volumes of 99% ethanol precipitated biotinylated RNA, which was washed twice with 80% ethanol, and finally re-suspended in 20 µl of H₂O. RNA was further purified under standard conditions for Proteinase K digestion, chloroform/phenol extraction, and ethanol precipitation as described by Sambrook J and Russell DW, *ibid*.

20

Similarly, any RNA can be labeled by digoxigenin using a DIG Chem-Link and Detection Set from Roche Diagnostics GmbH (Cat. No. 1 836 463, or Cat. No. 1 277 073). The handbook for this set is hereby incorporated herein by reference.

25

Example 19 – Subtraction of DNA/RNA hybrids

Pluralities of single-stranded DNA given as a “tester” can be subtracted by a biotinylated or digoxigenin-labeled “driver” as described by Carninci P et al., *Genome Res.* 10 (2000) 1617-1630, which is hereby incorporated herein by reference. In brief, depending on the stringency desired for the experiments, hybridizations were carried out at Rot values of between 5 to 10 (mild subtraction) or 500 (strong subtraction) in a

30

buffer containing 80% formamide (from deionized stock), 250 mM NaCl, 25 mM HEPES (pH 7.5), and 5 mM EDTA at 42°C. Hybridization times varied depending on the Rot values applied, where commonly about 500 ng of tester were used. The ratio of tester to driver can vary as well depending on the experimental needs. Hybridizations were terminated by ethanol precipitation under standard conditions, and DNA/RNA hybrids were re-suspended in 50 µl of H₂O. Hybrids between the labeled driver and the tester were separated from free tester molecules by the means of a high affinity binding substance directed against biotin or digoxigenin. In the case of biotin, streptavidin-coated magnetic beads (CPG Inc.) were applied as already described in Example 14. Similarly, digoxigenin-labeled hybrids can be separated under similar conditions using an immobilized anti-digoxigenin antibody, available from Roche Diagnostics GmbH as a Fab fragment (Cat. No. 1 093 274). Tester molecules isolated during the subtraction step were further purified by Proteinase K digestion, phenol/chloroform extraction, and ethanol precipitation to concentrate the material, followed by gel filtration on a G50 column (Takara), and further ethanol precipitation.

Example 20 – Preparation of circular single-stranded DNA

Single-stranded circular DNA to perform the invention can be prepared by an alternative approach involving a strand-specific amplification of one strand and its religation into circular single-stranded DNA followed by digestion of remaining double-stranded DNA by a double-stranded DNA specific endonuclease. As plasmid DNA prepared by a standard method is commonly obtained as supercoiled DNA, it is advisable to relax the plasmid DNA by treatment with Topoisomerase II (Amersham Biosciences, Code Number E78303Y) before the amplification step. In a standard reaction 600 ng of plasmid DNA are incubated with 2 units of Topoisomerase II at 30°C over 15 min according to the maker's instructions, which are hereby incorporated herein by reference. The reactions were terminated by Proteinase K digestion, followed by phenol/chloroform extraction and ethanol precipitation under standard conditions.

30

For the amplification/ligation reaction, up to 500 ng of the aforementioned plasmid DNA from library G2 or G3 were placed into a 0.5 ml tube as template for the

amplification reaction. As primers for the use of vector pFLCII the following oligonucleotides were prepared:

Primer RNAI

5'

5 AGGAGAGGTCTAGACCACTTTGTACAAGAAAGCTGGGTGGATCCGGACTGT
TTTT (SEQ ID NO: 18)

Primer RNAII

5' - CTCGACCTCGAGAGCCTGCTTTTTTTGTACAAACTTGTGGCCCGGTACCC
(SEQ ID NO: 19)

10 Ori-R

5'- GTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATT (SEQ ID NO:
20)

Amp-R 5'- GCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGG (SEQ
ID NO: 21)

15 Although the reaction can be performed with one primer only, it is preferable to use up
to four primers at the same time to assure complete synthesis of the second strand. For
each primer 10 μ M of the oligonucleotide is added to the aforementioned template, and
all components are heated to 95°C for 5 min to separate the strands and allow for the
annealing of the primers; after the denaturation, samples were placed immediately on
20 ice. To initiate the reaction the following solutions were added: 10X reaction buffer 5
 μ l, 10 mM dNTP mix 1.5 μ l (0.3 mM final concentration), T7 DNA Polymerase (5
units, USB), and deionized H₂O to a final volume of 50 μ l. After an incubation of the
reaction mixture at 37°C for 20 minutes, heating at 70°C for 10 min terminated the
reaction. In this setup, the reaction allows only for one reaction cycle. Alternatively the
25 reaction can be performed as a cycle by adding to the aforementioned template and
primers 10X reaction buffer 5 μ l, 10 mM dNTP mix 1.5 μ l (0.3 mM final concentration),
and deionized H₂O up to a final volume of 50 μ l. After a hot start at 95°C for 1 min, 5
units of Deep Vent Polymerase (New England Biolabs), and 5 units of Tsc DNA ligase
(Roche Diagnostics GmbH) were added to initiate the reaction. The reaction was run
30 with some 10-20 cycles of 95°C for 1 min, 55°C for 30 sec, and 72°C for 8 min. After
completion of the reaction, remaining primers were removed by chromatography on S-
400 columns (Amersham Biosciences) according to the maker's instruction, and

remaining enzymatic activities were digested with Proteinase K (Qiagen) followed by extraction with phenol/chloroform. DNA was precipitated with ethanol under standard conditions, and the sample was re-suspended in 20 µl H₂O.

- 5 Independent from the approach chosen, remaining double-stranded DNA within the reaction mixture was destroyed by DSN treatment. The aforementioned reaction mixture was first incubated at 65°C for 5 min to dissociate secondary structures, which may have formed within single-stranded DNA molecules. After heat treatment, the sample was immediately placed on ice. For digestion with DSN, 20 µl of the sample
10 were mixed with 2.5 µl of 10X DSN buffer (Evrogen, Cat.# EA001) and 1 µl of the single-stranded-DNA binding protein T4gene-32 (4.5 µg/µl, USB). After the solution was adjusted to a final volume of 24 µl with water, 1 µl of a DSN enzyme stock (1 unit per µl, Evrogen, Cat.# EA001) was added and the reaction mixture was incubated at 37°C for 1 hour. Addition of 1 µl of a 0.5M EDTA stock solution terminated the
15 reaction, before the volume was adjusted with water to 50 µl, and 1 µl of 10% of SDS was added. Remaining DSN activity was destroyed by Proteinase K treatment, for which 2 µl of a Proteinase K enzyme solution (20 µg/µl, Qiagen, Cat. NO. 19131) were added, and the reaction mixture was incubated at 45°C for 2 hours. After the Proteinase K treatment the reaction mixture was extracted with equal volumes of
20 phenol/chloroform and chloroform under standard conditions. Single-stranded DNA was precipitated out of the aqueous phase by adding 1 µl of a 2 µg/µl glycogen solution, 2.5 µl of 5 M NaCl, and 150 µl of absolute ethanol. After incubation at minus 20°C for 30 min, DNA was collected by centrifugation. The DNA pellet was washed twice with 80% ethanol before the DNA was finally dissolved in 40 µl of water. The quality of the
25 DNA can be tested in a 0.8% agarose gel under standard conditions.

Claims

1. A method for isolating nucleic acid molecules whose sequences include a portion complimentary to each other and a mutually unrelated portion, comprising the steps of:
 - 5 hybridizing a single-stranded linear nucleic acid molecule with a single-stranded circular nucleic acid molecule so as to form a hybrid molecule;
 filling in a single-stranded region of the hybrid circular nucleic acid which region is not hybridized with the single-stranded linear nucleic acid molecule with nucleic acids so as to form a double-stranded hybrid which is double stranded except at
10 any single-stranded loop portion; and
 recovering a hybrid molecule that has a single-stranded loop portion resulting from an unpaired region in either of the hybridized linear nucleic molecule acid or the hybridized circular nucleic acid molecule or both.
- 15 2. The method of claim 1, further comprising the step of attaching at least one short oligonucleotide that functions as a primer to the single-stranded circular nucleic acid molecule in preparation for the filling-in step.
- 20 3. The method of claim 1 or 2, wherein the filling-in step is carried out using a polymerase as well as the hybridized linear nuclear acid molecule or the short oligonucleotide as a primer.
- 25 4. The method of any of claims 1-3, further comprising the step of removing a portion of the single-stranded linear nucleic acid molecule which portion has not hybridized to the single-stranded circular nucleic acid molecule by means of an exonuclease.
5. The method of claim 4, wherein the exonuclease is chosen from the group consisting of exonuclease VII, exonuclease T, exonuclease I, and any mixture thereof.
- 30 6. The method of any of claims 1 to 5, wherein the step of recovering the hybrid molecule, which is double-stranded except at the single-stranded loop portion, comprises the steps of attaching a substance that preferentially binds to a single-

stranded nucleic acid and has a tag molecule to the single-stranded loop portion of the hybrid molecule, and retrieving a hybrid molecule having the single-stranded loop portion using the tag molecule associated with the single-stranded binding substance by binding the tag molecule to a matrix that has binding affinity to the tag molecule.

5

7. The method of claim 6, wherein the substance that preferentially binds to a single-stranded nucleic acid is a protein or an oligonucleotide, and the tag molecule is chosen from the group of biotin, digoxigenin, an antibody, and an antigen.

10 8. The method of claim 7, wherein the matrix has one chosen from the group consisting of avidin, streptavidin, digoxigenin-binding molecule, an antigen, or an antibody.

9. The method of any of claims 1-8, wherein the single-stranded linear nucleic acid molecule is RNA and the single-stranded circular nucleic acid is DNA.

15

10. The method of claim 9, wherein the RNA is naturally occurring RNA.

11. The method of claim 9, wherein the RNA is prepared *in vitro*.

20 12. The method of claim 9, wherein the RNA is prepared from a nucleic acid cloned into a vector.

13. The method of claim 12, further comprising the step of removing a portion of the RNA derived from vector sequences by means of an RNase.

25

14. The method of claim 13, wherein the RNase cleaves the portion of the RNA that is hybridized to a short DNA oligonucleotide.

15. The method of claims 13 or 14, wherein the RNase is chosen from the group
30 consisting of RNase H, Hybridase, and any mixture thereof.

16. The method of any of claims 9-15, further comprising the step of removing RNA from the hybrid molecule by means of an RNase after the recovering step using a single-stranded nucleic acid binding substance.
- 5 17. The method of claim 16, wherein the RNase is chosen from the group consisting of RNase H, RNase I, and any mixture thereof.
18. The method of any of claims 1-19, further comprising the step of cloning the hybrid molecule after the recovering step.
- 10 19. The method of any of claims 1-18, wherein the single-stranded circular nucleic acid molecule comprises genetic elements required for cloning.
20. The method of any of claims 1-19, wherein the single-stranded circular nucleic acid molecule comprises an ori for initiation of replication.
- 15 21. The method of any of claims 1-20, wherein the single-stranded circular nucleic acid molecule comprises one or more selection markers used for cloning.
- 20 22. The method of claim 21, wherein the selection marker for cloning is at least one chosen from the group of Ampicillin, Tetracycline, Chloramphenicol, Kanamycin, Apramycin, Zeocin, and a mixture thereof.
23. The method of any of claims 1-22, wherein the nucleic acid molecules whose sequences include a portion complimentary to each other and a mutually unrelated portion are derived for alternatively spliced transcripts.
- 25 24. The method of any of claims 1-23, wherein the hybrid molecule is analyzed by partial or full-length sequencing.
- 30 25. The method of any of claims 1-24, further comprising the steps of obtaining a single-stranded nucleic acid molecule from the hybrid molecule and analyzing the

single-stranded nucleic acid molecule by hybridization to an immobilized nucleic acid molecule.

26. The method of claims 25, wherein the immobilized nucleic acid molecule is placed
5 on an array.

27. The method of claim 26, wherein the array comprises oligonucleotides or printed DNA fragments.

10 28. The method of any of claims 25-27, wherein the immobilized nucleic acid molecule comprises exon or intron specific information.

29. The method of any of claims 27-30, wherein the immobilized nucleic acid molecule is tiled nucleic acid molecules derived from genomic or cDNA sequence information.
15

30. The method of claim 25, wherein the single-stranded nucleic acid molecule is used as a hybridization probe in hybridization experiments.

31. The method of claim 30, wherein the hybridization probe is used for identification
20 of related nucleic acid molecules.

32. The method of claim 30 or 31, wherein the hybridization probe is used for enriching related nucleic acid molecules.

25 33. A method for preparing circular single-stranded nucleic acid molecules, comprising the steps of:

separating a first double-stranded circular nucleic acid molecule into two single-stranded circular nucleic acid molecules;

hybridizing at least one primer to one of the two single-stranded circular nucleic
30 acid molecules;

synthesizing a second strand for the one of the two single-stranded circular nucleic acid molecules using the primer so as to form a second double-stranded circular nucleic acid molecule; and

5 separating a single-stranded circular nucleic acid molecule from the second double-stranded circular nucleic acid molecule.

34. The method of claim 33, wherein the two strands of the double-stranded circular nucleic acid molecule are separated by heat or alkali treatment.

10 35. The method of claim 33 or 34, wherein the synthesis of the second strand is carried out using a polymerase.

36. The method of claim 35, wherein the polymerase is a DNA polymerase having no strand displacement activity.

15

37. The method of claim 35, wherein the DNA polymerase is chosen from the group consisting of T7 DNA polymerase, Vent polymerase, Deep Vent polymerase, and any mixture thereof.

20 38. The method of any of claims 33-37, wherein the second strand prepared by means of the polymerase is ligated by means of a ligase to the end of the primer to form a circular nucleic acid molecule.

39. The method of claim 38, wherein the ligase is chosen from the group consisting of
25 Tsc DNA Ligase, Ampligase, Pfu DNA ligase, Taq DNA Ligase, and any mixture thereof.

40. The method of any of claims 38, wherein the polymerase and the ligase are thermostable.

30

41. The method of claim 33, wherein the following steps are repeated to perform a chain reaction prior to the step of separating the first double-stranded circular nucleic acid molecule:

- 5 separating a double-stranded circular nucleic acid molecule into two single-stranded circular nucleic acid molecules by heat treatment;
- annealing at least one primer to one of the two single-stranded circular nucleic acid molecules;
- synthesizing a second strand by means of a polymerase as initiated by the at least one primer; and
- 10 ligating the ends of the second strand to form a double-stranded circular nucleic acid molecule.

42. A method for isolating nucleic acid isoforms using the the hybrid molecule that has a single-stranded loop portion obtained by the method of any of claims 1-41 from
15 fragmented genomic DNA, cDNA, full-length cDNA, mRNA and/or RNA.

43. The method of claim 42, wherein the isoforms are full-length cDNAs or a fragment thereof comprising an unpaired region.

20 44. A cloning vector comprising an isoform obtained according to claim 42 or 43.

45. A host cell comprising the vector of claim 44.

25 46. A method for preparing a polypeptide derived from an isoform obtained according to claim 42 or 43 *in vitro* or *in vivo*.

47. A method for identifying isoform polypeptides using information obtained from the hybrid molecule obtained according to any of claims 1-43.

30 48. The method of any of claims 1-41, wherein nucleic acid isoforms are further obtained from the hybrid molecule that is in turn obtained from a preparation of single-stranded circular nucleic acids.

49. The method of any of claims 1-41, further comprising the steps of obtaining a single-stranded nucleic acid molecule from the hybrid molecule as a nucleic acid molecule probe comprising at least one exon or intron.

5

50. The method of any of claims 1-41, further comprising the steps of obtaining a single-stranded nucleic acid molecule from the hybrid molecule as a nucleic acid molecule probe comprising cDNA..

10 51. A method for determining sequence variation of an isoform isolated according to claim 42 or 43, comprising the step of sequencing a full-length or partial sequence of the isoform.

15 52. The method of any of claims 1-41, wherein sequence information of sequence variants resulting from the hybrid molecule that has a single-stranded loop is used for designing sequencing primers.

20 53. The method of claim 42 or 43, wherein isoform sequencing data obtained from the isoforms are aligned to the genome, to genomic sequencing data and/or to cDNA sequencing data so as to obtain genetic information.

54. A method for identifying nucleic acids that, comprising the steps of:

forming a circular nucleic acid molecule from a first set of single-stranded linear nucleic acid molecules,

25 hybridizing a second set of single-stranded linear nucleic acid molecules with the circular nucleic acid molecule so as to form a circular hybrid; and

recovering a circular hybrid that has a single-stranded loop structure resulting from an unpaired region in either of the hybridized linear nucleic acid molecule or the hybridized circular nucleic acid molecule or both.

30

55. The method of claim 54, further comprising the step of attaching to the single-stranded circular nucleic acid molecule at least one short oligonucleotide that functions as a primer prior to or after the hybridizing step.

5 56. The method of claim 54 or 55, further comprising the step of filling in a single-stranded region of the circular hybrid to form using a polymerase a double-stranded circular hybrid which is double-stranded except at the loop portion.

10 57. The method of claim 56, the filling-in step is carried out using the hybridized nuclear acid or the short oligonucleotide as a primer.

15 58. The method of any of claims 54 to 57, wherein the step of recovering the circular hybrid comprises the steps of attaching a substance that preferentially binds to a single-stranded nucleic acid and that includes a tag molecule to the single-stranded loop portion of the circular hybrid, and screening a circular hybrid that binds to a matrix that has binding affinity to the tag molecule.

20 59. The method of claim 58, wherein the substance that preferentially binds to a single-stranded nucleic acid is a protein or an oligonucleotide, and the tag molecule is chosen from the group consisting of biotin, digoxigenin, an antibody, and an antigen.

25 60. The method of claim 58, wherein the matrix has one chosen from the group consisting of avidin, streptavidin, digoxigenin-binding molecule, an antigen, and an antibody.

61. The method of any of claims 54-60, wherein the linear nucleic acid molecule is RNA and the circular nucleic acid molecule is DNA.

30 62. A method for cloning a nucleic acid that comprises a sequence corresponding to a portion of a gene that is differentially spliced, comprising the steps of:

hybridizing a single-stranded linear nucleic acid molecule with a circular nucleic acid molecule to form a circular hybrid;

recovering a circular hybrid that has a single-stranded loop structure resulting from an unpaired region in either of the hybridized linear nucleic acid molecule or the hybridized circular nucleic acid molecule or both;

5 removing one strand from the circular hybrid to form a single-stranded circular nucleic acid; and

cloning the single-stranded circular nucleic acid in a biological host.

63. The method of claim 62, further comprising the step of attaching to the circular nucleic acid molecule at least one short oligonucleotide that functions as a primer prior
10 to or after the hybridizing step.

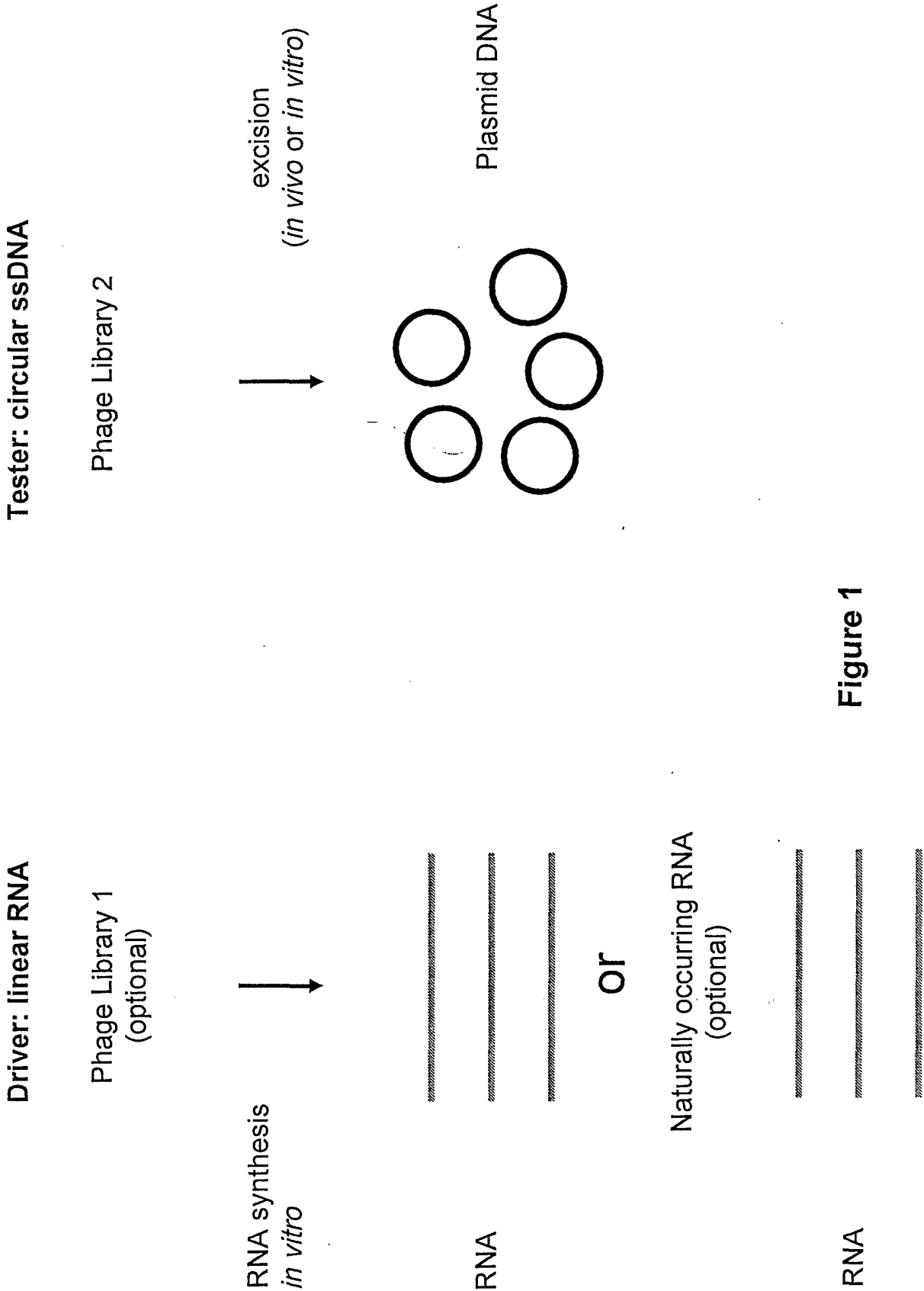
64. The method of claim 62 or 63, further comprising the step of filling in a single-stranded region of the circular hybrid to form using a polymerase a double-stranded circular hybrid which is double-stranded except at the single-stranded loop portion.
15

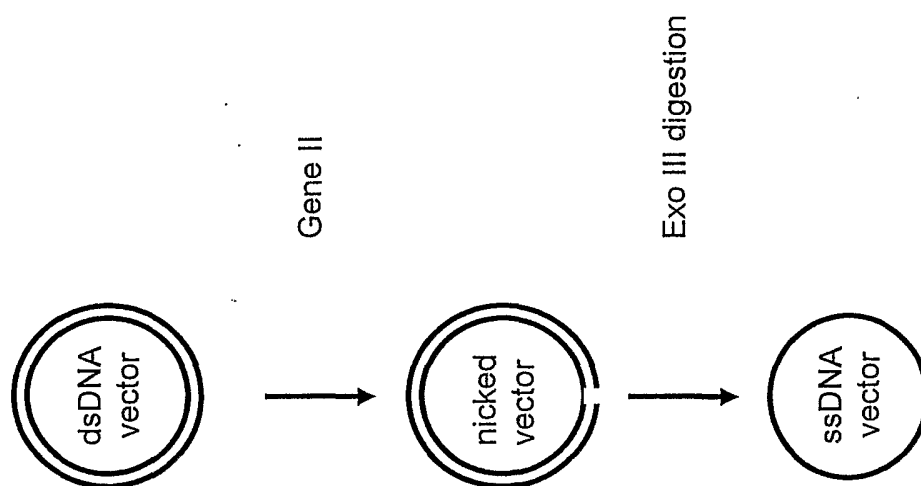
65. The method of claim 64, the filling-in step is carried out using the linear nucleic acid molecule hybridized with the circular nucleic acid molecule or the short oligonucleotide as a primer.

20 66. The method of any of claim 62 to 65, wherein the step of recovering the circular hybrid comprises the steps of attaching a substance that preferentially binds to a single-stranded nucleic acid molecule and that includes a tag molecule to any single-stranded loop portion of the circular hybrid that has formed a double-stranded hybrid, and screening a circular hybrid that binds to a matrix that has binding affinity to the tag
25 molecule.

67. The method of claim 66, wherein the substance that preferentially binds to a single-stranded nucleic acid is a protein or an oligonucleotide, and the tag molecule is chosen from the group consisting of biotin, digoxigenin, an antibody, and an antigen.
30

68. The method of claim 67, wherein the matrix comprises one chosen from the group consisting of avidin, streptavidin, digoxigenin-binding molecule, an antigen, or an antibody.
- 5 69. The method of any of claims 62-68, wherein the linear nucleic acid molecule is RNA and the circular nucleic acid molecule is DNA.



**Figure 2**

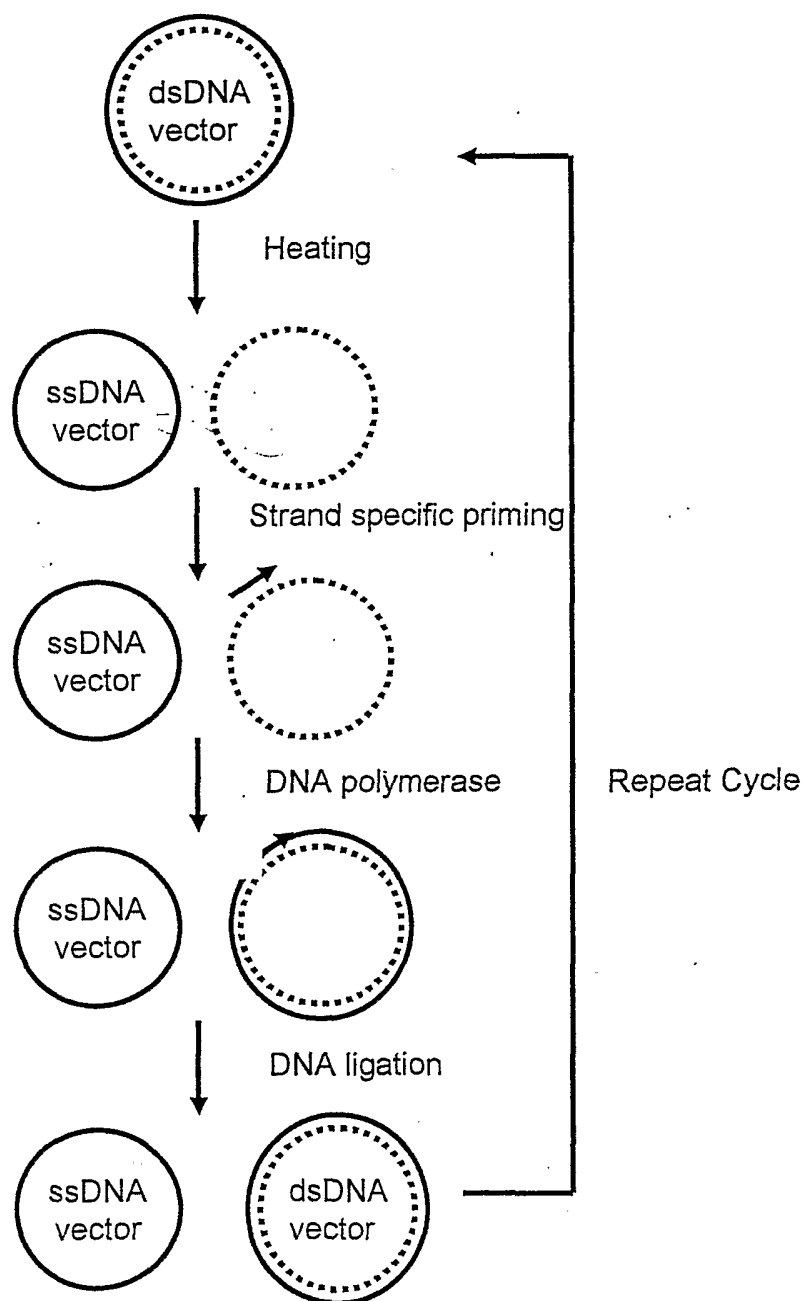


Figure 3

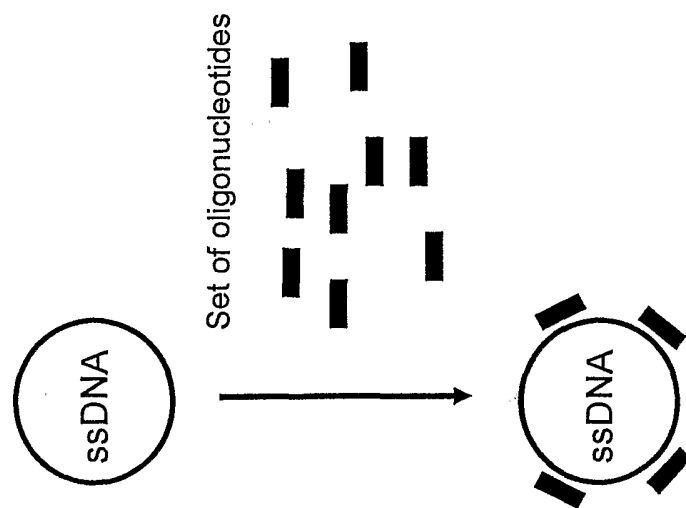


Figure 4

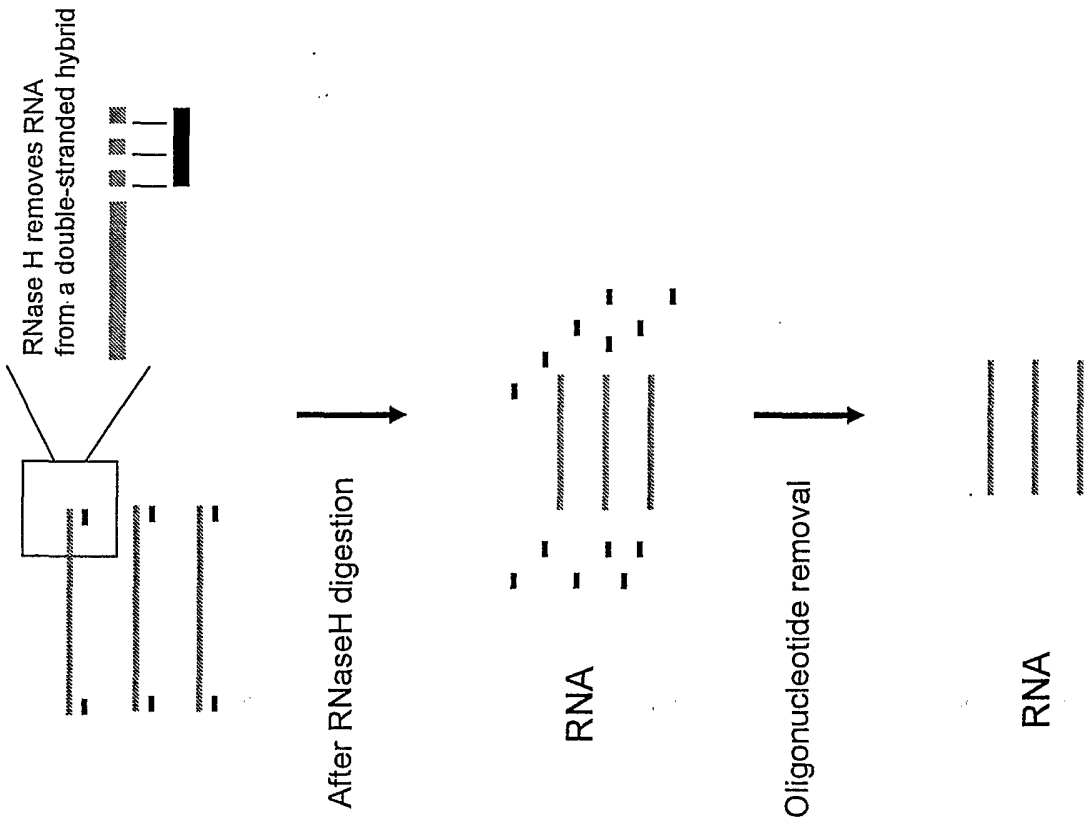


Figure 5

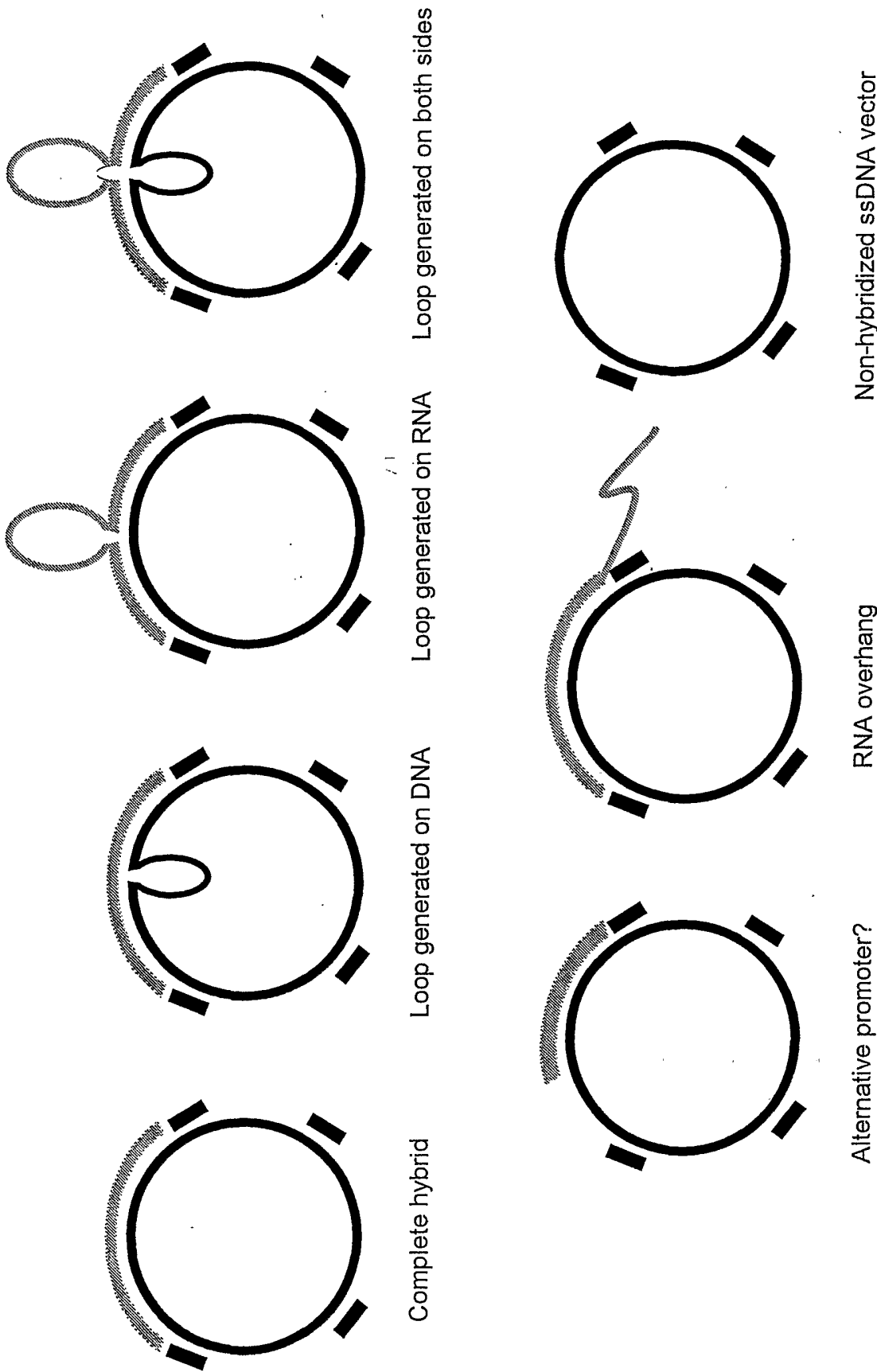


Figure 6

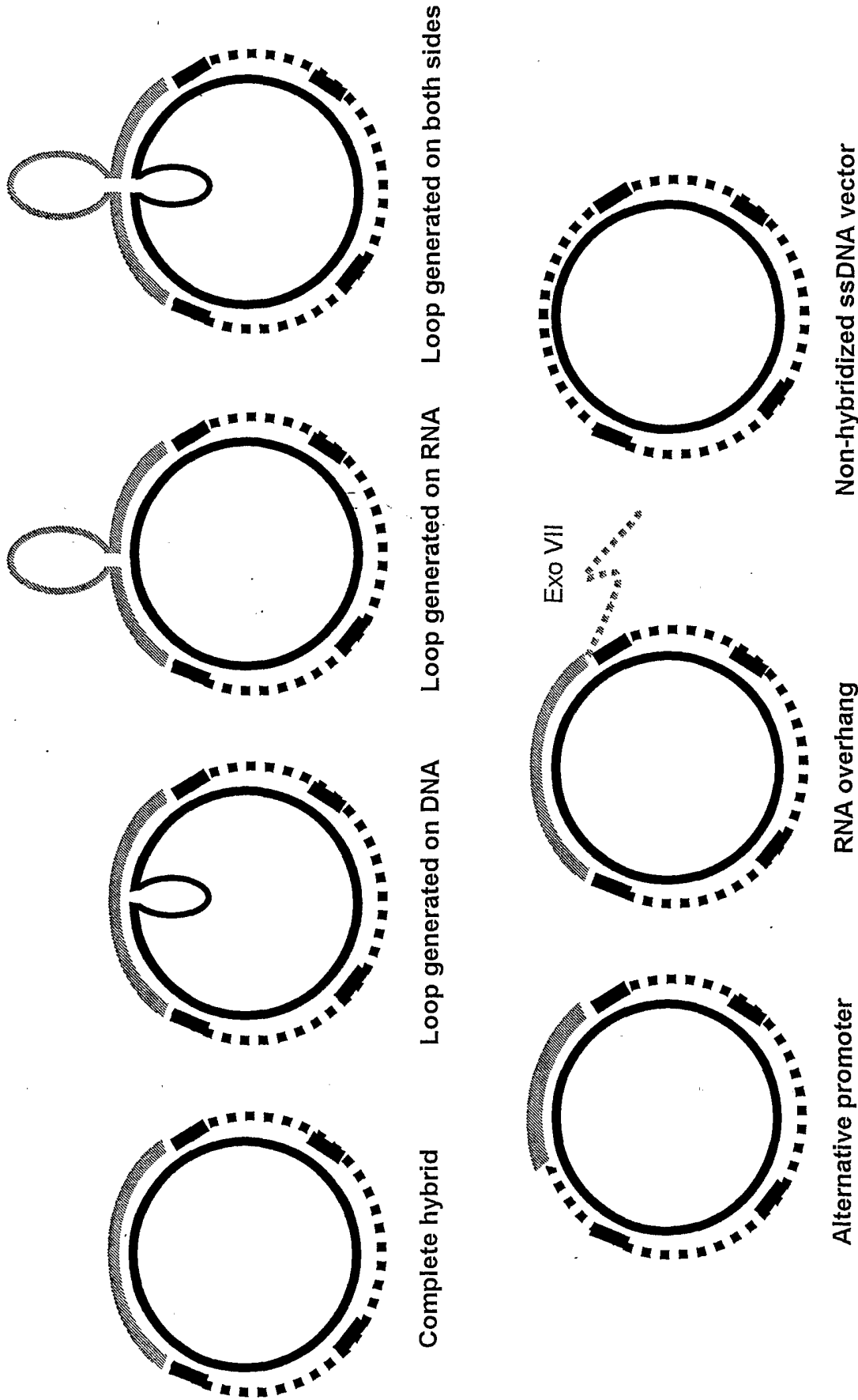


Figure 7

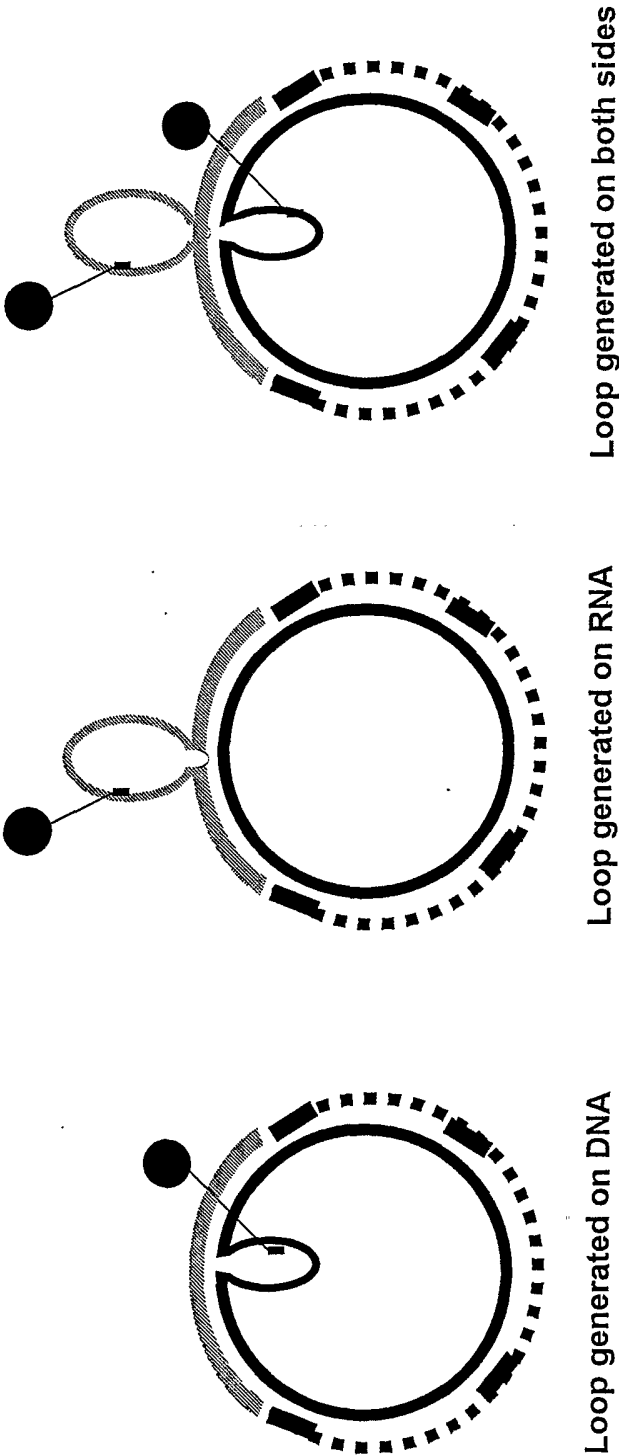
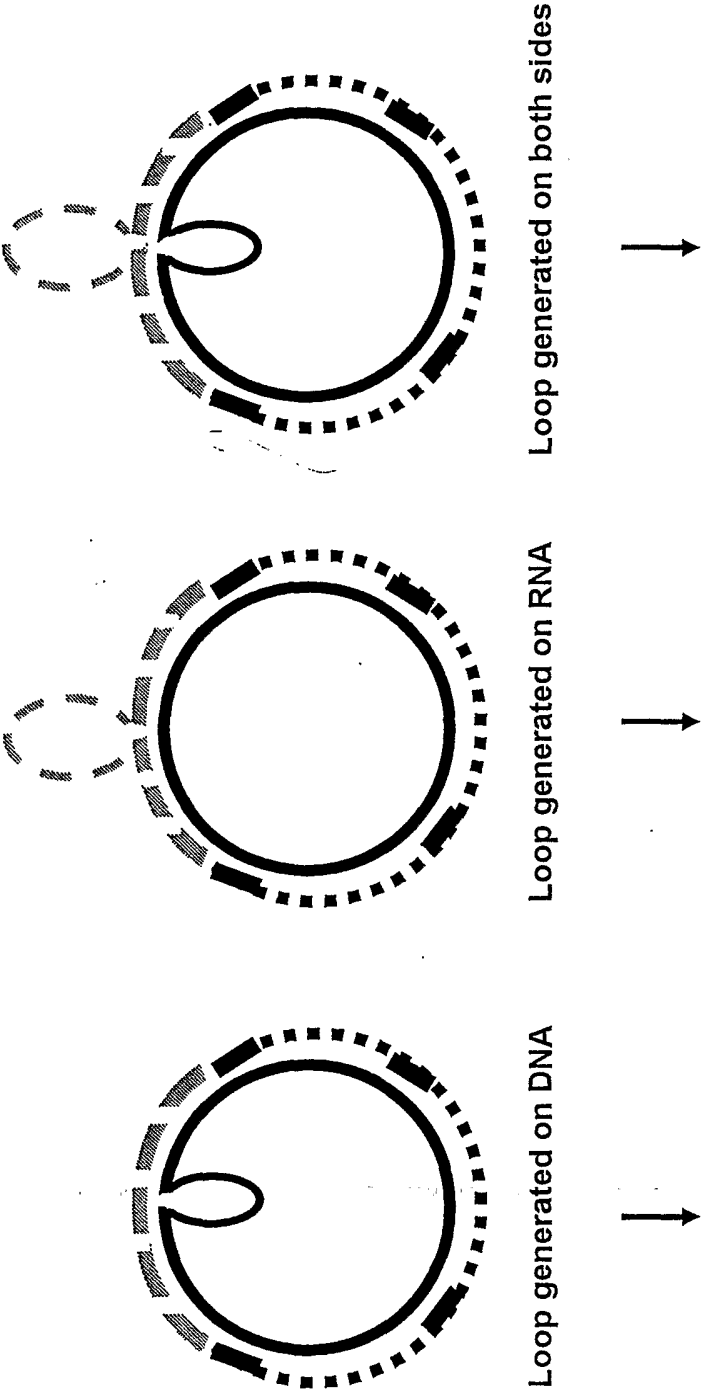


Figure 8



Transfer of single-stranded vector DNA into *E. coli*

Figure 9

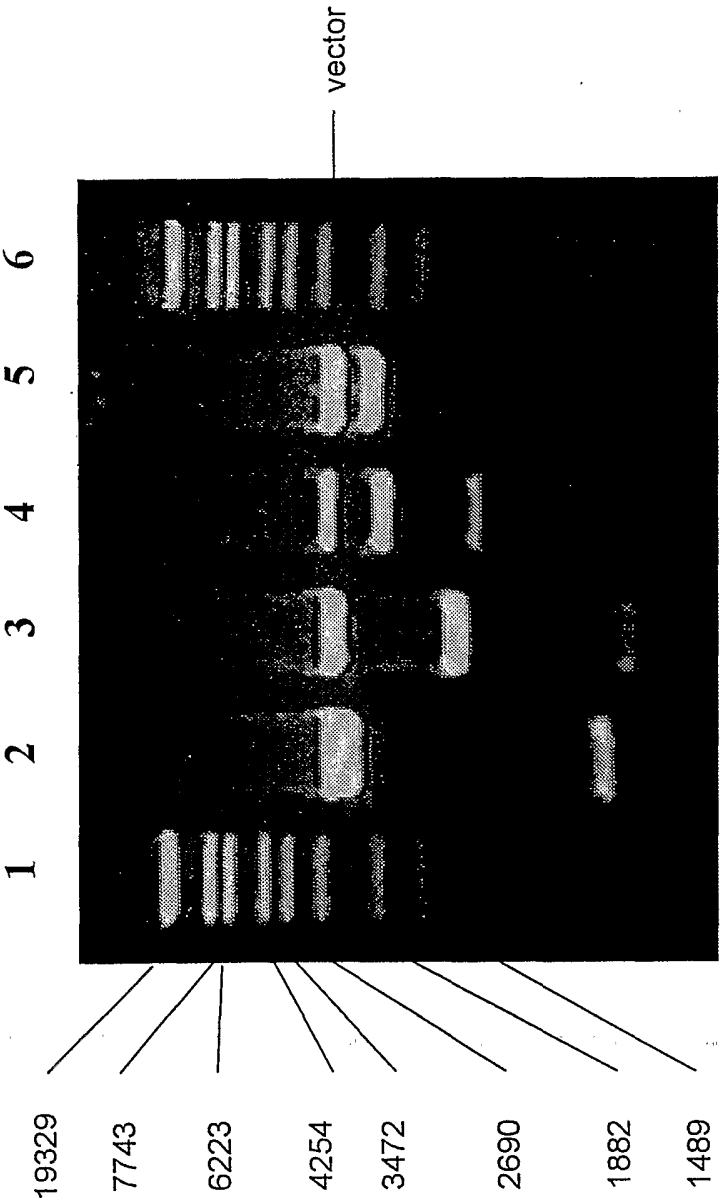


Figure 10

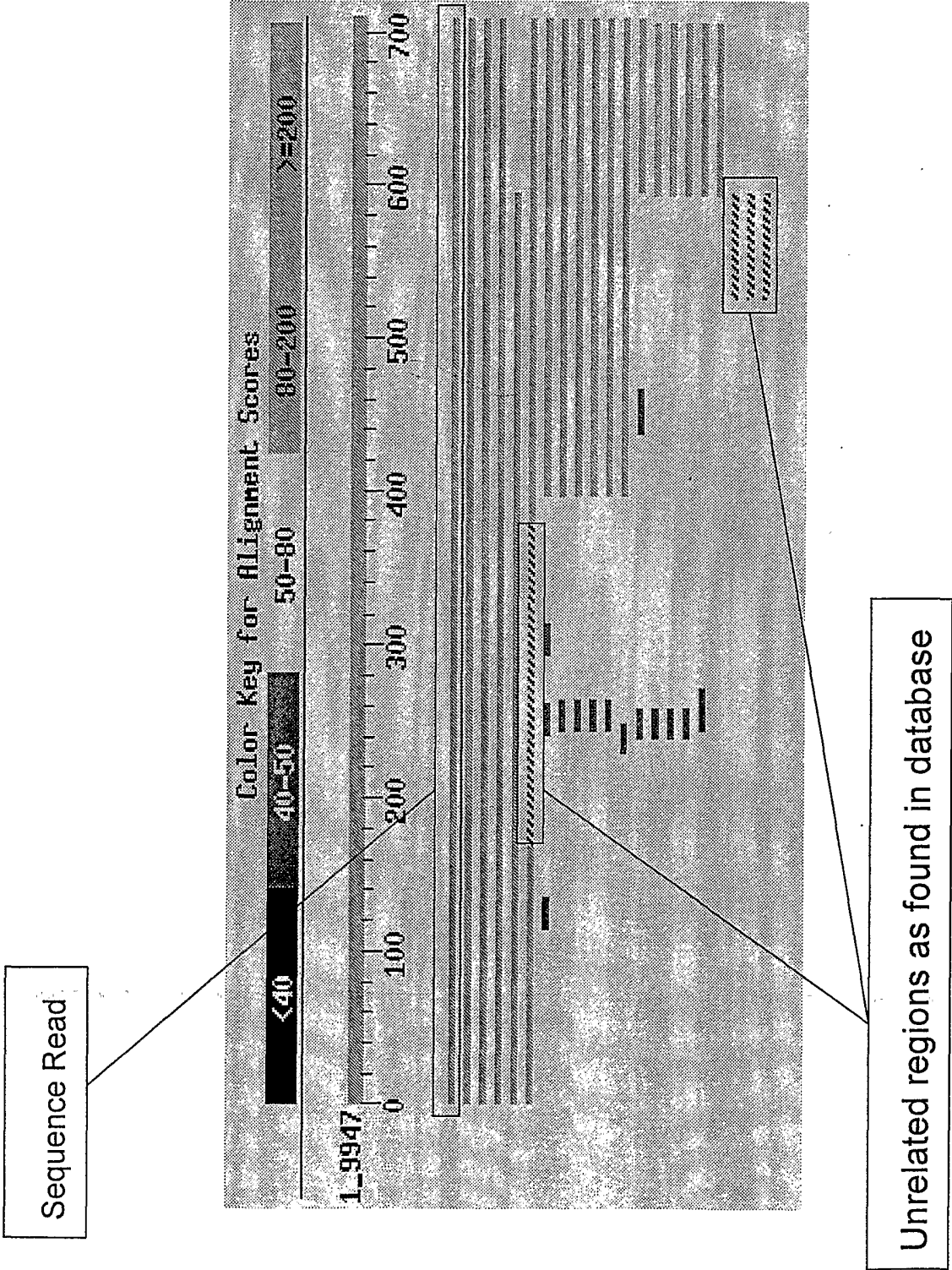


Figure 11

SEQUENCE LISTING

<110> KABUSHIKI KAISHA DNAFORM
RIKEN

<120> METHOD FOR ISOLATING NUCLEIC ACID ISOFORMS

<130> 1451

<150> JP2004-139869

<151> 2004-05-10

<160> 21

<170> PatentIn version 3.1

<210> 1

<211> 49

<212> DNA

<213> Artificial

<220>

<223> Primer T3GW1

<400> 1

gagagagaga attaacccctc actaaaggga caagtttgta caaaaaagc

49

<210> 2

<211> 47

<212> DNA

<213> Artificial

<220>

<223> Primer T7GW2

<400> 2

gagagagaga attaacctca ctaagggacc actttgtaca agaaagc

47

<210> 3

<211> 47

<212> DNA

<213> Artificial

<220>

<223> Primer T3GW2

<400> 3

gagagagaga attaacctca ctaagggacc actttgtaca agaaagc

47

<210> 4

<211> 48

<212> DNA

<213> Artificial

<220>

<223> Primer T7GW1

<400> 4

gagagagagt aatacgactc actatgggac aagtttgtac aaaaaagc

48

<210> 5

<211> 55

<212> DNA

<213> Artificial

<220>

<223> Primer

<400> 5
aggagaggtc tagaccactt tgtacaagaa agctgggtgg atccggactg ttttt 55

<210> 6
<211> 49
<212> DNA
<213> Artificial

<220>
<223> Primer RNAII

<400> 6
ctcgacctcg agagcctgct tttttgtaca aacttggtgc ccggtaccc 49

<210> 7
<211> 40
<212> DNA
<213> Artificial

<220>
<223> Ori-F

<400> 7
aatcgacgct caagtcagag gtggcgaaac ccgacaggac 40

<210> 8
<211> 40
<212> DNA
<213> Artificial

<220>
<223> Amp-F

<400> 8
ccttcggtcc tccgatcggt gtcagaagta agttggccgc 40

<210> 9
<211> 49
<212> DNA
<213> Artificial

<220>
<223> FLCII-1

<400> 9
gggtaccggg ccacaagttt gtacaaaaaa gcaggctctc gaggtcgag 49

<210> 10
<211> 55
<212> DNA
<213> Artificial

<220>
<223> FLCII-2

<400> 10
aaaaacagtc cggatccacc cagctttctt gtacaaagtg gtctagacct ctcct 55

<210> 11
<211> 40
<212> DNA
<213> Artificial

<220>

<223> Ori-F

<400> 11

aatcgacgct caagtcagag gtggcgaaac ccgacaggac

40

<210> 12

<211> 40

<212> DNA

<213> Artificial

<220>

<223> Amp-F

<400> 12

ccttcggtcc tccgatcgtt gtcagaagta agttggccgc

40

<210> 13

<211> 49

<212> DNA

<213> Artificial

<220>

<223> FLCII-1

<400> 13

gggtaccggg ccacaagttt gtacaaaaaa gcaggctctc gaggtcgag

49

<210> 14

<211> 55

<212> DNA

<213> Artificial

<220>

<223> FLCII-2

<400> 14

aaaaacagtc cggatccacc cagctttctt gtacaaagtg gtctagacct ctcct

55

<210> 15

<211> 17

<212> DNA

<213> Artificial

<220>

<223> M13 Reverse primer

<400> 15

caggaaacag ctatgac

17

<210> 16

<211> 16

<212> DNA

<213> Artificial

<220>

<223> M13 (-20) Forward primer

<400> 16

gtaaaacgac ggccag

16

<210> 17

<211> 799

<212> DNA

<213> Homo sapiens

<220>
<221> misc_feature
<222> (677)..(677)
<223> n is any nucleotide

<220>
<221> misc_feature
<222> (682)..(682)
<223> n is any nucleotide

<400> 17
accagggggg atcctgggat tctgcctttc cggaaggtcg ggctggctct cggtgctagg 60
ccgccgttct aggcccaagt cgggattcca ggccctgctc caagccaagt ccgtagccgc 120
gacgacacag cggggatcgc agaggacgcg gccgcagctc gttgggggtg ctgagccttg 180
gagggcggca gccgggggtgc gcaggccgtg acggcccctc ccccgccctg agcgcagcgg 240
cgaaggacga cagccccgga ccggaggagg tgaagcgtc acgtaactgc cccgggtgat 300
gactcaccgc ctctcctggc agatttgagg tgcgtgctgc tgttgctgca gggcggacgt 360
cgggtgcccc agagcgggtc tgcggcgacc gcttgcttgc ccttggttcg gcggccccgg 420
ggtttccga gagcgttctc ttgtaacgt ctgcggcctg cagctcgag cgagctccgg 480
cttctcccat gaaggtgtct ctgggtaacg gcgacatggg cgtctctgcc cacctgcagc 540
cttgcaagtc tggaactacg cggtttttta ccagcaacac gcacagttcg gttgtattgc 600
aaggctttga tcagctcaga atagagggcc tactttgtga cgtgacgctg gtacctggcg 660
atggagagga aatcttnccg gntcatagag ccatgatggc gtctgccagt gattatttta 720
aggcatgtt cactggagga atgaaagaaa aagatttgat gtgcatcaag cttcatgggg 780
tgaacaaagt tggctctgaa 799

<210> 18
<211> 55
<212> DNA
<213> Artificial

<220>
<223> Primer RNAI

<400> 18
aggagaggtc tagaccatt tgtacaagaa agctgggtgg atccggactg ttttt 55

<210> 19
<211> 49
<212> DNA
<213> Artificial

<220>
<223> Primer RNAII

<400> 19
ctcgacctcg agagcctgct tttttgtaca aacttggtgc ccggtaccc 49

<210> 20
<211> 40
<212> DNA
<213> Artificial

<220>

<223> Ori-R

<400> 20

gtcctgtcgg gtttcgccac ctctgacttg agcgtcgatt

40

<210> 21

<211> 40

<212> DNA

<213> Artificial

<220>

<223> Amp-R

<400> 21

gcggccaact tacttctgac aacgatcgga ggaccgaagg

40

INTERNATIONAL SEARCH REPORT

Intern: Application No
PCT/JP2004/010174

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12Q1/68

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12Q

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

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 02/31190 A (GENSET SA ; THILL GILBERT (FR)) 18 April 2002 (2002-04-18) page 1 - page 38; example 2	1-32, 42-69
Y	US 2003/165931 A1 (EDON FLORENCE ET AL) 4 September 2003 (2003-09-04) the whole document	1-32, 42-69
Y	US 5 846 721 A (BONALDO MARIA DE FATIMA ET AL) 8 December 1998 (1998-12-08) the whole document	1-69
P,X	WO 2004/053163 A (HAYASHIZAKI YOSHIHIDE ; RIKEN (JP); CARNINCI PIERO (JP); DNAFORM KK (J) 24 June 2004 (2004-06-24) cited in the application the whole document	1-32, 42-69
	-/--	



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *&* document member of the same patent family

Date of the actual completion of the international search

22 March 2005

Date of mailing of the international search report

06 APR 2005

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Botz, J

INTERNATIONAL SEARCH REPORT

Intern I Application No
PCT/JP2004/010174

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	SCHWEIGHOFFER F ET AL: "QUALITATIVE GENE PROFILING: A NOVEL TOOL IN GENOMICS AND IN PHARMACOGENOMICS THAT DECIPHERS MESSENGER RNA ISOFORMS DIVERSITY" PHARMACOGENOMICS, ASHLEY PUBLICATIONS, GB, vol. 1, no. 2, 2000, pages 187-197, XP008024170 ISSN: 1462-2416 the whole document -----	1-32, 42-69
A	BRACCO L ET AL: "The relevance of alternative RNA splicing to pharmacogenomics" TRENDS IN BIOTECHNOLOGY, ELSEVIER PUBLICATIONS, CAMBRIDGE, GB, vol. 21, no. 8, August 2003 (2003-08), pages 346-353, XP004442025 ISSN: 0167-7799 cited in the application the whole document -----	1-32, 42-69
X	WO 02/00934 A (MOLECULAR STAGING, INC) 3 January 2002 (2002-01-03) pages 10,11,16; claims 1,11,13,15,24 -----	33-41
Y	WO 92/01813 A (SYNGENE, INC) 6 February 1992 (1992-02-06) the whole document -----	33-41
Y	DEAN F B ET AL: "Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification" GENOME RESEARCH, COLD SPRING HARBOR LABORATORY PRESS, US, vol. 11, no. 6, June 2001 (2001-06), pages 1095-1099, XP002223174 ISSN: 1088-9051 the whole document -----	33-41

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2004/010174

Box II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

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

2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

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

Box III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

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

see additional sheet

1. ☒ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

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

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

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

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

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

1. claims: 1-32, 42-69

Methods for (i) isolating nucleic acid molecules whose sequences include a portion complementary to each other and a mutually unrelated portion, (ii) isolating nucleic acid isoforms using the hybrid molecules that has a single stranded loop portion obtained by the method of any of claims 1 - 32 from fragmented genomic DNA, cDNA, full-length cDNA, mRNA and/or RNA (iii) preparing a polypeptide derived from an isoform obtained according to claims 42 or 43 (iv) determining sequence variations of an isoform isolated according to claim 42 or 43, comprising the step of sequencing a full-length or partial sequence of isoform (v) a method for identifying nucleic acids (vi) a method for cloning a nucleic acid (vii) a cloning vector (viii) a host cell;

2. claims: 33-41 (completely)

A method for preparing circular single-stranded nucleic molecules;

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/JP2004/010174

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 0231190	A	18-04-2002	AU 7663201 A	22-04-2002
			CA 2425545 A1	18-04-2002
			EP 1379683 A2	14-01-2004
			WO 0231190 A2	18-04-2002
			JP 2004511236 T	15-04-2004
			US 2004106140 A1	03-06-2004
			US 2002055109 A1	09-05-2002
US 2003165931	A1	04-09-2003	FR 2775984 A1	17-09-1999
			WO 2004040018 A2	13-05-2004
			AU 760656 B2	22-05-2003
			AU 2839999 A	27-09-1999
			CA 2323231 A1	16-09-1999
			EP 1062364 A1	27-12-2000
			WO 9946403 A1	16-09-1999
			JP 2002505887 T	26-02-2002
			US 6251590 B1	26-06-2001
			US 2004191828 A1	30-09-2004
US 5846721	A	08-12-1998	NONE	
WO 2004053163	A	24-06-2004	JP 2004187606 A	08-07-2004
			WO 2004053163 A1	24-06-2004
WO 0200934	A	03-01-2002	US 6323009 B1	27-11-2001
			AU 6872501 A	08-01-2002
			CA 2410951 A1	03-01-2002
			EP 1315833 A2	04-06-2003
			JP 2004512823 T	30-04-2004
			WO 0200934 A2	03-01-2002
			US 2003207267 A1	06-11-2003
WO 9201813	A	06-02-1992	AU 649066 B2	12-05-1994
			AU 8417391 A	18-02-1992
			CA 2087256 A1	26-01-1992
			EP 0542874 A1	26-05-1993
			JP 6500014 T	06-01-1994
			NO 930235 A	22-01-1993
			WO 9201813 A1	06-02-1992